MODELING DATA NETWORKS

SIDNEY RESNICK

ABSTRACT. Data networks offer a fascinating, if somewhat potentially frustrating, setting for many applied probability and extreme value techniques to be applied. We survey some of the basic models and statistical techniques for fitting the models. We point out some of the shortcomings in the models and in the statistical techniques. The required range of techniques is broad. The ability to contribute in internet time is questionable.

CONTENTS

Sidney Resnick's research was partially supported by NSF grant DMS-97-04982 at Cornell University.

1. Introduction.

The story begins around 1993 with the publication of what is now known as the Bellcore study ([54, 100, 164]). Traditional queueing models had thrived on assumptions of exponentially bounded tails, Poisson inputs and lots of independence. Collected network data studied at what was then Bellcore (now Telcordia) exhibited properties which were inconsistent with traditional queueing models. These anomalies were also found in world wide web downloads in the Boston University study ([36, 28, 29, 30, 33, 31, 32]). The unusual properties found in the data traces included:

- self-similarity (ss) and long-range dependence (LRD) of various transmission rates:
  - packet counts per unit time,
  - www bits/time.
- heavy tails of quantities such as
  - file sizes,
  - transmission rates,
  - transmission durations,
  - CPU job completion times,

– call lengths

The Bellcore study in early 90's resulted in a paradigm shift worthy of a sociological study to understand the frenzy to jump on and off various bandwagons but after some resistance to the presence of long range dependence, there was widespread acceptance of the statement that *packet counts per unit time exhibit self similarity and long range dependence.* Research goals then shifted from detection of the phenomena to greater understanding of the causes. The challenges were:

- Explain the origins and effects of long-range dependence and self-similarity.
- Understand some connections between self-similarity, long range dependence and heavy tails. Use these connections to find an explanation for the perceived long range dependence in traffic measurements.
- Begin to understand the effect of network protocols and architecture on traffic. The simplest models, such as the featured infinite source Poisson model, pretend protocols and controls are absent. This is an ambitious goal.
- Say something useful for the purposes of capacity planning.

1.1. **The infinite node Poisson model.** Attempts to explain long range dependence and self-similarity in traffic rates centered around the paradigm: *heavy tailed file sizes cause LRD in network traffic*. Specific models must be used to explain this and the two most effective and simple models were:

- Superposition of on/off processes ([157, 110, 109, 153, 164, 75, 74, 88, 120]). This is described as follows: imagine a source/destination pair. The source sends at unit rate for a random length of time to the destination and then is silent or inactive for a random period. Then the source sends again and when finished is silent. And so on. So the transmission schedule of the source follows an alternating renewal or on/off structure. Now imagine the traffic generated by many source/destination pairs being superimposed and this yields the overall traffic.
- The infinite source Poisson model, sometimes called the $M/G/\infty$ input model ([72, 76, 90, 89, 109, 142, 121, 137]). Imagine infinitely many potential users connected to a single server which processes work at constant rate r. At a Poisson time point, some user begins transmitting work to the server at constant (ugh!) rate which, without loss of generality, we take to be rate 1. The length of the transmission is random with heavy tailed distribution. The length of the transmission may be considered to be the size of the file needing transmission.

Both models have their adherents and the two models are asymptotically equivalent in a manner nobody (to date) has made fully transparent. We will focus on the infinite source Poisson model.

Some good news about the model:

- It is somewhat flexible and certainly simple.
- Since each node transmits at unit rate, the overall transmission rate at time t is simply the number of active users $N(t)$ at t. From classical $M/G/\infty$ queueing theory, we know $N(t)$ is a Poisson random variable with mean $\lambda\mu_{\text{on}}$ where λ is the rate parameter of the Poisson process and μ_{on} is the mean file size or mean transmission length.
- The length of each transmission is random and heavy tailed.
- The model offers a very simple explanation of long range dependence being caused by heavy tailed file sizes.
- The model predicts traffic aggregated over users and accumulated over time $[0, T]$ is approximated by either a Gaussian process (fractional Brownian motion or FBM) or a heavy tailed

stable Lévy motion ([109]). Thus the two approximations are very different in character but at least both are self-similar.

Some less good news about the model:

- The model does not fit collected data traces all that well.
  - The constant transmission rate assumption is clearly wrong. Each of us knows from personal experience that downloads and uploads do not proceed at constant rate.
  - Not all times of transmissions are Poisson. Identifying Poisson time points in the data can be problematic. Some are machine triggered and these will certainly not be Poisson. While network engineers rightly believe in the *invariant* that behavior associated with humans acting independently can be modeled as a Poisson process, it is highly unlikely that, for example, subsidiary downloads triggered by going to the CNN website (imagine the calls to DoubleClick's ads) would follow a Poisson pattern.
- There is no hope that this simple model can successfully match fine time scale behavior observed below, say, 100 milliseconds. Below this time scale threshold, observational studies speculate that traffic exhibits multifractal characteristics.
- The model does not take into account admission and congestion controls such as TCP. How can one incorporate a complex object such as a control mechanism into an informative probability model?

1.2. **Broad Issues (BI's) to consider for data network modeling.** Before considering further the infinite node Poisson model, consider the following issues related to data network modeling and data analysis.

BI 1. The problem of research time scales: How do the disciplines of applied probability, statistics and applied mathematics make contributions to data network analysis and planning in *internet time*? The rapid pace of development makes it difficult for mathematical research to find a niche. If you propose a research project lasting perhaps 2–3 years to an internet engineer, their eyes will glaze over. The world will be a different place by the time the project is completed and your project results may or may not have applicability. How many of us remember using Mosaic as our browser? The year 1994 may be fresh in some of our minds but is the dark ages of the internet. Several of the data sets analyzed in [72] are quite old (of the order of 7 years and the clock is still ticking) and this raises the question of their relevance. Are these measurements an accurate reflection of current traffic? Maybe not. If not, are the methods collected and developed to analyze the old data sets at least useful for analyzing current measurements? Hopefully.

Is the most meaningful contribution we can make that we simply help cause and justify paradigm shifts with explanations which may lag behind empirical, experimental and heuristic developments?

Consider the following sobering thought. When the monolithic AT&T spun off Lucent and Lucent wound up with Bell Laboratories, the newly constituted AT&T Labs-Research in Florham Park did not stock the building with researchers and engineers from the applied probability community. (My informal count in 1999 of people whose primary affiliation was applied probability was 2, though of course there are many broad and talented people who defy easy classification.) Why did the expertise of the applied probability community not appeal to the organizers of the Florham Park facility?

BI 2. Insider vs Outsider: Do you

- Analyze data that is already available, say on the web (e.g., the ITA web site at http://ita.ee. lbl.gov/html/traces.html) or that you have bootlegged by hook or by crook? This is often what academics (including me) have done. However, such data may be old (dangerous when the internet time clock is ticking), badly suited for the purposes of the study, or just plain dirty. Note, for example, the UCB data analyzed in [72] is NOT very suitable for testing the Poisson assumption.

OR

- Design a network experiment to get the data you want. This typically requires cooperative net administrators and some hardware and software expertise more typically found in the computer science and electrical engineering communities. It may require you to go beyond your local area network to something of the scale of World Net, UUNet etc.

BI 3. How can you DISCERN THE INFLUENCE OF NETWORK ARCHITECTURES AND PROTOCOLS? How do you model the complexities of something like TCP [92] realistically? Can you get a reasonably accurate model which encourages analytic analysis, or if mathematical analysis is too difficult, is the model accurate enough that simulation and experimentation yield fruitful information? Some beginning but not entirely satisfactory attempts to grapple with the dynamics of TCP are contained in [17, 116, 117, 152, 118, 67, 6].

BI 4. A BROAD VARIETY OF TECHNIQUES is useful and a wide background is required. Teamwork may be the way to go. Here is a partial list of skills and expertise that your team's toolbox should have:

- Applied Probability & Statistics:
  - Stochastic processes: FBM, Lévy stable motion and fractional motions,
  - Poisson point process theory,
  - weak convergence,
  - heavy tailed analysis,
  - long-range dependence,
  - self-similarity & multifractality,
  - extreme value theory and associated statistical techniques,
  - estimation methods such as maximum likelihood, exploratory tools using graphical techniques,
  - time series analysis,
  - queueing theory.
- Applied Mathematics:
  - wavelets (seem to be the right tool for examining phenomena on different time scales),
  - numerical methods,
  - design and implementation of simulation tools,
  - computing.

BI 5. BLACK BOX VS STRUCTURAL MODELING: The philosophy behind black box modeling is to provide a broad class of models with enough parameters to fit a variety of data sets. The emphasis is on finding something that fits according to some criterion. For example, the classical Box-Jenkins ARMA approach to fitting time series models is to difference the data until it looks stationary and then fit an ARMA(p, q) model of the form

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + \sum_{i=0}^{q} \theta_i Z_{n-i},$$

where one specifies p, q, ϕ's and θ's to yield a model matching the L_2 sample moments of the data. Traditionally, $\{Z_n\}$ is white noise.

One may try to adapt this class of models to a heavy tailed context in which case it is natural to suppose $\{Z_n\}$ is iid and heavy tailed. However, while such highly structured methods may have some appeal in economics and finance, in other contexts the approach may have drawbacks.

Consider the following problems with the approach.

- The dependence structure in heavy tailed models is complex and there is no reason why a highly structured linear model will model the dependence correctly. Generally you do not get good fits for dependent heavy tailed data with ARMA modeling and one could go out on a limb and assert that *the only heavy tailed data that ARMA fits are simulated data from the ARMA model.*
- Even if the black box approach of fitting something like an ARMA worked acceptably, there would be a tendency to say "so what", since this does not provide fundamental insights into system dynamics. A method which finds a pattern in a bunch of numbers but ignores physics, structure and system dynamics is not so revealing.
- In a rapidly changing environment, spending excessive time just fitting the data to a black box model may not be so useful, since the next data set generated from similar mechanisms may not be fit to the same model. One should not overemphasize this point, but keep in mind the goal is to understand system dynamics and not just to successfully fit a model to the data.

These objections have caused the internet community to emphasize structural modeling which incorporates idealized features of the network. The typical internet trace collects packet headers which contains rather detailed information, and this can hopefully be utilized to good effect.

2. How do heavy tails cause long range dependence?

Understanding the connection between heavy tails and long range dependence requires a context. For the simplest explanations one can choose either the superposition of on/off processes or the infinite node Poisson model, and our preference is for the latter.

2.1. The infinite node Poisson model.

The simplest model which explains the paradigm that heavy tails induce long range dependence is the infinite source Poisson model, sometimes called the M/G/∞ input model.

In this model, there is potentially an infinite number of sources capable of sending work to the server. Imagine that transmission sources turn on and initiate *sessions* or *connections* at Poisson time points $\{\Gamma_k\}$ with rate λ. The lengths of sessions $\{L_n\}$ are iid non-negative random variables with common distribution F_{on} and during a session, work is transmitted to the server at constant rate. As a normalization, we assume the transmission rate is 1. Assume

$$1 - F_{\text{on}}(t) := \bar{F}_{\text{on}}(t) = t^{-\alpha} L(t), \quad t \to \infty; \tag{2.1}$$

for some slowly varying function L, that is,

$$\lim_{t\to\infty} \frac{\bar{F}_{\text{on}}(tx)}{\bar{F}_{\text{on}}(t)} = x^{-\alpha}, \quad x > 0.$$

In practice, empirical estimates of α usually range between 1 and 2 ([164, 101]). However, studies of file sizes sometimes report measurements of $\alpha < 1$ ([142, 7]). The assumption of a fixed unit

transmission rate is one of the more unrealistic aspects of the model, which accords neither with anyone's personal experience nor with measurement studies. Later, in Section 8, we mention how to modify the model in the interests of greater realism so that either

(i) transmission rates are random and possibly dependent on the size of the file to be transmitted; or

(ii) it is assumed that cumulative input from a source follows a random multifractal process.

However, for the present, in the interests of simplicity and for tractability, the fixed transmission rate will be assumed.

Note that in the case of $0 < \alpha < 1$, which we will term the *very heavy tailed* case, both the mean and the variance of F_{on} are infinite. In case $1 < \alpha < 2$, termed merely the *heavy tailed* case, the variance of F_{on} is infinite but

$$\mu_{\text{on}} = E(L_1) = \int_0^\infty \bar{F}_{\text{on}}(t)dt < \infty.$$

The processes of primary interest for describing this system are the following:

$$\begin{aligned} N(t) &= \text{number of sessions in progress at } t \\ &= \text{number of busy servers in the M/G/}\infty \text{ model} \\ &= \sum_{k=1}^{\infty} 1_{[\Gamma_k \le t < \Gamma_k + L_k]} \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} A(t) &= \int_0^t N(s)ds = \text{cumulative input in } [0,t], \\ r &= \text{release rate or the rate at which the server works off the offered load.} \end{aligned} \tag{2.3}$$

Note that expressing $A(t)$ as an integral gives $N(t)$ the interpretation of "instantaneous input rate at time t". So realizations of $N(t)$ correspond to data traces of "packet counts per unit time".

Stability requires us to assume that the long term input rate should be less than the output rate, so we require

$$\lambda \mu_{\text{on}} < r.$$

This means the content or buffer level process $\{X(t), t \ge 0\}$, which satisfies

$$dX(t) = N(t)dt - r1_{[X(t)>0]}dt,$$

is regenerative with finite mean regeneration times and achieves a stationary distribution.

2.2. **Connection between heavy tails and long range dependence.** The common explanation for long range dependence in the total transmission rate by the system is that *high variability causes long range dependence*, where we understand that high variability means heavy tails. The long range dependence resulting from the heavy tailed distribution F_{on} can be easily seen for the infinite node Poisson model.

Assume that $1 < \alpha < 2$. To make our argument transparent, we consider the following background. For each t, $N(t)$ is a Poisson random variable. Why? When $1 < \alpha < 2$, $N(\cdot)$ has a stationary version on $\mathbb{R}$, the whole real line. Assume

$$\sum_k \epsilon_{\Gamma_k} = \text{PRM}(\lambda dt)$$

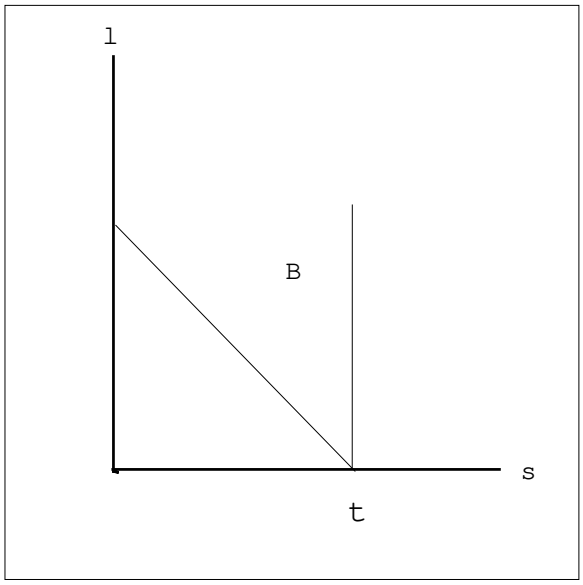

FIGURE 1. The region B

is a homogeneous Poisson random measure on $\mathbb{R}$, with rate λ. Then

$$M := \sum_k \epsilon_{(\Gamma_k, L_k)} = \text{PRM}(\lambda dt \times F_{\text{on}}) \tag{2.4}$$

is a two dimensional Poisson random measure on $\mathbb{R} \times [0, \infty)$ (eg, [139]) with mean measure $\lambda dt \times F_{\text{on}}(dx)$, and

$$\begin{aligned} N(t) =& \sum_k 1_{[\Gamma_k \le t < \Gamma_k + L_k]} \\ =& M(\{(s,l) : s \le t < s + l\} = M(B) \end{aligned}$$

is Poisson because it is the two dimensional Poisson process M evaluated on the region B. See the gorgeous Figure 1. Note B is the region in the (s,l)-plane to the left of the vertical line through $(t,0)$ and above the -45 degree line through $(t,0)$. The mean of $M(B)$ is

$$\begin{aligned} E\Big(M(\{(s,l) : s \le t < s + l\})\Big) &= \iint_{\{(s,l): s \le t < s+l\}} \lambda ds \bar{F}_{\text{on}}(dl) \\ &= \int_{s=-\infty}^{t} \bar{F}_{\text{on}}(t-s) \lambda ds = \lambda \mu_{\text{on}}. \end{aligned} \tag{2.5}$$

Understanding the relation between $\{N(t)\}$ and the random measure M allows us to compute easily the covariance function. Refer to Figure 2. Recall that $N(t)$ corresponds to points to the left of the vertical through $(t,0)$ and above the -45-degree line through $(t,0)$, with a similar interpretation for $N(t+s)$. The process $\{N(t), t \in \mathbb{R}\}$ is stationary with covariance function

$$\begin{aligned} &\text{Cov}(N(t), N(t+s)) \\ &= \text{Cov}\big(M(A_1) + M(A_2), M(A_2) + M(A_3)\big), \end{aligned}$$

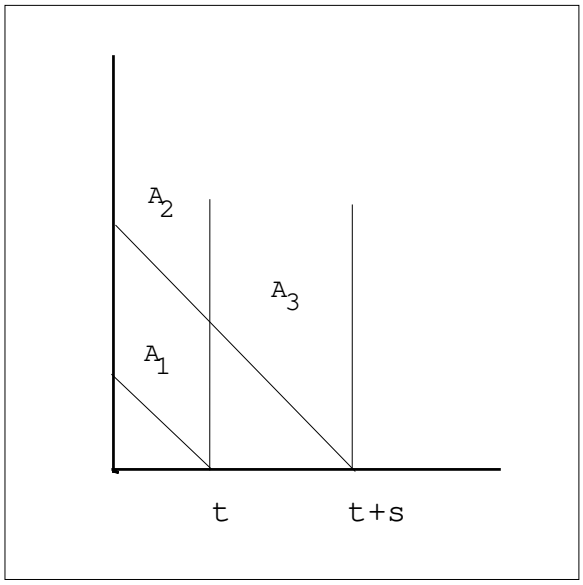

FIGURE 2. The regions A_1, A_2, A_3

and because $M(A_1)$ and $M(A_3)$ are independent, the previous expression reduces to

$$=\mathrm{Cov}\big(M(A_2), M(A_2)\big) = \mathrm{Var}\big(M(A_2)\big).$$

For a Poisson random variable, the mean and the variance are equal, and therefore the above equals

$$\begin{aligned} &= E\big(M(A_2)\big) = \int_{u=-\infty}^{t} \lambda du\ \bar{F}_{\mathrm{on}}(t+s-u) \\ &= \lambda \int_{s}^{\infty} \bar{F}_{\mathrm{on}}(v)dv \sim \frac{\lambda}{\alpha-1} s^{-(\alpha-1)} L(s), \end{aligned}$$

where we used Karamata's theorem for regularly varying functions ([42, 14, 139]). To summarize, we find that

$$\begin{aligned} \mathrm{Cov}(N(t), N(t+s)) &= \lambda \int_{s}^{\infty} \bar{F}_{\mathrm{on}}(v)dv \\ &\sim \frac{\lambda}{\alpha-1} s^{-(\alpha-1)} L(s) \\ &= \frac{\lambda}{\alpha-1} s \bar{F}_{\mathrm{on}}(s), \quad s \to \infty. \end{aligned} \tag{2.6}$$

The slow decay of the covariance as a function of the lag s characterizes *long range dependence*.

3. Further implications of the simple model: Is network traffic stable Lévy motion or fractional Brownian motion?

Our simple Poisson based model offers a compelling explanation of how heavy tailed file sizes induce long range dependence in the traffic rates. To decide if our model is an accurate enough reflection of reality, however, we need to see how well data measurements fit the model. So we require a partial catalogue of features of the model, in order to see if such features are found in

data measurements. In this section, based on [109], we analyze what the model predicts about the cumulative traffic process.

3.1. **Background.** There is strong belief in the self-similar nature of aggregate traffic rates, at least at time scales above a certain threshold. Empirical [164, 7] and theoretical [157, 74, 75, 76] evidence supports the heavy tailed explanation of the self-similarity. However, various authors diverge in their conclusions about the marginal distributions of cumulative traffic. There exists theoretical interest in the conclusion that marginal distributions should be heavy tailed [93, 137] and spirited interest and evidence for Gaussian marginal distributions [101]. The latter result coincides with empirical observations that traffic looks Gaussian on links which are heavily loaded.

The infinite source Poisson model with heavy tailed file sizes allows cumulative traffic at large time scales to look either heavy tailed or Gaussian, depending on whether the rate at which transmissions are initiated (crudely referred to as the *connection rate*) is moderate or quite large.

The process describing offered traffic is $A(t)$, the cumulative input in $[0,t]$ by all sources. Recall from (2.2) and (2.3) that the model assumes unit rate transmissions and $A(t)$ is the integral of $N(s)$ over $[0,t]$. For large T, we think of $(A(Tt), t \geq 0)$ as the process on large time scales. The results show that if the connection rate $\lambda(\cdot)$ is allowed to depend on T in such a way that it has a growth rate in T which is moderate (in a manner to be made precise), then $A(T\cdot)$ looks like an α-stable Lévy motion, while if the connection rate grows faster than a critical value, $A(T\cdot)$ looks like a fractional Brownian motion. We make these statements precise by adopting a heavy traffic outlook and think of a family of models indexed by T, where the T-th model has connection rate $\lambda(T)$ and file size distribution $F_{\rm on}$. Depending on growth rates, the T-th model is approximated by either Lévy stable motion or FBM.

Let $(\Gamma_k, -\infty < k < \infty)$ be the points of the rate λ homogeneous Poisson process on $\mathbb{R}$, labeled so that $\Gamma_0 < 0 < \Gamma_1$ and hence $\{-\Gamma_0, \Gamma_1, (\Gamma_{k+1} - \Gamma_k, k \neq 0)\}$ are iid exponentially distributed random variables with parameter λ. The random measure which counts the points is denoted by $\sum_{k=-\infty}^{\infty} \epsilon_{\Gamma_k}$ and is a Poisson random measure with mean measure $\lambda\mathbb{L}$, where $\mathbb{L}$ stands for Lebesgue measure. The communication system has an infinite number of nodes or *sources*, and at time Γ_k a connection is made and some node begins a transmission at constant rate to the server. As a normalization, this constant rate is taken to be unity. The lengths of transmissions are random variables L_k. Assume $L_{\rm on}, L_1, L_2, \ldots$ are iid and independent of (Γ_k), and

$$P(L_{\rm on} > x) = \bar{F}_{\rm on}(x) = x^{-\alpha} L(x), \quad x > 0,\ 1 < \alpha < 2\,, \tag{3.1}$$

where L is a slowly varying function. Since $\alpha \in (1,2)$, the variance of $L_{\rm on}$ is infinite and its mean $\mu_{\rm on}$ is finite. We will need the quantile function

$$b(t) = \left(1/\bar{F}_{\rm on}\right)^{\leftarrow}(t) =: \inf\{x : \frac{1}{1 - F_{\rm on}(x)} \geq t\}, \quad t > 0, \tag{3.2}$$

which is regularly varying with index $1/\alpha$. Recall the two dimensional Poisson random measure M defined by (2.4) which is a counting function on $\mathbb{R} \times [0,\infty]$ corresponding to the points $\{(\Gamma_k, L_k)\}$ and has mean measure $\lambda\mathbb{L} \times F_{\rm on}$; cf. [139].

To remind us we consider the T-th model, we sometimes subscript quantities by T. So for example, the number of active sources at t, or the overall transmission rate at t is denoted by either $N(t)$ or $N_T(t)$. We will consider a family of Poisson processes indexed by the scaling parameter $T > 0$ such that the intensity $\lambda = \lambda(T)$ goes to infinity as $T \to \infty$. The intensity $\lambda = \lambda(T)$ will be referred to as the *connection rate* for the T-th model.

Recall that heavy tailed transmission times L_k induce long range dependence in N; the precise expression of this is (2.6). High variability in transmission times causes long range dependence in the rate at which work is offered to the system.

3.2. **The critical input rate.** Recall that $\lambda = \lambda(T)$ is the parameter governing the connection rate in the T-th model, and suppose $\lambda = \lambda(T)$ is a non-decreasing function of T. We phrase our condition first in terms of the quantile function b defined in (3.2). The asymptotic behavior of $A_T(\cdot)$ depends on whether

$$\text{Slow Growth Condition 1:} \quad \lim_{T\to\infty} \frac{b(\lambda T)}{T} = 0,$$

or

$$\text{Fast Growth Condition 2:} \quad \lim_{T\to\infty} \frac{b(\lambda T)}{T} = \infty$$

holds. Notice that $b(\cdot)$ is regularly varying with index $1/\alpha$.

There is an alternative, more intuitive, way to express the conditions.

Lemma 1. *Assume F_{on} satisfies* (3.1). *Consider the stationary version of the input rate $N_T(\cdot)$.*

(1) *The slow growth condition* 1 *is equivalent to either of the two conditions*

$$\lim_{T\to\infty} \lambda T\, \bar{F}_{\text{on}}(T) = 0 \quad \textit{or} \quad \lim_{T\to\infty} Cov\,(N_T(0), N_T(T)) = 0. \tag{3.3}$$

(2) *The fast growth condition* 2 *is equivalent to either of the two conditions*

$$\lim_{T\to\infty} \lambda T \bar{F}_{\text{on}}(T) = \infty \quad \textit{or} \quad \lim_{T\to\infty} Cov\,(N_T(0), N_T(T)) = \infty. \tag{3.4}$$

If we think of the scaled process $N_T(t) = N(Tt)$, then the covariance appearing in (3.3) and (3.4) is the lag 1 covariance of $N_T(\cdot)$. So, as we proceed through our family of models indexed by T, under slow growth, the lag 1 covariance is diminishing at large scales, and under fast growth the lag 1 covariance is getting very strong.

Proof. In the case of Condition 1, there exists a function $0 < \epsilon(T) \to 0$ such that $T\epsilon(T) \to \infty$ and $b(\lambda T) = T\epsilon(T)$. Thus

$$\lambda T \sim 1/\bar{F}_{\text{on}}(T\epsilon(T))\,. \tag{3.5}$$

Therefore, Condition 1 implies

$$\lambda T\ \bar{F}_{\text{on}}(T) \sim \bar{F}_{\text{on}}(T)/\bar{F}_{\text{on}}(T\epsilon(T)) \ \to 0\,. \tag{3.6}$$

Conversely, if $\delta(T) := \lambda T \bar{F}_{\text{on}}(T) \to 0$, then using $b^{\leftarrow}(T) \sim 1/\bar{F}_{\text{on}}(T)$, we get

$$\frac{b(\lambda T)}{T} \sim \frac{b(\delta(T) b^{\leftarrow}(T))}{b(b^{\leftarrow}(T))} \to 0\,,$$

and so Condition 1 and (3.6) are equivalent. Similarly, Condition 2 is the same as

$$\lambda T\ \bar{F}_{\text{on}}(T) \to \infty\,. \tag{3.7}$$

To get the equivalence in terms of the covariances, use (2.6). □

We will need the following facts proven in [72]. If Condition 1 holds, then

$$\lim_{T\to\infty} \frac{\lambda T^2 \bar{F}_{\text{on}}(T)}{b(\lambda T)} = 0, \tag{3.8}$$

and if Condition 2 holds, this limit is infinite.

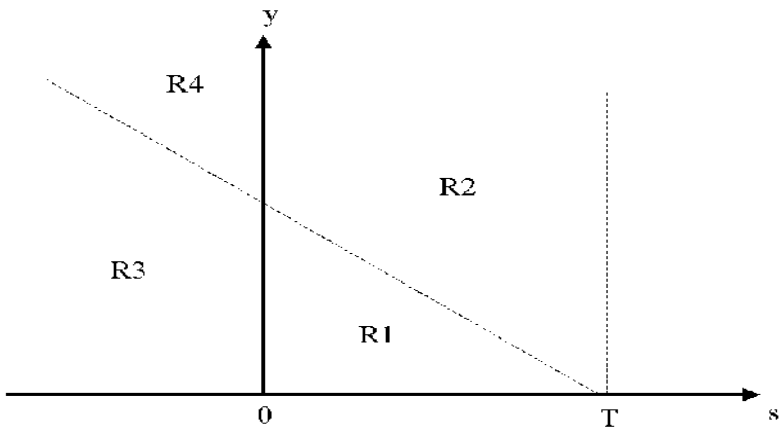

FIGURE 3. The regions R_1, R_2, R_3, R_4.

3.3. **α–stable approximations for the infinite source Poisson model under slow growth.** We now assume Condition 1 holds and show why at large time scales, A is approximately an α-stable Lévy motion; that is, a process with stationary independent increments with α-stable marginals ([151, 64]). The following is the result under the slow growth condition.

Theorem 1. *If Condition* 1 *holds, then the process* $(A(Tt), t \geq 0)$ *describing the total accumulated input in* $[0, Tt]$, $t \geq 0$, *satisfies the limit relation*

$$X^{(T)}(\cdot) := \frac{A(T\cdot) - T\lambda\mu_{\rm on}(\cdot)}{b(\lambda T)} \stackrel{fidi}{\to} X_\alpha(\cdot), \tag{3.9}$$

where $X_\alpha(\cdot)$ *is an* α*-stable Lévy motion. Here* $\stackrel{fidi}{\to}$ *denotes convergence of the finite dimensional distributions.*

Remark. The mode of converge cannot be extended to J_1 convergence in the Skorokhod space $D[0,\infty)$. This follows, for example, from [93] who show that a sequence of processes with a.s. continuous sample paths cannot converge in distribution in $(D[0,\infty), J_1)$ to a process with a.s. discontinuous sample paths. A thorough discussion of this phonomena is in [162]; see also [137].

Here is an outline of some elements of the proof.

3.3.1. *The basic decomposition.* We start by giving a useful decomposition of the random variable $A(T)$ corresponding to a decomposition of $(-\infty, T] \times [0,\infty)$:

$$\begin{aligned} R_1 :=& \{(s,y) : 0 < s \leq T,\ 0 < y,\ s + y \leq T\}, \\ R_2 :=& \{(s,y) : 0 < s \leq T,\ T < s + y\}, \\ R_3 :=& \{(s,y) : s \leq 0,\ 0 < s + y \leq T\}, \\ R_4 :=& \{(s,y) : s \leq 0,\ T < s + y\}, \end{aligned} \tag{3.10}$$

(see Figure 3). Rewrite $A(T)$, using (2.3), as

$$\begin{aligned} A(T) =& \sum_k L_k 1_{[(\Gamma_k, L_k)\in R_1]} + \sum_k (T-\Gamma_k) 1_{[(\Gamma_k, L_k)\in R_2]} + \sum_k (L_k+\Gamma_k) 1_{[(\Gamma_k, L_k)\in R_3]} \\ &+ \sum_k T 1_{[(\Gamma_k, L_k)\in R_4]} \\ =:& A_1 + A_2 + A_3 + A_4 \,. \end{aligned} \tag{3.11}$$

Recall the definition of the PRM M from (2.4) with mean measure $\lambda\mathbb{L} \times F_{\rm on}$. Note that A_i is a function of the points of M in region R_i, and since the R_i's are disjoint, $A_i, i = 1, \ldots, 4$, are mutually independent. Calculations as in (2.5) and use of Karamata's theorem give that as $T \to \infty$

$$\begin{aligned} \lambda m_1 := EM(R_1) =& \lambda \int_0^T F_{\rm on}(T-s)\, ds \sim \lambda T, \\ \lambda m_2 := EM(R_2) =& \lambda \int_0^T \bar{F}_{\rm on}(T-s)\, ds \sim \lambda\, \mu_{\rm on}, \\ \lambda m_3 := EM(R_3) =& \lambda \int_{s=-\infty}^0 \int_{y=-s}^{-s+T} F_{\rm on}(dy)\, ds \sim \lambda \mu_{\rm on} \\ \lambda m_4 := EM(R_4) =& \lambda \int_{s=-\infty}^0 \int_{y=-s+T}^{\infty} F_{\rm on}(dy)\, ds = \lambda \int_T^\infty \bar{F}_{\rm on}(u) du \\ \sim& \lambda T \bar{F}_{\rm on}(T)/(\alpha-1) \to 0. \end{aligned} \tag{3.12}$$

So the mean measure $EM(\cdot)$ restricted to R_i is finite for $i = 1, \ldots, 4$, which implies that the points of $M\big|_{R_i}$ can be represented as a Poisson number of iid random vectors ([140, page 341]):

$$M\big|_{R_i} \stackrel{d}{=} \sum_{k=1}^{P_i} \epsilon_{(t_{k,i}, j_{k,i})}\,, \quad i = 1, \ldots, 4\,,$$

where P_i is a Poisson random variable with mean λm_i, which is independent of the iid pairs $(t_{k,i}, j_{k,i})$, $k \geq 1$, with common distribution

$$\frac{\lambda \mathbb{L}(ds) F_{\rm on}(dy)}{\lambda m_i}\Big|_{R_i} = \frac{\mathbb{L}(ds) F_{\rm on}(dy)}{m_i}\Big|_{R_i}, \tag{3.13}$$

for $i = 1, \ldots, 4$. Notice that the distributions of $((t_{k,i}, j_{k,i}))$ are independent of λ, which only enters into the specification of the mean of P_i, $i = 1, \ldots, 4$. This means that for fixed T, we can represent the A_i's as sums of a Poisson number of iid random variables,

$$\begin{aligned} A_1 &\stackrel{d}{=} \textstyle\sum_{k=1}^{P_1} j_{k,1}\,, & A_2 &\stackrel{d}{=} \textstyle\sum_{k=1}^{P_2} (T - t_{k,2}), \\ A_3 &\stackrel{d}{=} \textstyle\sum_{k=1}^{P_3} (j_{k,3} + t_{k,3}), & A_4 &\stackrel{d}{=} \textstyle\sum_{k=1}^{P_4} T. \end{aligned} \tag{3.14}$$

3.3.2. *Moments of the summands.* We need information about the moments of the summands in (3.14). All the variables are bounded by T, so all moments exist, and we derive the asymptotic form of the moments as $T \to \infty$. For notational ease, we often suppress the dependence on T in the notation, and let (t_i, j_i) be random variables with the same distribution as $(t_{k,i}, j_{k,i})$, for $i = 1, \ldots, 4$.

From (3.12) and (3.13), observe that for $l \geq 1$,

$$\begin{aligned} Ej_1^l =& \int_0^T \int_{y=0}^{T-s} y^l \frac{\mathbb{L}(ds)F_{\rm on}(dy)}{m_1} \sim \frac{1}{T}\int_{s=0}^T \int_{y=0}^{T-s} y^l \; F_{\rm on}(dy)\; ds \\ =& \frac{1}{T}\int_{s=0}^T \left(\int_{y=0}^s y^l F_{\rm on}(dy)\right) ds. \end{aligned} \tag{3.15}$$

For $l=1$, since $\int_0^s yF_{\rm on}(dy) \to \mu_{\rm on}$, we have

$$Ej_1 \to \mu_{\rm on}\,. \tag{3.16}$$

For $l > \alpha$, we have, using first a change of variables and then Karamata's theorem, that

$$\frac{Ej_1^l}{T^l \bar F_{\rm on}(T)} \sim \int_0^1 \int_0^s y^l \; \frac{F_{\rm on}(Tdy)}{\bar F_{\rm on}(T)}\; ds \sim \int_0^1 \int_0^s y^l \; \alpha y^{-1-\alpha}\; dy\; ds = \frac{\alpha}{(l-\alpha)(l-\alpha+1)}. \tag{3.17}$$

We also note that (3.16) and (3.17) imply

$$\frac{\mathrm{Var}(j_1)}{T^2 \bar F_{\rm on}(T)} \sim \int_0^1 \int_0^s y^2 \; \alpha y^{-1-\alpha}\; dy\; ds \; = \frac{\alpha}{(2-\alpha)(3-\alpha)} =: \; \sigma_1^2 \tag{3.18}$$

and

$$\limsup_{T\to\infty} \frac{E|j_1 - Ej_1|^3}{T^3 \bar F_{\rm on}(T)} \leq \limsup_{T\to\infty} \frac{4\Big(Ej_1^3 + \big(Ej_1\big)^3\Big)}{T^3 \bar F_{\rm on}(T)} = \text{const}\,. \tag{3.19}$$

Similar calculations for $T - t_2$ give that for $l \geq 1$,

$$\begin{aligned} E(T-t_2)^l =& \int_{s=0}^T \int_{y=T-s}^\infty (T-s)^l \frac{\mathbb{L}(ds)F_{\rm on}(dy)}{m_2} \sim \frac{1}{\mu_{\rm on}} \int_{s=0}^T \int_{y=T-s}^\infty (T-s)^l \; F_{\rm on}(dy)\; ds \\ =& \frac{1}{\mu_{\rm on}} \int_0^T u^l \bar F_{\rm on}(u) du, \end{aligned}$$

and therefore, for $l \geq 1$, as $T \to \infty$, from Karamata's theorem

$$\begin{aligned} \frac{E(T-t_2)^l}{T^{l+1}\bar F_{\rm on}(T)} \sim& \frac{1}{\mu_{\rm on}} \int_0^1 x^l \frac{\bar F_{\rm on}(Tx)}{\bar F_{\rm on}(T)} dx \\ \sim& \frac{1}{\mu_{\rm on}} \int_0^1 x^{l-\alpha} dx = \frac{1}{\mu_{\rm on}(l-\alpha+1)}. \end{aligned} \tag{3.20}$$

This implies that

$$\frac{\mathrm{Var}(T-t_2)}{T^3 \bar F_{\rm on}(T)} \sim \frac{1}{\mu_{\rm on}(3-\alpha)} =: \; \sigma_2^2 \tag{3.21}$$

and

$$\limsup_{T\to\infty} \frac{E|T - t_2 - E(T-t_2)|^3}{T^4 \bar F_{\rm on}(T)} \leq \text{const}\,. \tag{3.22}$$

Finally,

$$E(j_3 + t_3)^l = \int_{s=-\infty}^0 \int_{y=-s}^{T-s} (y+s)^l \frac{\mathbb{L}(ds)F_{\rm on}(dy)}{m_3} \sim \frac{1}{\mu_{\rm on}} \int_{-\infty}^0 \int_{y=-s}^{T-s} (y+s)^l \; F_{\rm on}(dy)\; ds\,,$$

and thus

$$\begin{aligned}\frac{E(j_3+t_3)^l}{T^{l+1}\bar{F}_{\rm on}(T)} &\sim \frac{1}{\mu_{\rm on}}\int_{s=-\infty}^{0}\int_{y=-s}^{1-s}(y+s)^l\,\frac{F_{\rm on}(Tdy)}{\bar{F}_{\rm on}(T)}\,ds \\ &\sim \frac{1}{\mu_{\rm on}}\int_{s=-\infty}^{0}\int_{y=-s}^{1-s}(y+s)^l\,\alpha y^{-1-\alpha}\,dy\,ds.\end{aligned} \tag{3.23}$$

It follows that

$$\frac{\mathrm{Var}(j_3+t_3)}{T^3\bar{F}_{\rm on}(T)} \sim \frac{1}{\mu_{\rm on}}\int_{s=-\infty}^{0}\int_{y=-s}^{1-s}(y+s)^2\,\alpha y^{-1-\alpha}\,dy\,ds =: \sigma_3^2$$

and

$$\limsup_{T\to\infty}\frac{E|j_3-t_3-E(j_3-t_3)|^3}{T^4\bar{F}_{\rm on})T)} \le \text{const}\,.$$

To compute σ_3^2, observe that

$$\begin{aligned}\sigma_3^2 :=& \frac{1}{\mu_{\rm on}}\int_{s=0}^{\infty}\int_{y=s}^{1+s}(y-s)^2\alpha y^{-\alpha-1}dyds \\ =& \frac{1}{\mu_{\rm on}}\int_{y=0}^{1}\alpha y^{-\alpha-1}\Big[\int_{s=0}^{y}(y-s)^2ds\Big]dy \\ &+\frac{1}{\mu_{\rm on}}\int_{y=1}^{\infty}\alpha y^{-\alpha-1}\Big[\int_{s=y-1}^{y}(y-s)^2ds\Big]dy \\ =& \frac{1}{\mu_{\rm on}}\int_{y=0}^{1}\alpha y^{-\alpha-1}\Big[\frac{y^3}{3}\Big]dy+\frac{1}{\mu_{\rm on}}\int_{y=1}^{\infty}\alpha y^{-\alpha-1}\Big[\frac{1}{3}\Big]dy \\ =& \frac{1}{\mu_{\rm on}}\Big[\frac{\alpha}{3(3-\alpha)}+\frac{1}{3}\Big]=\frac{1}{\mu_{\rm on}(3-\alpha)}.\end{aligned} \tag{3.24}$$

This may not be remarkably thrilling, but is quite useful.

3.3.3. *α-stable limits: one dimensional convergence.* We show under Condition 1 that $A(T)$ is asymptotically an α-stable random variable by showing that $A_1(T)=A_1$ is asymptotically stable and $A_i(T)=A_i$, $i=2,3,4$ are asymptotically negligible.

It is relatively easy to see that

$$A_i/b(\lambda T)\xrightarrow{P}0\,,\quad i=2,3,4\,. \tag{3.25}$$

We restrict ourselves to the case $i=2$; a similar argument works for $i=3,4$. By (3.20), (3.8) and Condition 1,

$$EA_2 = EP_2\,E(T-t_2) = [\lambda m_2]\,E(T-t_2) \sim (\text{const})\;\lambda T^2\bar{F}_{\rm on}(T) = o(b(\lambda T))\,.$$

Thus it remains to consider A_1. Recall the representation of A_1 given in (3.14). We start with the following decomposition:

$$A_1-\lambda\mu_{\rm on}T = \sum_{k=1}^{P_1}(j_{k,1}-Ej_1)+Ej_1\,[P_1-EP_1]+[EA_1-\lambda\mu_{\rm on}T] = A_{11}+A_{12}+A_{13}\,.$$

By (3.16), $Ej_1 \sim \mu_{\rm on}$. Since P_1 is Poisson with mean $\lambda m_1 \to \infty$, it satisfies the central limit theorem, i.e.

$$[\lambda m_1]^{-1/2}\,[P_1 - \lambda m_1] \stackrel{d}{\to} N(0,1)\,. \tag{3.26}$$

We conclude that

$$A_{12} = O_P([\lambda T]^{1/2}) = o_P(b(\lambda T)), \tag{3.27}$$

since

$$\lim_{T\to\infty} \frac{\sqrt{\lambda T}}{b(\lambda T)} = \lim_{s\to\infty} \frac{s^{1/2}}{b(s)} = \lim_{s\to\infty} s^{1/2-1/\alpha}/L(s)$$

and $1 < \alpha < 2$ implies $\frac{1}{2} < \frac{1}{\alpha} < 1$.

By (3.14) and (3.26), A_{11} is a sum of approximately $\lambda m_1 \sim \lambda T$ iid summands. Under Condition 1, $b(\lambda T)/T \to 0$, so that for any $x > 0$ fixed, we eventually have $T - b(\lambda T)x > 0$. Therefore, from (3.13)

$$\begin{aligned}
\lambda T P(j_1 > b(\lambda T)\,x) =& \lambda T \iint_{\substack{0\le s\le T\\ 0\le s+y\le T\\ y> b(\lambda T)x}} \frac{ds F_{\rm on}(dy)}{m_1} = \lambda T \int_{s=0}^{T-b(\lambda T)x} ds \int_{y=b(\lambda T)x}^{T-s} \frac{F_{\rm on}(dy)}{m_1} \\
=& \lambda T\Big[\frac{1}{m_1}\bar F_{\rm on}(b(\lambda T)x)(T - b(\lambda T)x) - \frac{1}{m_1}\int_0^{T-b(\lambda T)x} \bar F_{\rm on}(T-s)ds\Big] \\
\sim& \left(1 - \frac{b(\lambda T)x}{T}\right)\ \lambda T\ \bar F_{\rm on}(b(\lambda T)\,x) - \frac{b(\lambda T)}{T}\ \int_x^{T/b(\lambda T)} \lambda T\ \bar F_{\rm on}(b(\lambda T)\,s)\ ds \\
\sim& x^{-\alpha}\,.
\end{aligned}$$

Following an argument using point processes ([139, 127, 138, 128, 137, 108, 107]) we get for $t \ge 0$,

$$Y^{(T)}(\cdot) := \big(b(\lambda T)\big)^{-1} \sum_{k=1}^{[\lambda T\cdot]} (j_{k,1} - Ej_1) \Rightarrow X_\alpha(\cdot) \quad \text{in } \mathbb{D}[0,\infty), \tag{3.28}$$

where the limit is a totally skewed α-stable Lévy random motion ($p = 1,\ q = 0$). In fact, by independence, we may couple (3.26) and (3.28) to get joint convergence

$$\left(Y^{(T)}(\cdot), \frac{P_1}{\lambda T}\right) \Rightarrow (X_\alpha(\cdot), 1) \quad \text{in } \mathbb{D}[0,\infty)\times\mathbb{R}.$$

Using composition and the continuous mapping theorem, one obtains

$$\big(b(\lambda T)\big)^{-1} A_{11} = Y^{(T)}(P_1/(\lambda T)) = \big(b(\lambda T)\big)^{-1} \sum_{i=1}^{P_1} (j_{k,1} - Ej_1) \Rightarrow X_\alpha(1). \tag{3.29}$$

It remains to consider A_{13}. By (3.15) and Karamata's theorem,

$$\begin{aligned}
A_{13} =& E(A_1) - \lambda\mu_{\rm on} T = Ej_1\ EP_1 - \lambda T \mu_{\rm on} \\
=& \lambda \int_0^T \left[\int_0^s y\ F_{\rm on}(dy) - \mu_{\rm on}\right] ds = -\lambda \int_0^T \int_s^\infty y\ F_{\rm on}(dy)\ ds \\
\sim& -\,(\text{const})\ \lambda T^2 \bar F_{\rm on}(T) = o(b(\lambda T))\,.
\end{aligned} \tag{3.30}$$

The last limit relation follows from (3.8). Combining the limit relations (3.25), (3.27), (3.29) and (3.30), we conclude that $A(T)$ has the desired α-stable limit.

3.3.4. *α–stable limits: finite dimensional convergence.* We restrict ourselves to a sketch of the convergence of the 2-dimensional distributions; the general case is analogous. Suppose $t_1 < t_2$. The same arguments as for the one dimensional convergence show that it suffices to consider the joint convergence of $[b(\lambda T)]^{-1}(A_1(Tt_i) - \lambda T t_i\, \mu_{\rm on})$, $i = 1, 2$. We can write

$$\begin{aligned} A_1(Tt_2) =& A_1(Tt_1) + \sum_{Tt_1 < \Gamma_k \le Tt_2} L_k 1_{[\Gamma_k + L_k \le Tt_2]} + \sum_{\Gamma_k \le Tt_1} L_k 1_{[Tt_1 < \Gamma_k + L_k \le Tt_2]} \\ =:& A_1(Tt_1) + A_{21}(T(t_2 - t_1)) + A_{22}\,. \end{aligned}$$

Observe that $A_1(Tt_1)$ and $A_{21}(T(t_2 - t_1))$ are independent and that $A_{21}(T(t_2 - t_1)) \stackrel{d}{=} A_1(T(t_2 - t_1))$. Hence the proof of the 2-dimensional distributions follows from the 1-dimensional convergence if one can show that $[b(\lambda T)]^{-1} A_{22} \stackrel{P}{\to} 0$. This follows by arguments similar to earlier ones as one shows $EA_{22} = o(b(\lambda T))$ using (3.8). □

3.4. **FBM approximations for the infinite source Poisson model under fast growth.** We now study why fast connection rates associated with strong correlations of $N_T(\cdot)$ imply fractional Brownian motion limits.

Recall that a mean-zero Gaussian process $(B_H(t), t \ge 0)$ with a.s. continuous sample paths is called *fractional Brownian motion* if it has covariance structure

$$\mathrm{Cov}(B_H(t), B_H(s)) = \frac{\sigma_H^2}{2}\left(|t|^{2H} + |s|^{2H} - |t - s|^{2H}\right) \quad \text{for some } \sigma_H > 0,\ H \in (0, 1).$$

The case $H = 1/2$ corresponds to Brownian motion and, if $H \in (1/2, 1)$, the autocovariance function of the increment process $(B_H(t) - B_H(t-1))_{t=1,2,\ldots}$, the so-called *fractional Gaussian noise*, exhibits long range dependence. The following is the result under the fast growth condition.

Theorem 2. *If Condition* 2 *holds, then the process* $(A(Tt), t \ge 0)$ *describing the total accumulated input in* $[0, Tt]$, $t \ge 0$, *satisfies the limit relation*

$$\frac{A(T\cdot) - \lambda \mu_{\rm on} T(\cdot)}{[\lambda T^3 \bar{F}_{\rm on}(T)\sigma^2]^{1/2}} \Rightarrow B_H(\cdot).$$

Here $\Rightarrow$ *denotes weak convergence in* $(\mathbb{D}[0, \infty), J_1)$, B_H *is standard fractional Brownian motion,* $H = (3 - \alpha)/2$ *and* σ^2 *is given by*

$$\begin{aligned} \sigma^2 =& \frac{\alpha}{(2 - \alpha)(3 - \alpha)} + \frac{2}{\mu_{\rm on}(3 - \alpha)} + \frac{1}{\alpha - 1} \\ =& \frac{1}{3 - \alpha}\Big[\frac{\alpha}{2 - \alpha} + \frac{2}{\mu_{\rm on}}\Big] + \frac{1}{\alpha - 1}. \end{aligned} \tag{3.31}$$

Remark. Notice that $H = (3 - \alpha)/2 \in (0.5, 1)$. Hence the corresponding fractional Gaussian noise sequence of B_H exhibits long range dependence. This is in contrast to Theorem 1 where the limiting process, α-stable Lévy motion, has independent increments.

As for Theorem 1, the decomposition of Section 3.3.1 will be the key for deriving the Gaussian limit. We omit details and refer to [109]. Unlike the slow growth case, each $A_i, i = 1, \ldots, 4$ contributes to the Gaussian limits. Fast growth means there are more contributions in $[0, T]$ making it difficult for a single contribution to dominate as is typical of heavy tailed limits. The moment conditions of Subsection 3.3.2 allow one to prove asymptotic normality using the Lyapunov condition ([64, page 286], [141, page 319]).

3.5. **Covariance calculations for the infinite source Poisson model.** It is interesting to note that the second order structure of

$$A_T(T\cdot) = \int_0^{T\cdot} N_T(s)ds$$

converges to that of FBM. This happens *even under the slow growth condition when* $A_T(T\cdot)$ *is approximated by Lévy stable motion*. This emphasizes how dangerous it is to judge a process [book] by its second order properties [cover].

For $0 \le s < t$ write (suppressing the subscript T)

$$\begin{aligned}
E\big(A(Tt) - A(Ts)\big)^2 =& E\int_{Ts}^{Tt} N(x)dx \int_{Ts}^{Tt} N(y)dy \\
=& T^2 \int_{x=s}^{t}\int_{y=s}^{t} E\big(N(Tx)N(Ty)\big)dxdy \\
=& 2T^2 \iint_{s\le x<y\le t} E\big(N(Tx)N(Ty)\big)dxdy \\
=& 2T^2 \int_{x=s}^{t}\int_{y=x}^{t} \Big(\mathrm{Cov}\big(N(Tx),N(Ty)\big) + \lambda^2\mu_{\mathrm{on}}^2\Big)dxdy \\
=& 2T^2 \left\{ \int_{x=s}^{t}\int_{y=x}^{t} \left(\lambda \int_{w=T(y-x)}^{\infty} \bar{F}_{\mathrm{on}}(w)dw\right)dxdy + \int_{x=s}^{t}\int_{y=x}^{t} \lambda^2\mu_{\mathrm{on}}^2 dxdy \right\} \\
=& I_{s,t} + II_{s,t},
\end{aligned}$$

where we used (2.6). Integrate $II_{s,t}$ to get

$$II_{s,t} = 2T^2\lambda^2\mu_{\mathrm{on}}^2 \frac{(t-s)^2}{2} = T^2\lambda^2\mu_{\mathrm{on}}^2(t-s)^2.$$

For $I_{s,t}$ we have

$$\begin{aligned}
I_{s,t} =& 2T^3 \int_{x=s}^{t}\int_{y=x}^{t} \left(\lambda\int_{y-x}^{\infty} \bar{F}_{\mathrm{on}}(Tw)dw\right)dxdy \\
=& 2T^3\lambda \int_{x=s}^{t}\int_{y=0}^{t-x} \left(\int_{w=y}^{\infty} \bar{F}_{\mathrm{on}}(Tw)dw\right)dxdy,
\end{aligned}$$

and as $T\to\infty$, this is asymptotic to

$$\begin{aligned}
\sim& 2T^{3-\alpha}\lambda L(T) \int_{x=s}^{t} \left(\int_{y=0}^{t-x} \frac{y^{-(\alpha-1)}}{\alpha-1}dy\right)dx \\
=& 2T^{3-\alpha}L(T)\lambda \int_{x=0}^{t-3} \frac{x^{2-\alpha}}{(2-\alpha)(\alpha-1)}dx = 2T^{3-\alpha}\frac{L(T)\lambda(t-s)^{3-\alpha}}{(2-\alpha)(3-\alpha)(\alpha-1)}.
\end{aligned}$$

We conclude

$$\lim_{T\to\infty} \frac{I_{s,t}}{\lambda T^3 \bar{F}_{\mathrm{on}}(T)} = \frac{2}{(3-\alpha)(2-\alpha)(\alpha-1)}(t-s)^{3-\alpha} = K(\alpha)(t-s)^{3-\alpha} \tag{3.32}$$

and

$$II_{s,t} = T^2\lambda^2\mu_{\mathrm{on}}^2(t-s)^2. \tag{3.33}$$

Now observe

$$\begin{aligned}\operatorname{Cov}\big(A(Ts),A(Tt)\big) =& E\big(A(Ts)A(Tt)\big) - EA(Ts)A(Tt)\\ =& E\big(A(Ts)A(Tt)\big) - Ts\lambda\mu_{\text{on}}Tt\lambda\mu_{\text{on}}\end{aligned}$$

(since $E\int_0^t N(u)du = \int_0^t EN(u)du = \int_0^t \lambda\mu_{\text{on}}du = t\lambda\mu_m$)

$$\begin{aligned}=& \frac{1}{2}\left(EA(Ts)^2 + EA(Tt)^2 - E\big(A(Tt)-A(Ts)\big)^2\right) - T^2\lambda^2\mu_{\text{on}}^2 st\\ =& \frac{1}{2}\left(I_{0,s}+I_{0,t}-I_{s,t}+II_{0,s}+II_{0,t}-II_{s,t}\right) - T^2\lambda^2\mu_{\text{on}}^2 st\\ =& \frac{1}{2}\left(I_{0,s}+I_{0,t}-I_{s,t}\right) + \left\{\frac{1}{2}T^2\lambda^2\mu_{\text{on}}^2[s^2+t^2-(t-s)^2] - T^2\lambda^2\mu_{\text{on}}^2 st\right\}\\ =& \frac{1}{2}\left(I_{0,s}+I_{0,t}-I_{s,t}\right) + 0,\end{aligned}$$

and so

$$\lim_{T\to\infty}\frac{\operatorname{Cov}\big(A(Ts),A(Tt)\big)}{\lambda T^3\bar{F}_{\text{on}}(T)} = K(\alpha)\left[s^{3-\alpha}+t^{3-\alpha}-(t-s)^{3-\alpha}\right].$$

If we set $3-\alpha = 2H$ or $\frac{3-\alpha}{2} = H$ then since $1<\alpha<2$, we get $\frac{1}{2}<H<1$.

The conclusion: The second order structure of

$$\frac{A_T(T\cdot)}{\sqrt{\lambda T^3\bar{F}_{\text{on}}(T)}}$$

converges to that of FBM with $H=\frac{3-\alpha}{2}$, but we do not have weak convergence under the slow growth condition.

4. Does the model fit the data? Checking for Poisson, independence, stationarity. Formal and informal statistical techniques.

Now we have a model with some fairly well understood characteristics. Does it do a satisfactory job of explaining reality? Here are several statistical techniques which help decide.

4.1. **How do you identify Poisson time points and validate the choice statistically?** Not all data is particularly well suited to identifying Poisson time points. Points of a homogeneous Poisson process have the following characteristics:

- Interpoint distances are iid,
- Interpoint distances are exponentially distributed.

The quick and dirty (Q&D) methods of checking these characteristics which are most common are

- Check that the sample auto-correlation function (acf) of interpoint distances is approximately 0 for a reasonable number of lags.
- Check the exponential distribution postulate by a qq-plot which plots theoretical quantiles of the exponential distribution against sample quantiles of the empirical distribution function to check if the plot is approximately linear.

The expectation is that

- behavior of lots of humans acting independently is often well modelled by a Poisson process, but

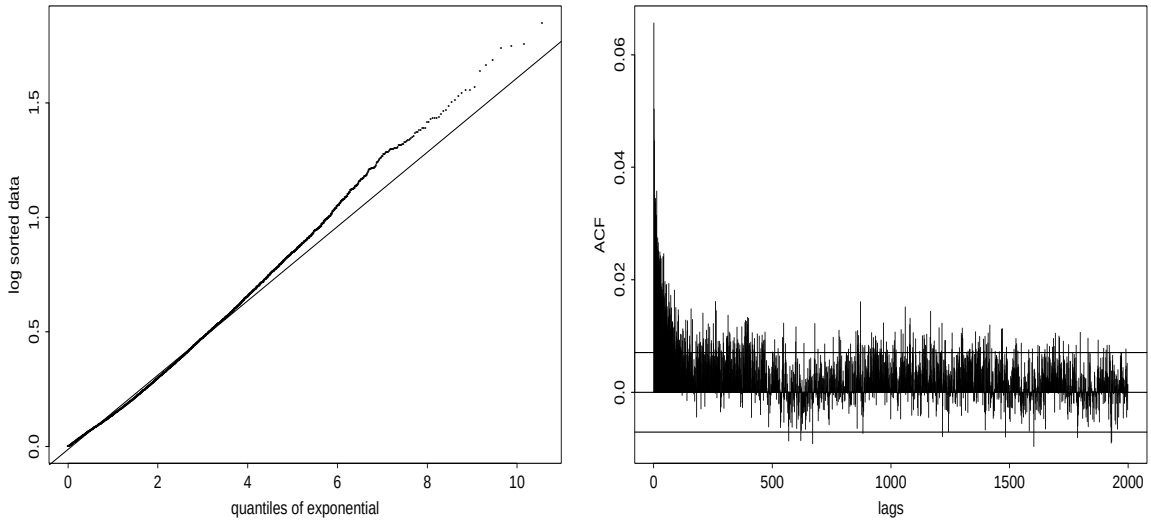

FIGURE 4. UCB inter-arrival times of requests: (*left*) qq-plot against exponential distribution, (*right*) autocorrelation function of interarrival times.

- Initiation times of machine triggered downloads or transmissions cannot be modelled as a Poisson process.

Example 1. THE UCB DATA; HTTP SESSIONS VIA MODEM. The UC Berkeley data is an 18 day trace collected in November 1996. It contains the home IP HTTP traffic processed by UC Berkeley during this period. It is available at http://ita.ee.lbl.gov/html/contrib/UCB.home-IP-HTTP.html. In [72], we analyzed a three hour peak portion of the data, with mean traffic rate of 341 kbit/s. The data content consists of: initiation time of a file transfer, file size, transfer times of a request, IP address of client. Due to the non-stationarity (more on this later) and the diurnal cycle, we restricted the analysis to several hours of peak traffic on a weekday, i.e. the period 5-8 p.m. on Thursday November 7. This part of the trace consists of about 80,000 requests. It is essential to select the period for analysis carefully.

The Figure 4 displays both the acf and qq-plots for times between requests. From the plots, the data does not look convincingly independent (too many spikes protruding from the magic window in the acf plot) nor exponentially distributed (the plot deviates alarmingly from a straight line). Further analysis of the data would be necessary to identify a subset of points as Poisson times. Probably a Poisson cluster process would better fit the data.

4.2. **Checking heavy tailed data for independence.**

4.2.1. *The sample autocorrelation function.* Although there are many pitfalls associated with using the sample auto-correlation function for non-Gaussian data ([61, 126, 25, 130, 131, 135, 41, 37]), it is still the most common method for checking for independence. The sample acf of the stationary sequence $X_1, X_2, \ldots$ is defined as

$$\hat{\rho}(h) = \frac{\sum_{i=1}^{n-h}(X_i - \bar{X})(X_{i+h} - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad h = 1, 2, \ldots,$$

and the method is to plot $\hat{\rho}(h)$, for various lags h and check if the values are all close to 0. Typically we plot at lags $h = 1, \ldots, 25$ and this should be adequate to assess evidence against independence. Of course, an essential point is to give meaning to the phrase *close to 0*. Consider the following two cases.

(a) The variances are finite: Then standard L_2 theory applies and Bartlett's formula from classical time series analysis ([18]) provides asymptotic normality for $\hat{\rho}(h)$ and under the null hypothesis of independence, one constructs for each $\hat{\rho}(h)$ a 95% confidence interval. This leads to the magic window that many mature statistical packages such as Splus automatically plot. If the acf spikes at various lags protrude from the magic window less than 5% of the lags, we shrug comfortably and think so far, so good. No evidence against independence has turned up using this technique.

(b) The variances are infinite: If the data is heavy tailed with infinite variance, the mathematical correlations do not exist. However, under the null hypothesis of independence, formulas of Davis and Resnick ([18, 38, 39, 40]) provide for asymptotic distributions of $\hat{\rho}(h)$ given by the distribution of the ratio of stable random variables. The distribution cannot be calculated explicitly but percentiles of the distribution can be easily simulated and incorporated into a routine. Then, a magic window is plotted and the procedure outlined in (a) can be carried out with this new magic window.

Example 2. Consider 4045 telephone call holding times indexed according to the time of initiation of the call. The range of the call holding data is (2288,11714735). Figure 5 shows the classical sample acf on the left for the call holding data side by side with the heavy tailed modification on the right which does not center the data by $\bar{X}$. The right graph of the heavy tailed sample acf has a dotted line drawn at height $h = 0.0035$ and the interval $[0, h]$ is a 95% confidence window analogous to the one given by Bartlett's formula in classical time series. The confidence window is drawn based on the assumption that the data is independent and has Pareto tails. According to [40, Theorem 3.3], h is given by

$$h = l\alpha^{1/\alpha}\frac{n^{-1/\alpha}}{\log n}$$

where l is the quantile satisfying

$$P[U/V \leq l] = .95$$

for U, V independent positive stable random variables with indices α and $\alpha/2$. In the case of the call holding data, we used the estimated value of $\alpha = .97$. The quantile l was estimated by simulation. The position of h relative to the heights of the sample heavy tailed acf values casts serious doubts on the assumption of independence. Figure 6 exhibits comparable graphs for the packet inter-arrival data. The right heavy tailed graph does not offer evidence against the hypothesis of independence.

4.2.2. *Two quick and dirty (Q&D) methods.* Here are two additional techniques which are often helpful.

(i) TRANSFORM. If data is heavy tailed, take a function of the data (say the log) to get a lighter tail. The advantage is that this puts you in a possibly more familiar domain and you can test for independence using your favorite method. You could, for example, use the classical acf method with Bartlett's formula. The disadvantage, is that you may obscure the importance of large values, which is exactly what makes the data interesting in the first place.

(ii) THE SUBSET METHOD. See ([61, 126, 136]). This is a special case of cross-validation. Split the data into, say, two subsets. Plot the acf of each half separately. If the data comes from an iid

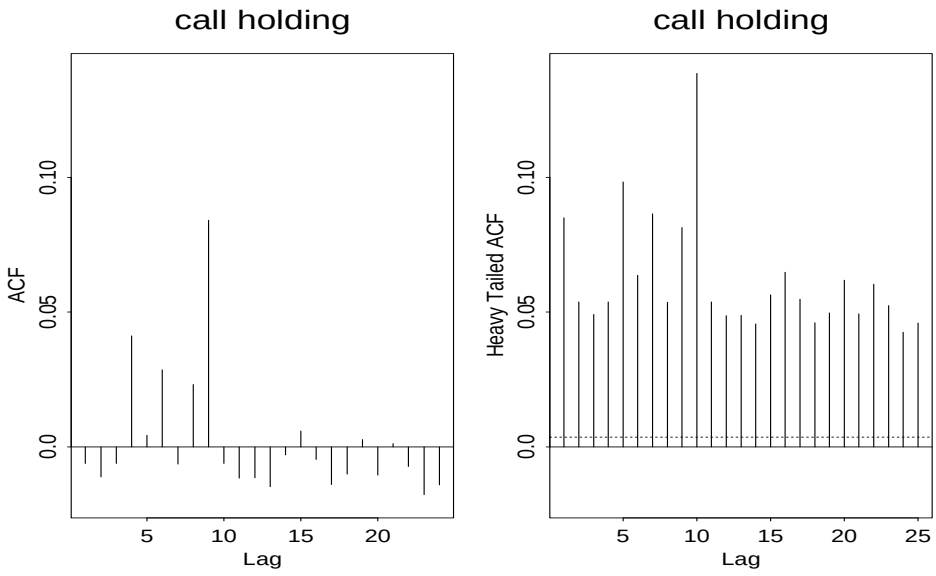

FIGURE 5. Classical acf (left) and heavy tailed modification (right) for call holding time data.

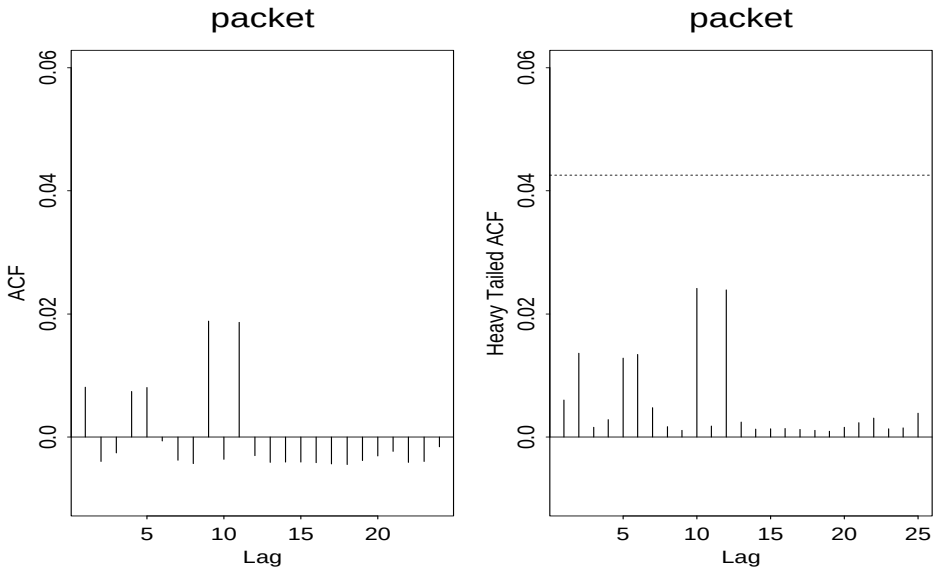

FIGURE 6. Classical acf (left) and heavy tailed modification (right) for packet inter-arrival data.

model, the two plots should look exactly the same even for moderate sample sizes of, for instance, 200. A discussion of how to make this more formal is in [136].

Example 3 (Silence)**.** This data consists of 1026 times between transmissions of packets at a terminal. It is clearly not independent. Figure 7 gives the time series plot and a qq-plot (more later) estimating α.

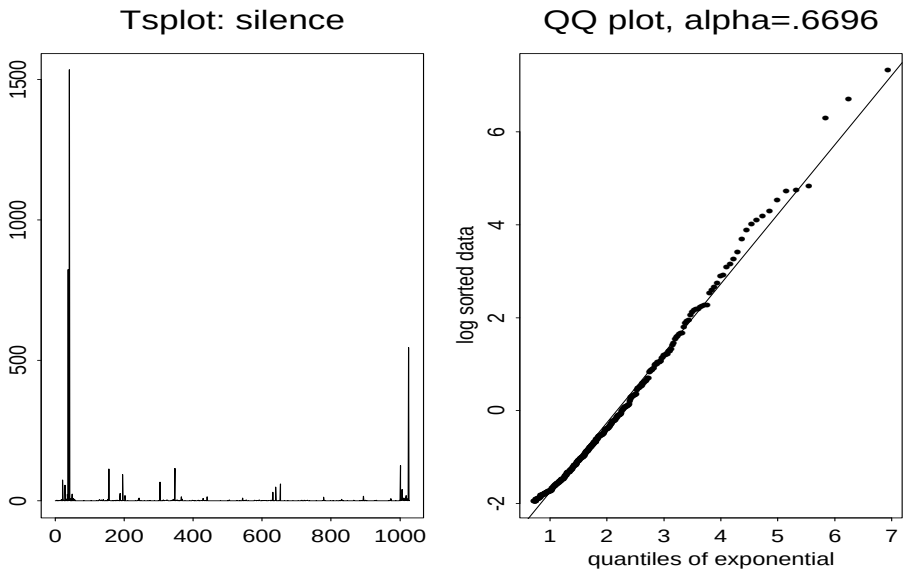

FIGURE 7. Silence: tsplot (left) and qq-plot (right)

Suppose, against good advice, that we try to fit a black box time series model to the data. Good available techniques ([61, 125]) for fitting linear models suggest an autoregression of order 9

$$X_n = \sum_{i=1}^{9} \phi_i X_{n-i} + Z_n, \quad n = 1, 2, \ldots,$$

and where $\{Z_n\}$ are iid. There are many good ways to estimate the coefficients. We used the linear programming method ([59, 60, 61, 58]). A goodness of fit test for the AR(9) model is to fit the coefficients and then estimate the residuals. If the model is correct, the residuals are close to iid. However, that is not the case here. Plotting the sample correlation functions of two disjoint blocks of length 400 gives very different plots, showing a serious problem with the fit.

4.3. **Stationarity.** Network traffic data is often not stationary over the course of, say, 24 hours. There are diurnal cycles and periods when intensity is high and periods when it is low. This is to be expected and discerning this problem does not require any particular genius.

Example 4 (Ericsson)**.** The Ericsson data trace consists of time stamps of starts and completions of the TCP connections that correspond to HTTP file transfers to and from a corporate Ericsson WWW server which holds home pages and information primarily directed to about 2000 company users at Ericsson facilities around the world. The recording started Thursday October 15, 1998 at 15:20, and ended Friday October 16 at 15:49. The information extracted from the data gives the times of connection starts, connection durations, number of bytes transferred (from server to user as well as the opposite direction), and client identification for each connection. The data set is quite nonstationary, and hence a more stationary subset needs to be selected.

Figure 9 displays traffic rates resulting from HTTP requests to the Erisson server over approximately a 24 hour period. The lack of stationarity is clear.

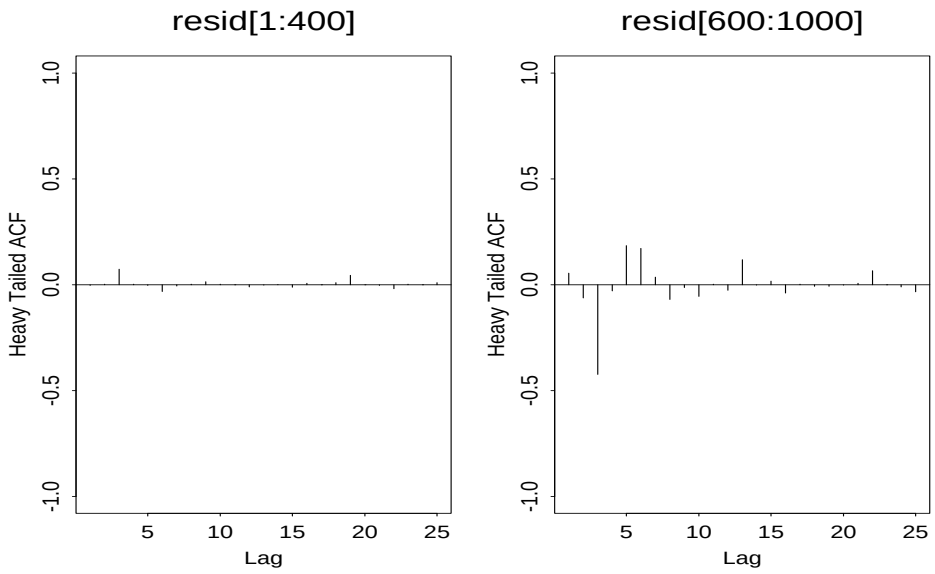

FIGURE 8. Sample acf for two disjoint subsets of SILENCE of length 400.

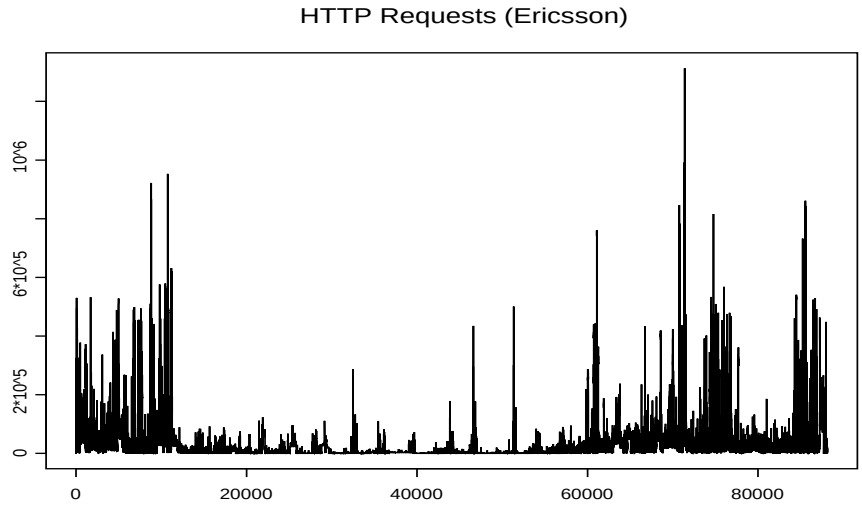

FIGURE 9. Ericsson traffic rates over 24 hours; number of bytes/second.

How does one cope with the lack of stationarity? To date, the most common method is to think that since there is so much data, we might as well take a block of the copious data which "looks stationary". A common rule of thumb is to restrict data to time spans of at most 4 hours.

Obviously, a more sophisticated approach would be to account for non-stationary behavior in the model.

5. Does the model fit the data? How to detect heavy tails.

There are a variety of techniques–none foolproof–which are available for deciding when a heavy tailed model is appropriate:

- Hill plots (a modified maximum likelihood estimation method) and refinements (smoothing, plotting on different scales).
- qq-plots of the empirical quantiles of the log-transformed data against corresponding exponential quantiles.
- Mean residual life plots from exteme value theory, whose popularity is puzzling since they do not work when $\alpha < 1$.
- Extreme value theory techniques such as
  - The Dekkers, Einmahl, de Haan moment estimator [51] of $1/\alpha$.
  - Peaks over threshold modeling leading to fitting a generalized Pareto distribution by the method of maximum likelihood to the exceedances in the data. There is excellent Splus software called EVIS (Extreme Values in Statistics) available (free) from Alexander McNeil's web site (http://www.math.ethz.ch/~mcneil/software.html) as well as the appealing package XTREMES by Reiss and Thomas [124]. A drawback to this, and many other similar methods, is sensitivity to the choice of threshold.

There is not universal agreement on terminology, but we will say that a random variable X has a heavy (right) tail if

$$P[X > x] \sim x^{-\alpha} L(x), \quad x \to \infty, \tag{5.1}$$

where $L(x)$ is slowly varying. The existing statistical techniques would be hard pressed to discern with discrimination the form of $L(x)$ and for many people in the engineering community the phrase *heavy tails* means Pareto or Pareto beyond some point or asymptotically Pareto; that is, $L(x)$ is constant. We will try to resist this simplification.

Recall the following cases:

(i) Very heavy tails: $0 < \alpha < 1$. In this case, assuming the random variable X is positive, both the mean and the second moment of X are infinite. This case is somewhat rare, but file size distributions have been fitted with α in this range. See [7, 142].

(ii) Heavy tails with infinite second moment: $1 < \alpha < 2$. This is a frequently observed case in which the mean is finite but the second moment is infinite. On/off cycle distributions have been fitted with this case [165], and it allows a stationary renewal process to be defined since finite means prevail.

(iii) Heavy tails with finite variance: $\alpha > 2$. This case is typical of financial data.

We now consider the methods outlined above in more detail. We suppose $\{X_n, n \geq 1\}$ is a stationary sequence and that

$$P[X_1 > x] = x^{-\alpha} L(x), \quad x \to \infty$$

where L is slowly varying and $\alpha > 0$. Consistency of estimates of α can usually be proven under just stationarity, but asymptotic normality usually requires the iid assumption or mixing conditions.

5.1. **The Hill estimator and Hill plot.** Let

$$X_{(1)} > X_{(2)} > \cdots > X_{(n)}$$

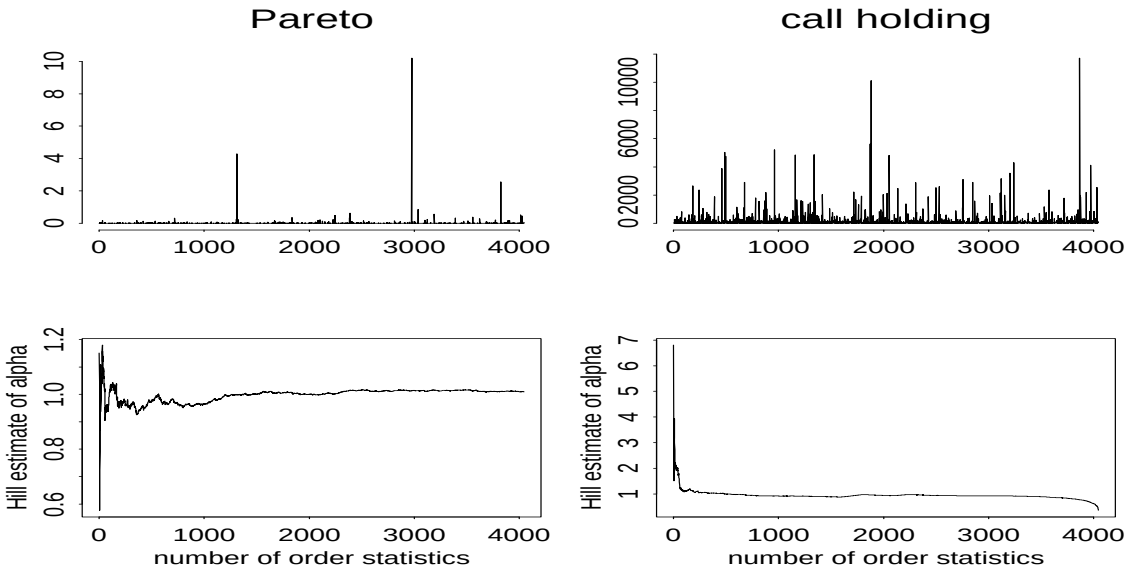

FIGURE 10. Time series and Hill plots for Pareto (left) and call holding (right) data.

be the order statistics of the sample $X_1, \ldots, X_n$. We pick $k < n$ and define the Hill estimator [78] of $1/\alpha$ based on $k+1$ upper order statistics to be

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}}.$$

The number of upper order statistics used in the estimation is $k+1$. The Hill plot is the plot of

$$((k, H_{k,n}^{-1}), 1 \leq k < n).$$

If the process is iid, or a linear MA(∞) or is an ARCH process from economics or more generally satisfies mixing conditions, then the Hill estimator is consistent for $1/\alpha$ in the sense that

$$H_{k,n} \xrightarrow{P} \alpha^{-1}$$

as $n \to \infty$, $k/n \to 0$. The Hill plot should have a stable regime sitting at height roughly α. There is a voluminous literature. See [106, 79, 134, 132, 133, 146, 48, 146, 147]. In the iid case, under a second order regular variation condition, $H_{k,n}$ is asymptotically normal with asymptotic mean $1/\alpha$ and asymptotic variance $1/\alpha^2$. See [66, 47, 34, 35, 49, 50, 122, 44].

5.1.1. *The Hill estimator in practice.* In practice, the Hill estimator is used as follows: We make the *Hill plot*, of

$$\{(k, H_{k,n}^{-1}), 1 \leq k \leq n\}$$

and hope the graph looks stable so you can pick out a value of α.

Sometimes this works beautifully and sometimes the plots are not very revealing. Consider Figure 10 which shows two cases where the procedure is heart-warming. The top row are time series plots. The top left plot is 4045 simulated observations from a Pareto distribution with $\alpha = 1$ and the top right plot is 4045 telephone call holding times indexed according to the time of initiation of the call. Both plots are scaled by division by 1000. The range of the Pareto data is (1.0001,

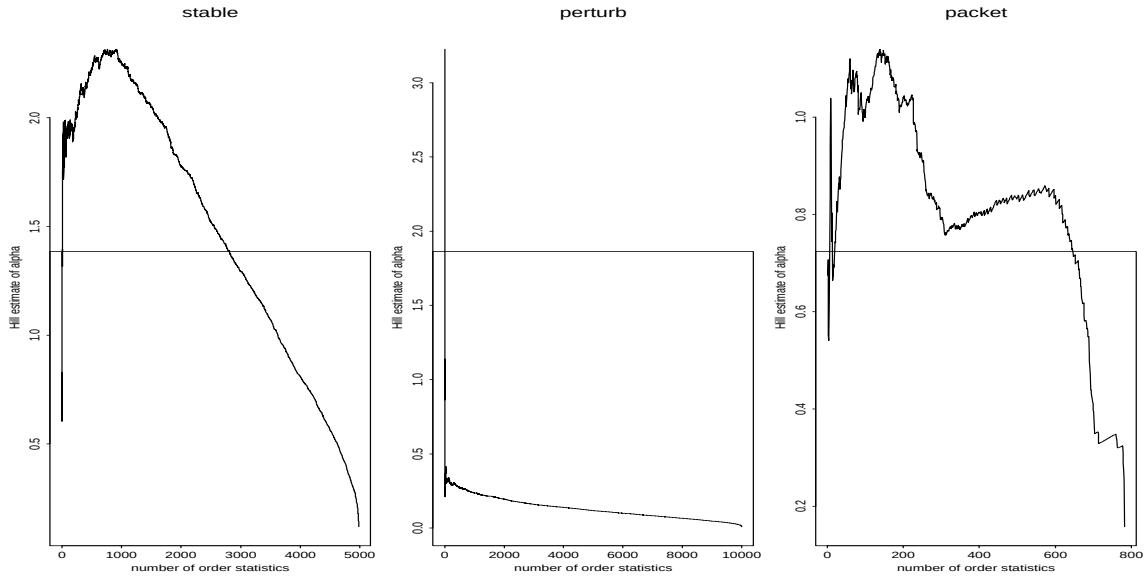

FIGURE 11. A Hill Horror Plot.

10206.477) and the range of the call holding data is (2288,11714735). The bottom two plots are Hill plots $\{(k, H_{k,n}^{-1}), 1 \leq k \leq 4045\}$, the bottom left plot being for the Pareto sample and the bottom right plot for the call holding times. After settling down, both Hill plots are gratifyingly stable and are in a tight neighborhood. The Hill plot for the Pareto seems to nail $\alpha = 1$ correctly and the estimate in the call holding example seems to be between .9 and 1. (So in this case, not only does the variance not exist but the mean appears to be infinite as well.) The Hill plots could be modified to include a confidence interval based on the asymptotic normality of the Hill estimator. McNeil's Hillplot function does just this.

The Hill plot is not always so revealing. Consider Figure 11, one of many Hill Horror Plots. The left plot is for a simulation of size 10,000 from a symmetric α–stable distribution with $\alpha = 1.7$. One would have to be paranormal to discern the correct answer of 1.7 from the plot. The middle plot is for a sample of size 10,000, called *perturb*, from the distribution tail

$$1 - F(x) \sim x^{-1}(\log x)^{10}, \quad x \to \infty,$$

so that $\alpha = 1$. The plot exhibits extreme bias and comes nowhere close to indicating the correct answer of 1. The problem, of course, is that the Hill estimator is designed for the Pareto distribution and thus does not know how to interpret information correctly from the factor $(\log x)^{10}$ and merely readjusts its estimate of α based on this factor rather than identifying the logarithmic perturbation. The third plot is 783 real data called *packet* representing inter-arrival times of packets to a server in a network. The problem here is that the graph is volatile and it is not easy to decide what the estimate should be. The sample size may just be too small.

A summary of difficulties when using the Hill estimator include:

(1) How do you get a point estimate from a graph? What value of k do you use?
(2) The graph may exhibit considerable volatility and/or the true answer may be hidden in the graph.

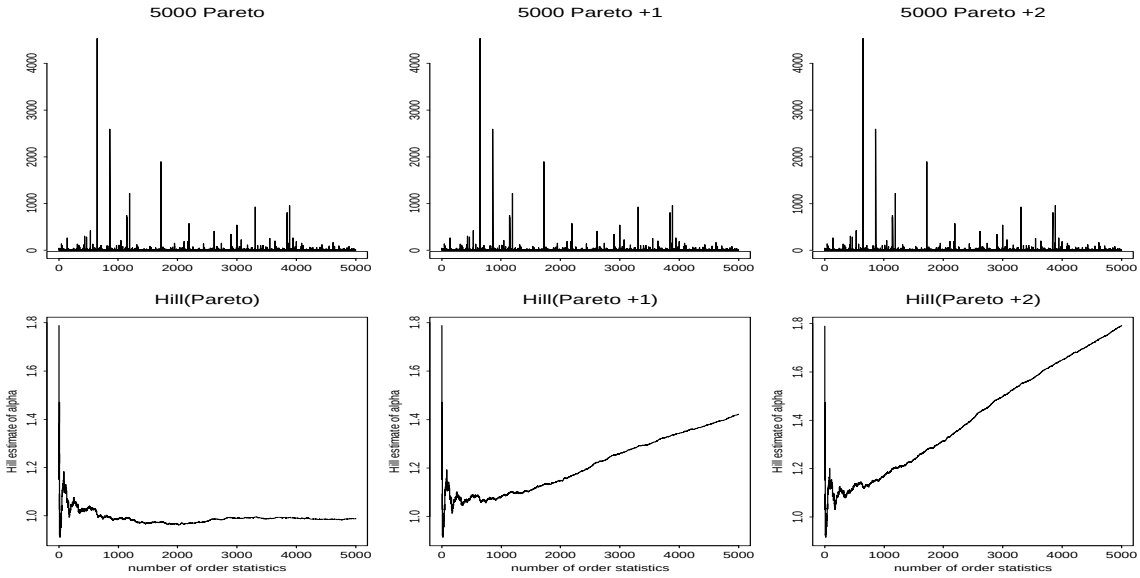

FIGURE 12. Lack of location invariance.

(3) The Hill estimate has optimality properties only when the underlying distribution is close to Pareto. If the distribution is far from Pareto, there may be outrageous error, even for sample sizes like 1,000,000.
(4) The Hill estimator is not location invariant. A shift in location does not affect the tail index but may throw the Hill estimator into a tizzy.

The lack of location invariance means the Hill estimator can be surprisingly sensitive to changes in location. Figure 12 illustrates this. The top plots are time series plots of 5000 iid Pareto observations where the true $\alpha = 1$. The two right plots on top have the Pareto observations shifted by 1 and then 2. The bottom two plots are the corresponding Hill plots. Shifting by larger and larger amounts soon produces a completely useless plot.

For point 1, several previous studies advocate choosing k to minimize the asymptotic mean squared error of Hill's estimator [73, 122]. In certain cases, the asymptotic form of this optimal k can be expressed but such a form requires one to know the distribution rather explicitly and it is not clear how much value one gets from an asymptotic formula. There are adaptive methods and bootstrap techniques which try to overcome these problems; it remains to be seen if they will enter the research community's toolbox.

For point 2, there are simple smoothing techniques which always help to overcome the volatility of the plot and plotting on a different scale frequently overcomes the difficulty associated with the stable example. These techniques are discussed in the next paragraph.

5.1.2. *Variant 1. The smooHill plot.* The Hill plot often exhibits extreme volatility which makes finding a stable regime in the plot more guesswork than science and to counteract this, Resnick and Stărică [134] developed a smoothing technique yielding the smooHill plot: Pick an integer u (usually 2 or 3) and define

$$smoo\ \hat{\alpha}_{k,n,u} = \frac{1}{\frac{1}{(u-1)k}\sum_{j=k+1}^{uk} H_{j,n}}.$$

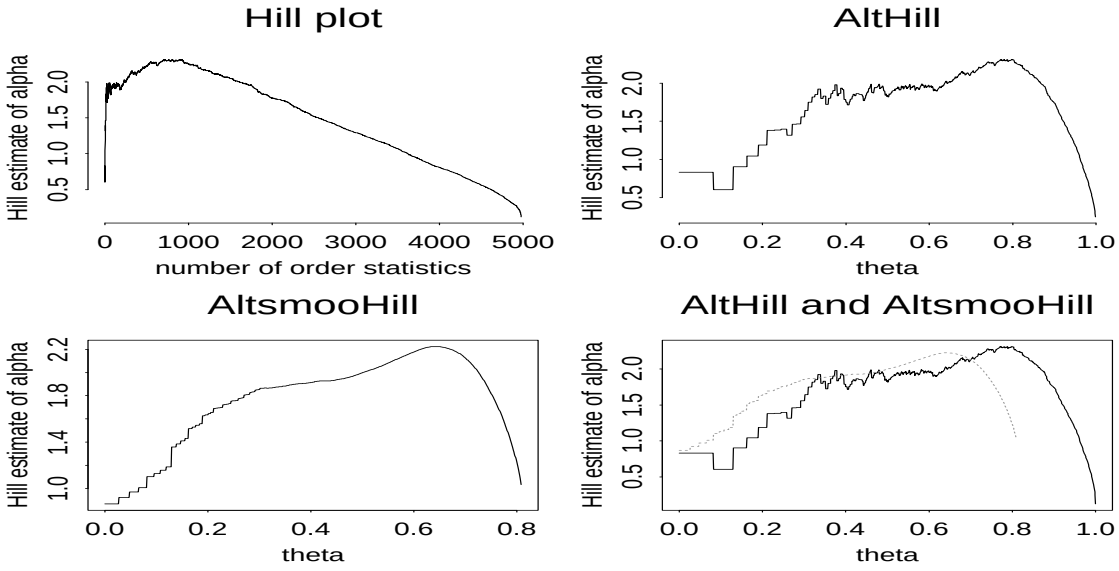

FIGURE 13. Stable, $\alpha = 1.7$

In the iid case when a second order regular variation condition holds, the asymptotic variance of $smooH_{k,n}$ is less than that of the Hill estimator, namely:

$$\frac{1}{\alpha^2}\frac{2}{u}(1 - \frac{\log u}{u}).$$

5.1.3. *Variant 2: Alt plotting; Changing the scale.* As an alternative to the Hill plot, it is sometimes useful to display the information provided by the Hill or smooHill estimation as

$$\{(\theta, H^{-1}_{\lceil n^\theta \rceil, n}), 0 \leq \theta \leq 1\},$$

and similarly for the smooHill plot, where we write $\lceil y \rceil$ for the smallest integer greater or equal to $y \geq 0$. We call such plots the *alternative Hill plot* abbreviated AltHill and the *alternative smoothed Hill plot* abbreviated AltsmooHill. The alternative display is sometimes revealing since the initial order statistics get shown more clearly and cover a bigger portion of the displayed space. Unless the distribution is Pareto, the AltHill plot spends more of the display space in a small neighborhood of α than the conventional Hill plot [52].

Figure 13 compares several Hill plots for 5000 observations from a stable distribution with $\alpha = 1.7$. Plotting on the usual scale is not revealing and the alt plot is more informative.

5.2. **Dynamic and static qq-plots.** Suppose $\{X_1, \ldots, X_n\}$ are iid with distribution F. As we did for the Hill plots, pick k upper order statistics

$$X_{(1)} > X_{(2)} > \cdots X_{(k)},$$

neglect the rest and consider these as the order statistics of exceedances above $X_{(k+1)}$. The distribution of these exceedances should be roughly Pareto if F is heavy tailed. Take the log transform to make this sample of exceedances approximately exponentially distributed and plot empirical quantiles of this sample of size k against the theoretical quantiles of the exponential density. This

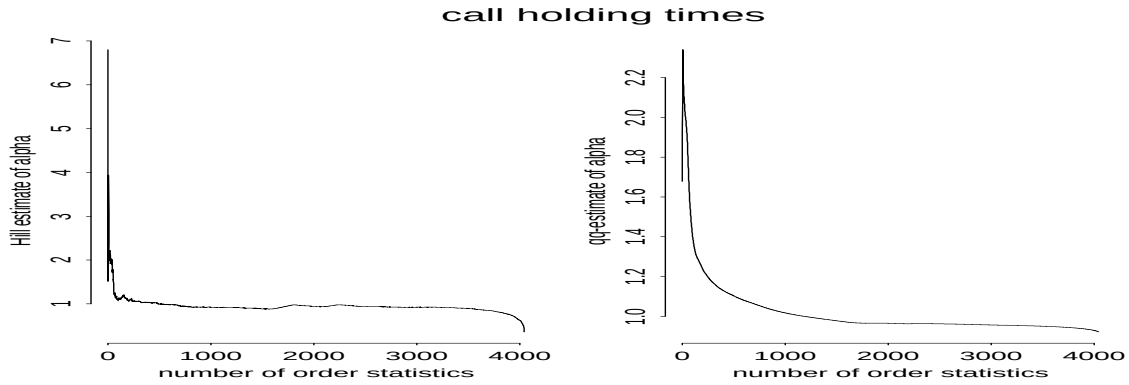

FIGURE 14. Hill and qq-plots for call holding times.

plot

$$\{(-\log(1-\frac{j}{k+1}), \log X_{(k-j+1)}, 1 \le j \le k\} \tag{5.2}$$

should yield approximately a straight line with slope=$1/\alpha$ if F satisfies (5.1). The slope of the least squares line through the points is an estimator called the qq-estimator [94]. Computing the slope, we find that the qq–estimator is given by

$$\widehat{\alpha^{-1}}_{k,n} = \frac{\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))\log(\frac{X_{(i)}}{X_{(k+1)}}) - \frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))H_{k,n}}{\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))^2 - (\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))^2}. \tag{5.3}$$

There are two different plots one can make based on the qq -estimator. There is the dynamic qq -plot obtained from plotting $\{(k, 1/\widehat{\alpha^{-1}}_{k,n}, 1 \le k \le n\}$, which is similar to the Hill plot. Another plot, the static qq-plot, is obtained by choosing and fixing k, plotting the points in (5.2) and putting the least squares line through the points while computing the slope as the estimate of α^{-1}.

The qq-estimator is consistent if $k \to \infty$ and $k/n \to 0$. Under an additional second order condition on $F(x)$ and further restriction on $k(n)$, it is asymptotically normal with asymptotic mean $1/\alpha$ and asymptotic variance $2/\alpha^2$. (See [94] and also [10].) This is larger than the asymptotic variance of the Hill estimator but bias and volatility of the plot seem to be more of an issue than asymptotic variance. The volatility of the qq-plot always seems to be less than that of the Hill estimator. As with the Hill estimator, sensitivity to choice of k is an important issue. It is rare that another technique gives a clearer answer to the question *Is the data heavy tailed?*

Figure 14 compares the Hill plot with the dynamic qq-plot for the call holding data and Figure 15 does the same thing for the packet inter-arrival data. Figure 16 gives two static qq–plots for the call holding data, one using $k = 3500$ and the other using $k = 1500$ yielding estimates of α of 0.95 and 0.98 respectively. For this data set, the estimators are unusually insensitive to the choice of k; the Hill plots and the dynamic qq-plots are quite stable and the static qq-plots do not change much as k varies. Figure 17 gives two static qq-plots for the packet inter-arrival data. The data set is only 783 in length and now there is some sensitivity to k.

The qq-plot is often a valuable technique to gauge if a heavy tailed model is appropriate. One can plot

$$\{(-\log(1-\frac{j}{n+1}), \log X_{(n-j+1)}), 1 \le j \le n\}$$

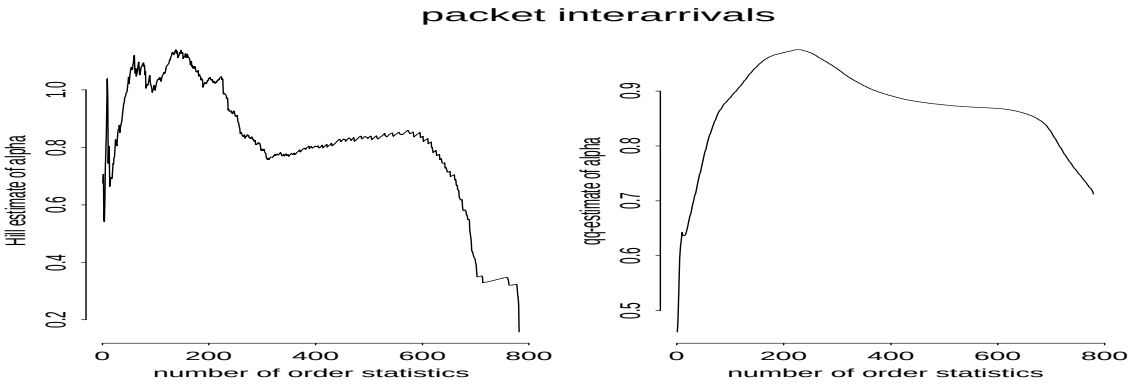

FIGURE 15. Hill and qq-plots for packet inter-arrivals.

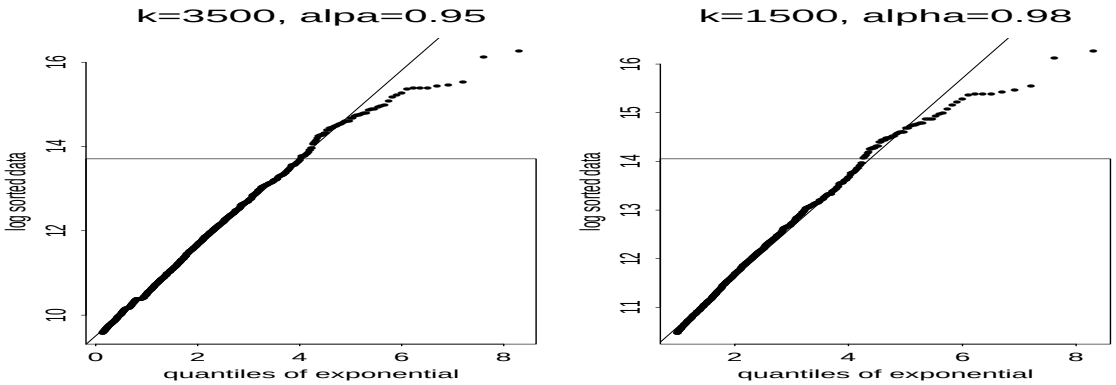

FIGURE 16. Static qq-plots for call holding times.

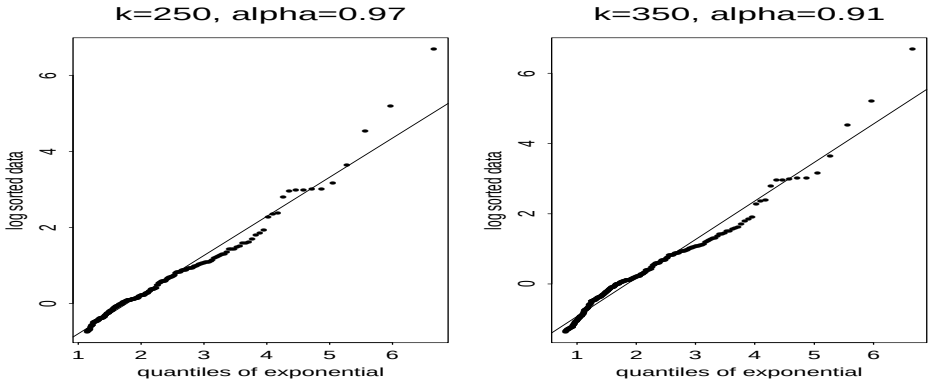

FIGURE 17. Static qq-plots for packet inter-arrivals.

using all the data. If the plot from some point on looks linear, this is evidence that a heavy tailed model is applicable. In many examples, there are a number of points at the end of the plot that do not seem to follow the line. The number of such points is of the order of 10.

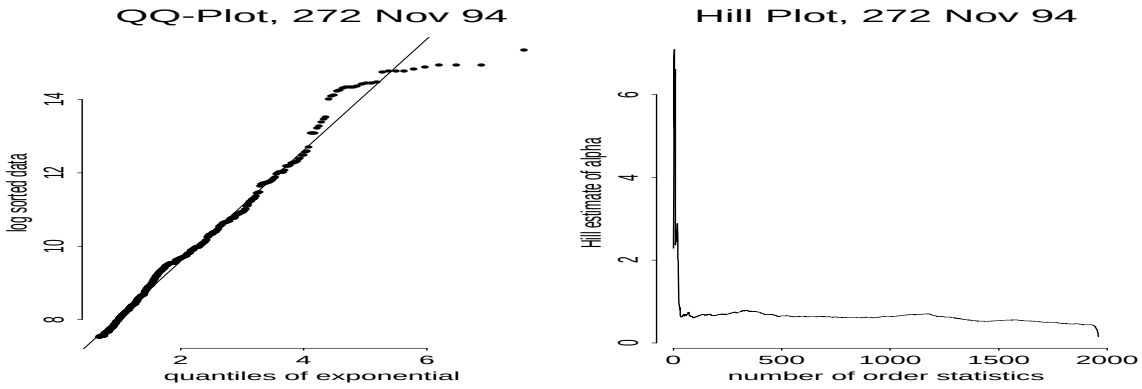

FIGURE 18. Hill and static qq-plots for BU file size data.

An interesting example of heavy tailed data is the Boston University study of web use from 1994 [36, 32, 30, 33, 31] which, among other things, kept track of sizes of files downloaded in a WWW session. For the month of November 1994 in the lab in Room 272, Figure 18 shows both a classical Hill plot and a static qq-plot. The mean of this data appears infinite since the estimate of $\alpha \approx 0.66$.

One example where the Hill plot requires careful inspection but the qq-plot is immediately informative is the following data set called *zjz*, which represents file sizes requested from a server at the University of North Carolina. The qq-plots of Figure 20 are informative, allowing an estimate of $\alpha = 1.42$ for $k = 18,000$. The size of the data set is 131,943. The Hill plots in Figure 19 are initially less clear since there is no clear region where the plot is stable. After the qq-estimate is obtained, the Hill-plot serves to confirm the qq-estimate.

5.3. **The Dekkers, Einmahl, De Haan moment estimator.** Ignoring location and scale, we know that the extreme value distributions [139, 42, 98, 24] can be parameterized as a one parameter family

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-\gamma^{-1}}\}, \quad \gamma \in \mathbb{R},\ 1+\gamma x > 0. \tag{5.4}$$

The Dekkers, Einmahl, De Haan moment estimator $\hat{\gamma}$ [49, 50, 51] is designed to estimate γ for a random sample in the domain of attraction of G_γ. When $\gamma > 0$, this is the same thing as estimating

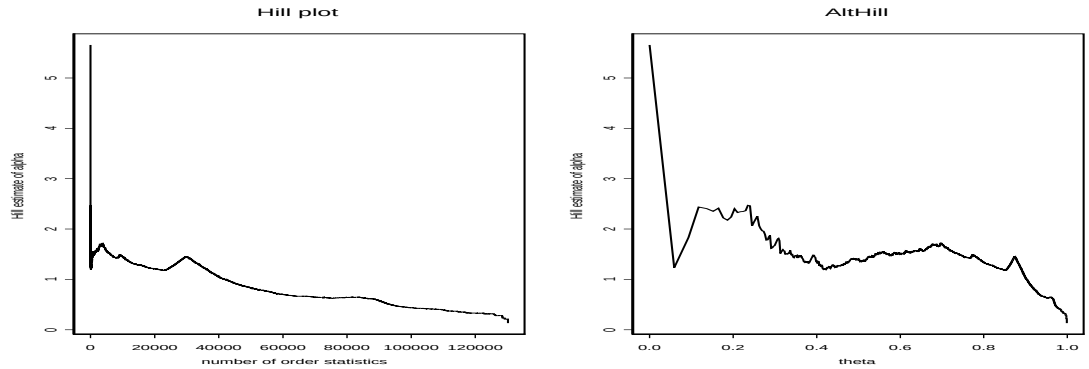

FIGURE 19. Hill and altHill plots for the zjz file size data.

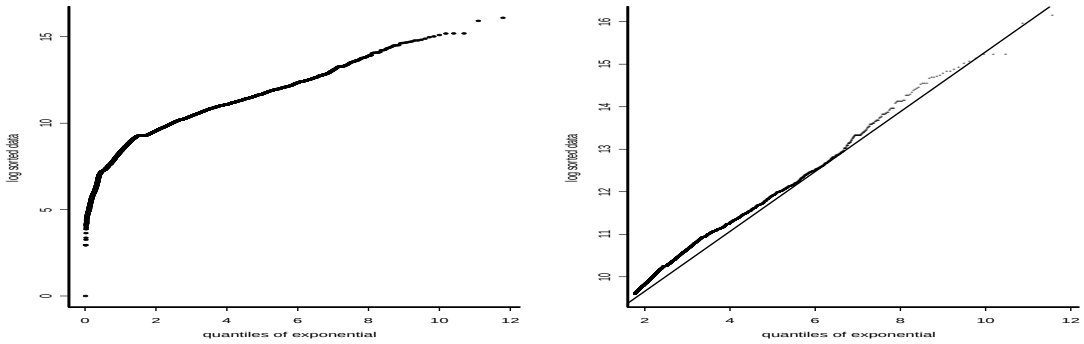

FIGURE 20. qq-plots for zjz.

$\gamma = 1/\alpha$. Recall that the exponential, normal, gamma densities and many others are in the $D(G_0)$, the domain of attraction of the Gumbel distribution. This provides another method of deciding when a distribution is heavy tailed or not. If $\hat{\gamma}$ is negative or very close to zero, there is considerable doubt that heavy tailed analysis should be applied.

The moment estimator is defined as follows: As usual, let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ be the order statistics from a random sample of size n. Define for $r = 1, 2$

$$H_{k,n}^{(r)} = \frac{1}{k} \sum_{i=1}^{k} \left(\log \frac{X_{(i)}}{X_{(k+1)}} \right)^r,$$

so that $H_{k,n}^{(1)}$ is the Hill estimator. Define

$$\hat{\gamma}_n = H_{k,n}^{(1)} + 1 - \frac{1/2}{1 - (H_{k,n}^{(1)})^2 / H_{k,n}^{(2)}}. \tag{5.5}$$

Then assuming $F \in D(G_\gamma)$, we have consistency

$$\hat{\gamma}_n \xrightarrow{P} \gamma,$$

as $n \to \infty$ and $k/n \to 0$. Furthermore under an additional condition and a further restriction on k,

$$\sqrt{k}(\hat{\gamma} - \gamma) \Rightarrow N,$$

where N is a normal random variable with 0 mean and variance

$$\sigma(\gamma) = \begin{cases} 1 + \gamma^2, & \text{if } \gamma \geq 0, \\ (1-\gamma)^2(1-2\gamma)\left(4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right), & \text{if } \gamma < 0. \end{cases}$$

The asymptotic variance of the moment estimator exceeds that of the Hill estimator when $\gamma > 0$, so from the point of view of asymptotic variance, there is no reason to prefer it. However, the moment estimator discerns a light tail more effectively than the Hill estimator, and thus it is often useful to apply the moment estimator to see if $\gamma \leq 0$, which would rule out heavy tail analysis.

Figure 21 compares the effectiveness of the moment estimator for discerning a light tail ($\gamma = 0$) with that of of the Hill estimator $\alpha = \infty$. The Hill estimator does not seem very reliable for this purpose. Both estimators are applied to the same random sample of size 1000 taken from an exponential distribution with unit mean.

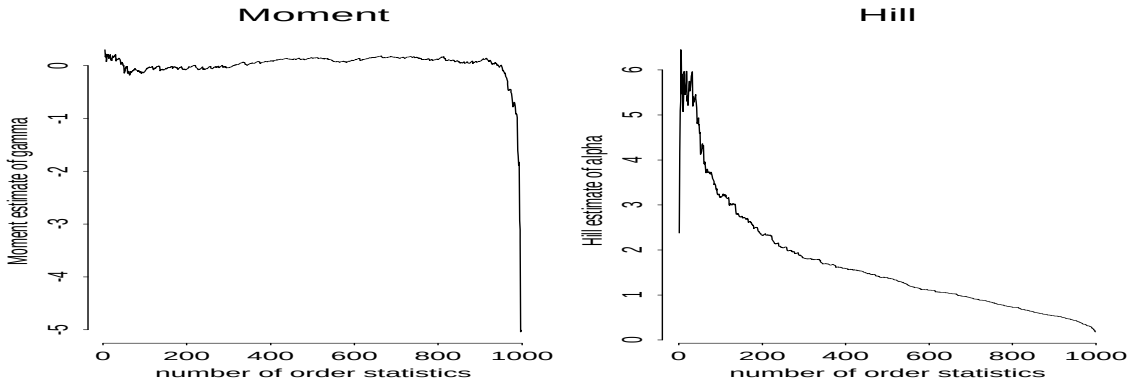

FIGURE 21. Unit exponential data.

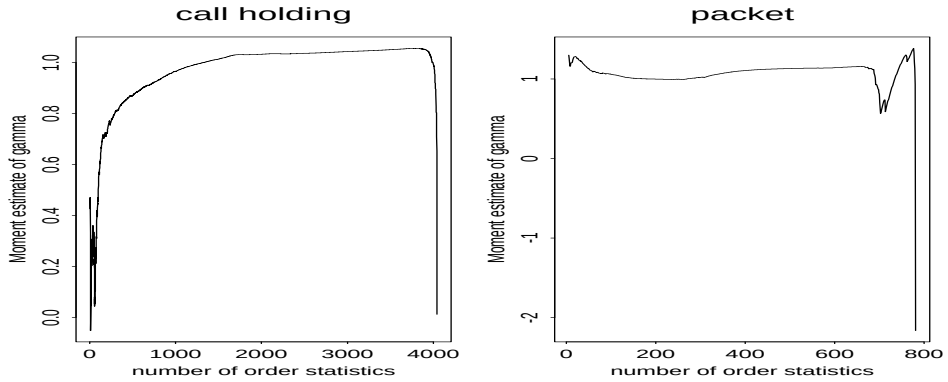

FIGURE 22. Moment estimator applied to call holding and packet data.

Figure 22 shows the moment estimator applied to the call holding data on the left and the packet inter-arrival data on the right. Keep in mind when comparing these graphs with previous graphs that $\gamma = 1/\alpha$.

5.4. **Peaks over threshold method.** A standard technique from extreme value theory is the peaks over threshold method [56, 124]. Assuming that the underlying distribution F is in a domain of attraction of an extreme value distribution G_γ given by (5.4), the excesses over a high threshold, are approximately distributed as a generalized Pareto distribution. In most cases, the parameters of the generalized Pareto can be estimated by maximum likelihood estimation. This certainly works in the heavy tailed case where F satisfies (5.1).

Example 5. The theory presented in the previous Section 3 concluded that if the model is correct, then the distribution of cumulative traffic aggregated over users and time is approximable either by FBM or by a stable Lévy motion. In [72] several data sets were examined to see if there were examples where cumulative input looked heavy tailed. The cumulative traffic was calculated over disjoint time blocks to get a sample, and then the POT method was used to estimate the extreme value index γ. In order to conclude the stable Lévy motion approximation is possibly correct, the POT estimation would need to find a positive value for γ which was greater than $1/2$ since this

would correspond to $0 < \alpha < 2$. Such a value was not found, as the summary from Table 1 below, taken from [72], demonstrates.

Data set	γ $(=1/\alpha)$
simM/G/∞	$-.13 \pm .03$
BUburst 10s	$-.36 \pm .13$
BUburst 1s	$.17 \pm .03$
UCB 10s	$.05 \pm .18$
UCB syn 10s	$-.60 \pm .14$
Munich lo TX	$.09 \pm .04$
Munich lo RX	$-.03 \pm .05$
Munich hi .1s	$.17 \pm .12$
Munich hi .01s	$.10 \pm .03$
Ericsson	$-.47 \pm .09$
Eri syn 1s	$-.31 \pm .12$

TABLE 1. Point estimates of the extreme value parameter for cumulative traffic $\pm$ "standard deviation".

6. Does the model fit the data? Long range dependence, self-similarity, Hurst phenomenon.

6.1. **Long range dependence.** A stationary L_2 sequence $\{\xi_n, n \geq 1\}$ possesses long range dependence (lrd) if

$$\mathrm{Cov}(\xi_n, \xi_{n+h}) \sim h^{-\beta} L(h), \quad h \to \infty \tag{6.1}$$

for $0 < \beta < 1$ and $L(\cdot)$ slowly varying [11]. Set $\gamma(h) = \mathrm{Cov}(\xi_n, \xi_{n+h})$ and $\rho(h) = \gamma(h)/\gamma(0)$ for the covariance and correlation functions of the stationary process $\{\xi_n\}$. There is no universal agreement about terminology, and sometimes long range dependence is taken to mean that covariances are not summable: $\sum_h |\gamma(h)| = \infty$, whereas short range dependence means $\sum_h |\gamma(h)| < \infty$.

Long range dependence, like the property of heavy tails, has acquired a mystical, almost religious, significance and generated controversy. Researchers argue over whether it exists, whether it matters if it exists or not, or whether analysts have been fooled into mistaking some other phenomena like shifting levels, undetected trend [95] or non-stationarity for long range dependence. Discussions about this have been going on since (at least) the mid 70's in hydrology ([16, 148, 149, 150, 15, 19, 13]) in finance ([111, 112]) and in data network modeling ([53, 77, 21, 22, 65, 99, 119, 154]). Think of it as one more modeling decision that needs to be made. Since long range dependence is an asymptotic property, models that possess long range dependence presumably have different asymptotic properties from those models where long range dependence is absent, although even this is sometimes disputed.

The definitions of short and long range dependence are designed so that the differences in properties show up clearest under block averaging [27]. Define the block average operator

$$A_m : \mathbb{R}^\infty \mapsto \mathbb{R}^\infty$$

by

$$A_m(x_1, x_2, \dots) = \left(\frac{1}{m} \sum_{i=(k-1)m+1}^{km} x_i, k \geq 1 \right),$$

so that A_m maps a sequence into the averages over successive blocks of length m. Suppose

$$\{\xi_n^{(m)}\} = A_m(\{\xi_n\})$$

is the block averaged $\{\xi_n\}$ sequence (which is still L_2 weakly stationary), and

$$\gamma^{(m)}(h) = \mathrm{Cov}(\xi_n^{(m)}, \xi_{n+h}^{(m)})$$

is the covariance function and $\rho^{(m)}(h)$ is the corresponding correlation function. Under short range dependence

$$\rho^{(m)}(h) \to 0, \quad |h| \geq 1, \ m \to \infty,$$

and the second order structure of $\{\sqrt{m}\xi_n^{(m)}, n \geq 1\}$ converges, as $m \to \infty$, to that of white noise. However, under long range dependence defined in (6.1),

$$\rho^{(m)}(h) \to \tilde{\rho}(h) = \frac{1}{2}\left[(h+1)^{2-\beta} - 2h^{2-\beta} + (h-1)^{2-\beta}\right], \quad m \to \infty, \tag{6.2}$$

which is the correlation function of *fractional Gaussian noise* (fgn). The correlation function of fgn exhibits slow decay characteristic of long range dependence since

$$\tilde{\rho}(h) \sim H(2H-1)h^{2H-2} = (1 - \frac{\beta}{2})(1-\beta)h^{-\beta}, \quad h \to \infty, \tag{6.3}$$

where $H = 1 - \frac{\beta}{2}$ is called the *Hurst* parameter (but be wary of this terminology when we get to the $R/\mathcal{S}$ statistic).

6.2. **Self-similarity in discrete time; connections with long range dependence.** There is a tendency to muddle long range dependence and the related notion of self-similarity, so some care in distinguishing between the concepts is worthwhile.

Let $\boldsymbol{X} = (X_1, X_2, \dots)$ be weakly stationary. Call $\boldsymbol{X}$ *distributionally self-similar* if there exists $H > 0$ such that for every $m = 1, 2, \dots$

$$m^{1-H} A_m \boldsymbol{X} \stackrel{d}{=} \boldsymbol{X} \tag{6.4}$$

as random elements of $\mathbb{R}^\infty$. This is equivalent to

$$m A_m \boldsymbol{X} \stackrel{d}{=} m^H \boldsymbol{X}$$

where $mA_m\boldsymbol{X}$ is the block aggregated process.

Call $\boldsymbol{X}$ *second order self-similar* if the second order structure of $m^{1-H}A_m\boldsymbol{X}$ and $\boldsymbol{X}$ are the same for every m. This means

$$\begin{aligned} &(i) \quad \rho((A_m\boldsymbol{X})_k, (A_m\boldsymbol{X})_{k+h}) = \rho(X_k, X_{k+h}) \\ &(ii) \quad \mathrm{Var}((m^{1-H}A_m\boldsymbol{X})_1) = \mathrm{Var}(X_1). \end{aligned}$$

It turns out that (ii) implies (i) and provided $1/2 < H < 1$, a second order self-similar process has the same correlation structure as fgn, and any Gaussian distributionally self-similar process with $1/2 < H < 1$ is fgn.

Asymptotic self-similarity in discrete time: Call the weakly stationary process $\boldsymbol{X} = (X_1, X_2, \ldots)$ *asymptotically second order self-similar* if as $m \to \infty$

$$\rho^{(m)}(h) = \rho((A_m\boldsymbol{X})_k, (A_m\boldsymbol{X})_{k+h}) \to \frac{1}{2}\left[(h+1)^{2H} - 2h^{2H} - (h-1)^{2H}\right], \quad h = 0, \pm 1, \ldots. \tag{6.5}$$

Call H the self-similarity parameter. Such processes have a correlation structure such that after aggregation over blocks of length m, the resulting process has only approximately the correlation structure of fgn. Any stationary process with long range dependence as defined in (6.1) is asymptotically second order self-similar. This includes fractionally differenced ARIMA models.

6.3. **Self-similarity in continuous time.** A continuous time stochastic process $(X(t)), t \geq 0)$ is self-similar with self-similarity parameter $H > 0$ (abbreviated H-ss) if for every $a > 0$

$$X(a\cdot) \stackrel{d}{=} a^H X(\cdot).$$

This means that the finite dimensional distributions of the process $(X(at), t \geq 0)$ with its dilated time are the same as those of the rescaled process $a^H X(\cdot)$ where the state space is stretched.

Brownian motion is H-ss with $H = \frac{1}{2}$, which immediately gives an example of a non-stationary process which is H-ss. In fact a strictly stationary process cannot be H-ss unless it is identically zero. However there is the following connection between stationarity and self-similarity [96, 97]:

(a) If $\{X(t), t > 0\}$ is H-ss, then $\{Y(t), t \in \mathbb{R}\}$ defined by

$$Y(t) = e^{-tH} X(e^t)$$

is strictly stationary.

(b) If $\{Y(t), t \in \mathbb{R}\}$ is strictly stationary, then $\{X(t), t > 0\}$ defined by

$$X(t) = t^H Y(\log t),$$

is H-ss.

The transformation applied to Brownian motion to get a stationary process yields the Ornstein-Uhlenbeck process.

Stationary increments: A process $\{Y(t), t \geq 0\}$ has *stationary increments* if for any $s > 0$

$$\{Y(s+t) - Y(s), t \geq 0\} \stackrel{d}{=} \{Y(t) - Y(0), t \geq 0\}.$$

A process Y which is both H-ss and has stationary increments will be abbreviated H-sssi.

Suppose $\{Y(t), t \geq 0\}$ is H-sssi and suppose $E(Y(t)^2) = t^{2H} E(Y(1)^2) < \infty$ and $EY(1) = 0$. Then

(a) If $H \in (0, 1]$ ([151])

$$\mathrm{Cov}\big(Y(t_1), Y(t_2)\big) = \frac{EY(1)^2}{2}\Big(t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H}\Big). \tag{6.6}$$

(b) Define for $j \geq 0$

$$X_j = Y(j+1) - Y(j).$$

Then $\{X_j, j \geq 0\}$ is weakly stationary and has the correlation structure of fgn.

Extremal processes [139] are good examples of H-ss processes which do not have stationary increments. The range of H can be $(0, \infty)$. There are many extremal processes which are not L_2 or even L_1. Self-similarity is a distributional property, not a moment property.

6.3.1. *Fractional Brownian motion.* For $0 < H \leq 1$, a Gaussian process with correlation function

$$\frac{\sigma^2}{2}\left(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}\right) \tag{6.7}$$

for $\sigma > 0$ will be called fractional Brownian motion (fbm) and denoted $(B_H(t), t \geq 0)$. This correlation function corresponds to the covariance in (6.6).

This helps explain the temptation to model traffic with fbm. Increments of fbm give fgn, a stationary sequence with long range dependence provided $H \in (0.5, 1)$.

6.3.2. *Asymptotic self-similarity in continuous time.* This is a notion developed by Lamperti [96, 97]. See also [55, 161, 160, 115, 114] for extensions.

Suppose $\{X(t), t \geq 0\}$ is a stochastic process such that $X(0) = 0$ and for each $t > 0$

$$X(t) \text{ is a non-degenerate random variable.} \tag{6.8}$$

Assume $\{\xi_n, n \geq 0\}$ is a sequence of random variables and $b(t), t \geq 0$, is a measurable, non-negative function such that as $s \to \infty$, $b(s) \to \infty$ and

$$\frac{\xi_{[st]}}{b(s)} \Rightarrow X(t) \tag{6.9}$$

in the sense of convergence of finite dimensional distributions. Then for some $H > 0$, $b(\cdot)$ is regularly varying with index H and $X(\cdot)$ is H-ss. Furthermore $X(\cdot)$ is H-sssi provided for $\{\xi_n, n \geq 0\}$ is stationary. This result, known as Lamperti's theorem, shows that limits of partial sum processes of stationary random variables will be self-similar.

Here is a summary and some additional facts: Suppose $\{Y(t), t \geq 0\}$ is H-sssi:

(a) If $EY(1)^2 < \infty$, then $0 < H \leq 1$.
(b) If $H = 1$, then $Y(t) = tY(1)$ almost surely.
(c) If $X_j = Y(j+1) - Y(j)$, for $j \geq 1$, then $\{X_j\}$ has the covariances of fgn and hence has long range dependence if $\frac{1}{2} < H < 1$.

This leads to methods of constructing long range dependent processes.

(i) Fractionally differenced ARMA processes (see [18]).
(ii) Increments of H-sssi processes with $\frac{1}{2} < H < 1$. Processes with the H-sssi property can be constructed by various means:
  (a) Stochastic integration (see [151]),
  (b) Lamperti's theorem.

6.4. **Should we be happy?** There are obvious problems with all this when you think it over:

(i) Long range dependence is defined by covariances for weakly stationary variables.
(ii) What happens if the variables have infinite variance? What if the process is non-Gaussian so that covariances are relatively uninformative?
(iii) There exist examples of a long range dependent sequence $\{X_n\}$ and a function $h : \mathbb{R} \mapsto \mathbb{R}$ such that $\{h(X_n)\}$ does not have long range dependence and vice versa. This is disturbing if you think of long range dependence as an intrinsic property describing dependence. The definition in terms of product moments falls short of capturing this.

6.5. **Statistical techniques and exploratory methods.** We mention some (not all) traditional techniques to identify when a model with long range dependence is appropriate as well as some recent wavelet methods. See [12, 5, 3, 1, 158, 159].

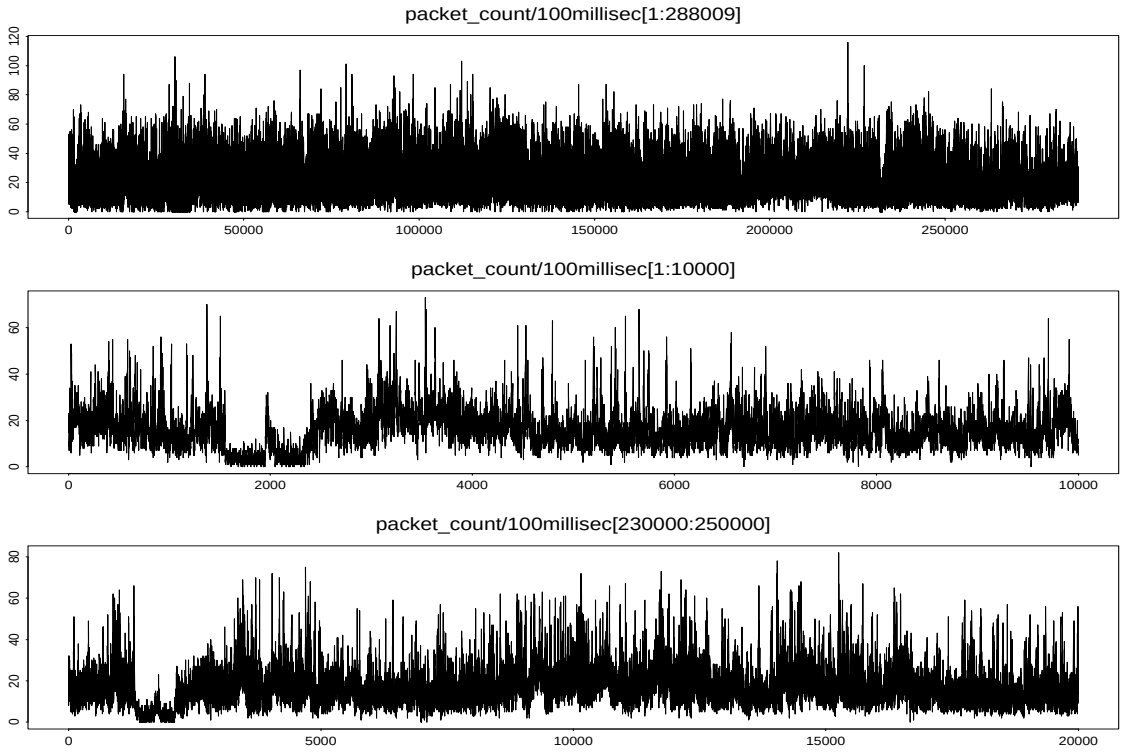

FIGURE 23. Time series plots for Company X data. Top: full data set. Middle: first 10,000 data. Bottom: last 20,000 observations.

6.5.1. *The sample acf.* The most common, ubiquitous Q&D method, assuming you are convinced the data comes from a stationary process, is to graph the sample acf. The plot should not decline rapidly. Classical time series data that one encounters in ARMA (Box-Jenkins) modeling exercises has a sample acf which is essentially zero after a few lags and acf plots of financial or teletraffic data are often in stark contrast.

Example 6 (Company X)**.** This trace is packet counts per 100 milliseconds=1/10 second for Financial Company X's wide area network link including USA–UK traffic. It consists of 288,009 observations corresponding to 8 hours of collection from 9 a.m.–5 p.m. Figure 23 shows time series plots. The top plot of the whole data set does not raise any alarms about lack of stationarity but this is partly due to the muddy plotting resulting from the abundance of data. The middle plot shows the first 10,000 and the bottom plot displays the last 20,000. With reduced data size, the last two plots raise some question whether stationarity is appropriate but this has not been pursued.

Figure 24 shows the acf plot for 2000 lags. There is little hurry for the plot to approach zero. (Don't try to model this with ARMA.)

Example 4 (continued). Recall that the Ericsson data did not look very stationary. In an attempt to counteract the non-stationarity, a subset of the last 28,122 observations was considered. The time series and acf plots are given in Figure 25. The acf plot is clearly not settling down as it would with an ARMA model, but is also not declining like a power law.

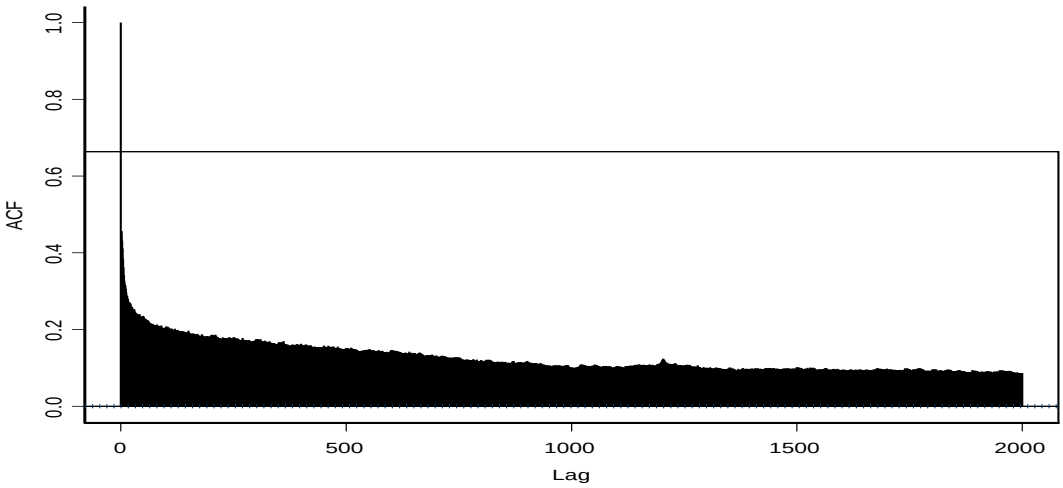

FIGURE 24. Sample autocorrelation plot for Company X data for 2000 lags.

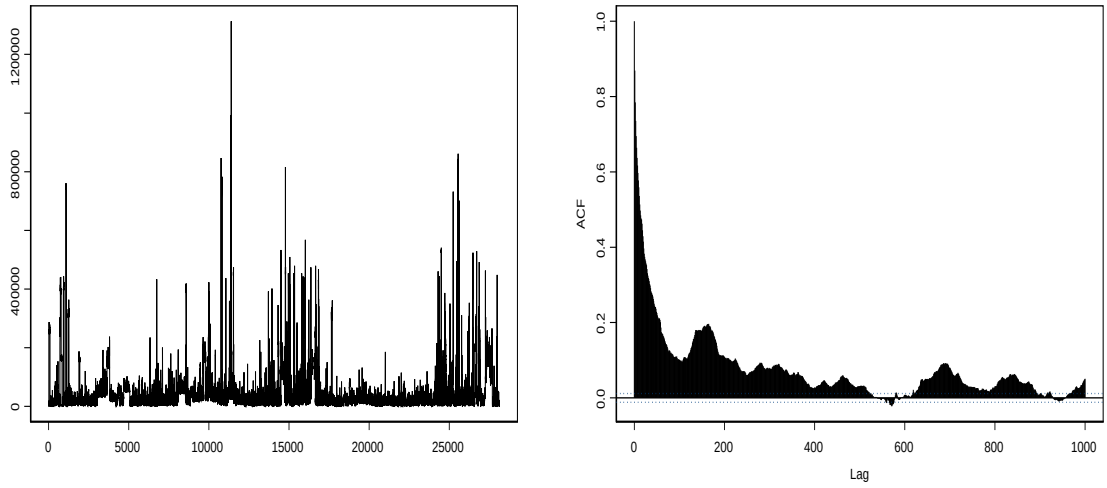

FIGURE 25. Time series plot (left) of the 28,122 last Company X data and acf plot (right) of these data.

6.5.2. *The variance-time plot.* Suppose $\boldsymbol{X} = \{X_n\}$ is a weakly stationary sequence with long range dependence whose acf satisfies (6.1). If we block average over a block of length m, the variance v_m of $(X_1 + \cdots + X_m)/m$ satisfies

$$v_m \sim \text{(const)}\ m^{-\beta} L(m). \tag{6.10}$$

Note for short range dependence

$$v_m \sim \text{ (const) } \frac{1}{m}. \tag{6.11}$$

We seek a graphical technique which capitalizes on the differences between (6.10) and (6.11) and which will identify $\beta \in (0,1)$, or at least make sure that the usual square root of n central limit behavior is not present.

Here is the method. Start from measurements $\{x_j, j = 1, \ldots n\}$.

(i) Compute the block averages $x_j^{(m)}, j = 1, \ldots, [n/m]$ over blocks of length m.

(ii) Use $(x_j^{(m)}, j = 1, \ldots, [n/m])$ to estimate v_m by computing the sample variance

$$\hat{v}_m = \frac{1}{[n/m]} \sum_{j=1}^{[n/m]} \left(x_j^{(m)} - \bar{x}^{(m)} \right)^2,$$

where

$$\bar{x}^{(m)} = \sum_{j=1}^{[n/m]} x_j^{(m)} / [n/m]$$

is the sample average.

(iii) Plot $\{(\log m, \log \hat{v}_m), m = 1, 2, \ldots\}$. Since

$$\log v_m \sim \text{ (const) } - \beta \log m,$$

the plot of $\{(\log m, \log \hat{v}_m), m \geq$ some threshold value $m_0\}$ should be roughly linear.

(iv) Put the least squares line through

$$\{(\log m, \log \hat{v}_m), m \geq m_0\}.$$

The threshold m_0 can be determined by inspection frequently (or at least hopefully). The slope of this line is an estimate of $-\beta$.

(v) If short-range behavior is present, the plot of $((\log m, \log \hat{v}_m), m \geq m_0)$ would be linear with slope approximately -1, so we also plot the line of that slope for comparison.

Example 6 (continued). Figure 26 gives the variance-time plot for the initial 20,000 measurements of the Company X data and yields an estimate of $\hat{\beta} = 0.196$.

6.5.3. *The Hurst phenomenon and the R/S statistic.* What attributes does one expect from data which is generated by a long range dependent process? Because of slowly decreasing covariances, such data will exhibit long periods where the level is high relative to the mean and long periods where it is low. (This may, of course, make long range dependence difficult to distinguish statistically from a shifting level model.)

The origin of the R/S statistic is in hydrology [80, 81] and comes from the following thought experiment (which to the unitiated may seem a bit like artificial mumbo jumbo): suppose inputs to a reservoir are $X_1, \ldots, X_T$ and think of X_i as the input at the i^{th} time interval. The level at time 0 is x and we are allowed to withdraw water at constant rate $r > 0$. In time periods $1, \ldots, T$ we want the reservoir never to be empty, never to overflow and we want to allow maximum possible usage of the water, which we interpret as meaning that the final water level at T should be x. Let R be the size of the reservoir. Find R, x, r.

Let $S_i = \sum_{j=1}^{i} X_j$ be the cumulative input in the first i periods and set $S_0 = 0$. The following are the constraints.

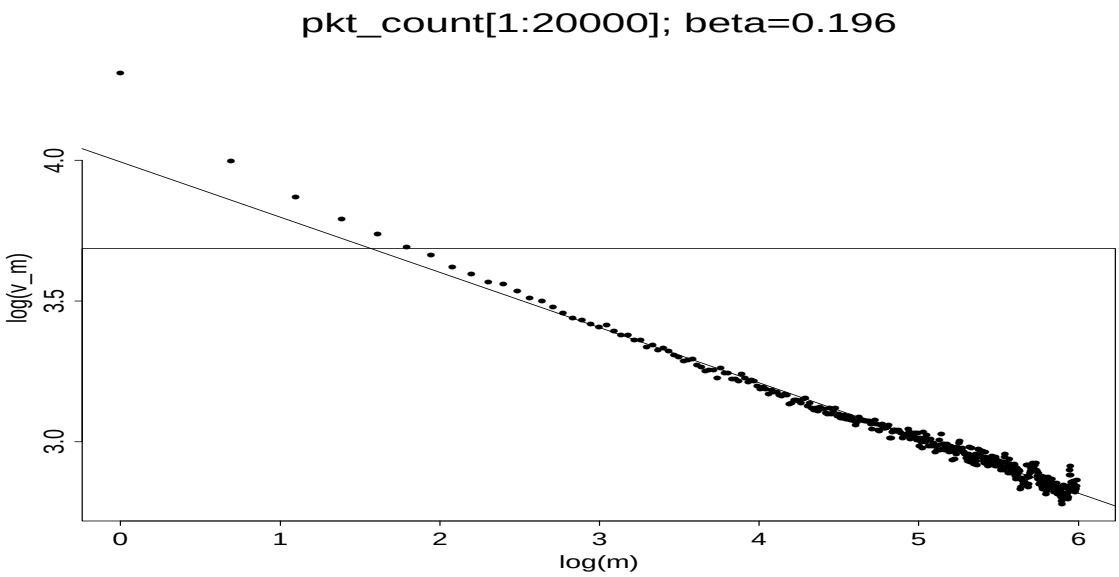

FIGURE 26. Variance-time plot of the initial 20,000 measurements in the Company X data.

(a) The initial amount equals the final amount in the reservoir:

$$x = x + S_T - Tr.$$

This immediately implies $r = \frac{S_T}{T} = \bar{X}_T$.

b) The never dry condition:

$$x + S_i - ir \geq 0, \quad i = 0, \ldots, T$$

or

$$x + \wedge_{i=0}^{T}(S_i - ir) \geq 0.$$

So we can set

$$x = -\wedge_{i=0}^{T}(S_i - \frac{i}{T}S_T).$$

(c) The condition that capacity R is never exceeded:

$$x + S_i - ir \leq R, \quad i = 0, \ldots, T$$

or

$$x + \vee_{i=0}^{T}(S_i - ir) \leq R$$

or

$$\vee_{i=0}^{T}(S_i - \frac{i}{T}S_T) - \wedge_{i=0}^{T}(S_i - \frac{i}{T}S_T) \leq R.$$

We set

$$R = R_T = \vee_{i=0}^{T}(S_i - \frac{i}{T}S_T) - \wedge_{i=0}^{T}(S_i - \frac{i}{T}S_T),$$

which is called the *adjusted range*.

To make a statistic which is scale invariant, we divide by

$$\mathcal{S} = \mathcal{S}_T = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(X_i - \bar{X}_T)^2}$$

and then $R/\mathcal{S}$ is scale and location invariant. $R/\mathcal{S}$ is called the *rescaled adjusted range*.

Hurst [80, 81] found empirically that for Nile river flows, as a function of T, $R/\mathcal{S}$ grew roughly like $(const)T^H$ where $H = 0.75$. However in 1951, W. Feller [63] proved that if $\{X_j\}$ are iid with finite variance, then

$$\frac{R_T/\mathcal{S}_T}{\sqrt{T}} \Rightarrow L, \quad T \to \infty$$

where L is a non-degenerate random variable, and thus, if Hurst's data had come from an iid model, the growth rate of $R/\mathcal{S}$ should be $T^{1/2}$. This discrepancy is known as the *Hurst phenomenon*.

The modern way to understand Feller's result is by means of the theory of weak convergence.

Theorem 3. *Suppose $\{X_n\}$ are random variables such that for some constants $\mu_n, b_n > 0$, there exists a process $X(\cdot)$ in $D[0,1]$ such that*

$$X_n(t) := \frac{\sum_{i=1}^{[nt]}(X_i - \mu_n)}{b_n} \Rightarrow X(t)$$

in $D[0,1]$. Define

$$W_j = \sum_{i=1}^{j} X_i - \frac{j}{n}\sum_{i=1}^{n} X_i = \sum_{i=1}^{j}(X_i - \bar{X}_n)$$

and

$$R_n = \bigvee_{j=0}^{n} W_j - \bigwedge_{j=0}^{n} W_j.$$

Assuming $\vee_{t=0}^{1}$ and $\wedge_{t=0}^{1}$ operators are almost surely continuous, we have

$$\frac{R_n}{b_n} \Rightarrow \bigvee_{0\le t\le 1}(X(t) - tX(1)) - \bigwedge_{0\le t\le 1}(X(t) - tX(1)).$$

Additionally, if for $\beta_n > 0$

$$\left(X_n(\cdot), \frac{\mathcal{S}_n}{\beta_n}\right) \Rightarrow (X(\cdot), \mathcal{S}),$$

then

$$\frac{R_n/\mathcal{S}_n}{b_n/\beta_n} \Rightarrow \mathcal{S}^{-1}\Big(\bigvee_{0\le t\le 1}(X(t) - tX(1)) - \bigwedge_{0\le t\le 1}(X(t) - tX(1))\Big).$$

For example, suppose in $D[0,1]$

$$X_n(t) \Rightarrow B(t),$$

a Brownian motion. Then $B^{(0)}(t) = B(t) - tB(1)$ is a Brownian bridge. When short range dependence is present we frequently have $b_n = \sqrt{n}$ and $\mathcal{S}_n^2 \to \sigma^2$, some constant; this would certainly be true in the iid, finite variance case. Then we would have

$$\frac{R_n/\mathcal{S}_n}{\sigma\sqrt{n}} \Longrightarrow \bigvee_{t=0}^{1} B^{(0)}(t) - \bigwedge_{t=0}^{1} B^{(0)}(t).$$

If we let $R_n^{(0)} = \frac{R_n/\mathcal{S}_n}{\sigma\sqrt{n}}$, so that $R_n/\mathcal{S}_n = \sigma R_n^{(0)}\sqrt{n}$, we would have

$$\log R_n/\mathcal{S}_n = \frac{1}{2}\log n + \log \sigma R_n^{(0)}.$$

Plotting $\{(\log n, \log R_n/\mathcal{S}_n), n \geq$ some threshold$\}$ should approximately give a line of slope 1/2, and plotting

$$\left(\frac{\log R_n/\mathcal{S}_n}{\log n}, n \geq 1\right)$$

should give a graph which settles in a neighborhood of 1/2.

However, this is frequently not the case.

6.5.4. *Trying to explain the Hurst phenomenon.* Suppose the following conditions hold:

(1) The $R/\mathcal{S}$ statistic has a limit distribution:

$$\frac{R_n/\mathcal{S}_n}{b_n} \Rightarrow \chi, \tag{6.12}$$

where χ is a non-degenerate random variable, and

(2) b_n is a scaling function satisfying

$$b_n \sim n^J L(n) \in RV_J; \tag{6.13}$$

that is $t^J L(t)$ is regularly varying with index J. The regular variation index J is called the *Hurst exponent* (as opposed to the Hurst parameter discussed just after (6.3)). One hopes J always equals the self-similarity parameter H, since the R/S procedure estimates J, but, alas, this is not always the case.

The $R/\mathcal{S}$ estimation procedure: To estimate J we follow the outline in the Brownian limit example above. Define

$$\chi_n = \frac{R_n/\mathcal{S}_n}{b_n},$$

and then $\chi_n \Rightarrow \chi$ and

$$\log \chi_n = \log(R_n/\mathcal{S}_n) - J\log n - \log L(n) \Rightarrow \log \chi,$$

so

$$\frac{\log(R_n/\mathcal{S}_n)}{\log n} - J = \frac{\log \chi_n + \log L(n)}{\log n} = o_p(1).$$

The strategy, then, is to plot

$$\left\{\left(m, \frac{\log(R_m/\mathcal{S}_m)}{\log m}\right), 1 \leq m \leq \text{ length(dataset)}\right\}$$

and this should asymptote at J.

To see how this works in practice, Figure 27 displays two $R/\mathcal{S}$ plots for simulated data. On the left is the $R/\mathcal{S}$ plot for 5000 N(0,1) iid variables and on the right is the $R/\mathcal{S}$ plot for 5000 iid stable variables with $\alpha = .2$ and skewness 1. Note both plots hug the height 1/2, which is the correct self-similarity parameter for normal but not the self-similarity parameter for stable, which would be 1/.2=5.

As a bonus, here is the $R/\mathcal{S}$ plot in Figure 28 for the last 5000 measurements in the Ericsson trace. The plot is definitely NOT indicating 1/2.

So, from these examples, the $R/\mathcal{S}$ statistic may detect something, but it may not be the self-similarity parameter. In particular, with the example of the iid stable random variables, $R/\mathcal{S}$

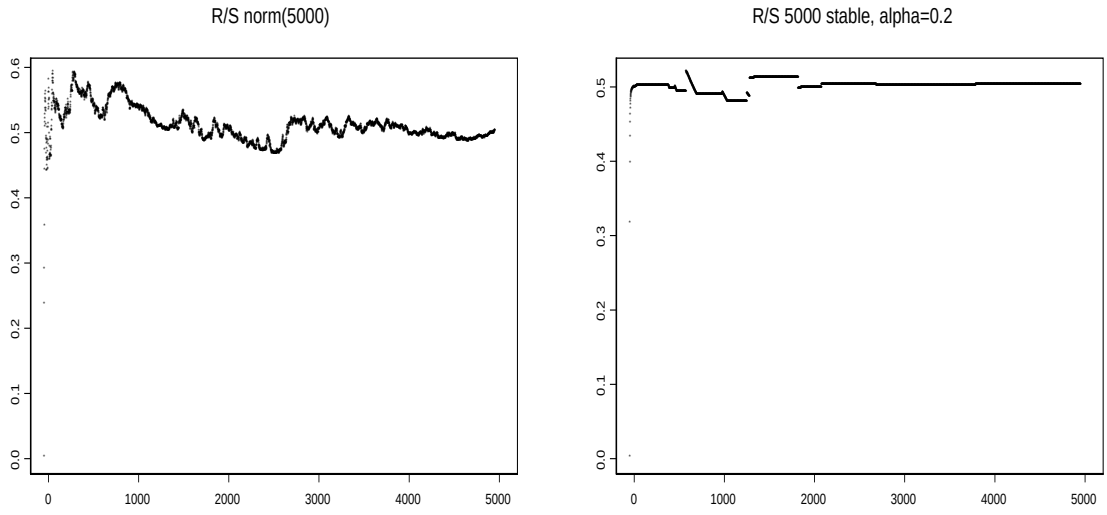

FIGURE 27. $R/\mathcal{S}$ plots $\{\big(m, (\log m)^{-1} \log(R_m/\mathcal{S}_m)\big), 1 \leq m \leq n\}$ for iid normal (left) and iid stable (right) with $\alpha = 0.2$ and skewness 1.

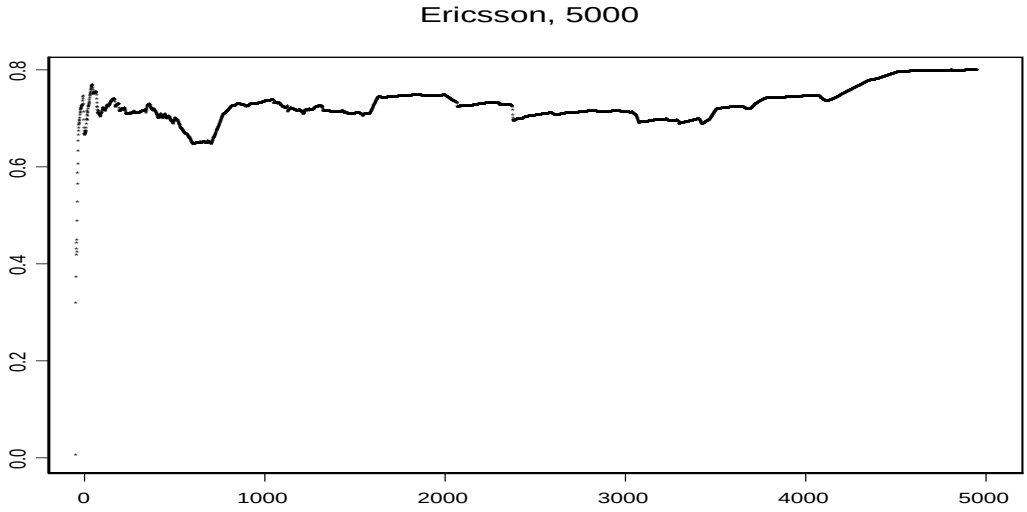

FIGURE 28. $R/\mathcal{S}$ plot $\{\big(m, (\log m)^{-1} \log(R_m/\mathcal{S}_m)\big), 1 \leq m \leq n\}$ for last 5000 measurements in the Ericsson trace.

definitely indicated 1/2 and not anything related to the self-similarity parameter. This requires some further explanation.

Notation: If $\{X(t), 0 \leq t \leq 1\}$ is a stochastic process such that $X(0) = 0$, we write

$$X^{(0)}(t) = X(t) - tX(1), \quad 0 \leq t \leq 1$$

for the bridge version. If B is a Brownian motion, $B^{(0)}$ is Brownian bridge and if X is a stable Lévy motion, then $X^{(0)}$ is a stable bridge. We also write

$$X^{\vee}(t) = \bigvee_{0\le s\le t} X(s), \quad X^{\wedge}(t) = \bigwedge_{0\le s\le t} X(s).$$

For instance, we have

$$\left(B^{(0)}\right)^{\vee}(1) - \left(B^{(0)}\right)^{\wedge}(1) = \bigvee_{0\le t\le 1} (B(t) - tB(1)) - \bigwedge_{0\le t\le 1} (B(t) - tB(1)).$$

Consider the following cases:

(1) If $\{X_n\}$ are iid with finite variance, since $R/\mathcal{S}$ is location and scale invariant, we might as well suppose $EX_n = 0, EX_n^2 = 1$. Then $\mathcal{S}_n \xrightarrow{P} 1$ and

$$\frac{R_n/\mathcal{S}_n}{\sqrt{n}} \Rightarrow (B^{(0)})^{\vee}(1) - (B^{(0)})^{\wedge}(1).$$

This certainly does not explain the Hurst phenomenon.

(2) Moran [113] speculated long range dependence could perhaps be explained by heavy tails and/or infinite variance but, alas, this is not true [104]. If $\{X_n\}$ are iid and in the domain of attraction of an α-stable distribution, then

$$\frac{R_n/\mathcal{S}_n}{\sqrt{n}} \Rightarrow \chi, \tag{6.14}$$

where χ is constructed from the stable bridge. This too does not explain the Hurst phenomenon; note the disturbing fact that $J = \frac{1}{2}$ but the self-similarity parameter $H = \frac{1}{\alpha}$.

(3) Suppose $\{Y(t), t \ge 0\}$ is H-sssi with nice paths, say in $D[0,\infty)$. Set $X_k = Y(k) - Y(k-1), k \ge 1$ and $S_n = \sum_{i=1}^n X_i$. Note $S_n = Y(n)$ since $Y(0) = 0$ from self-similarity. Then using self-similarity, it is easy to check

$$\begin{aligned}\frac{R_n}{n^H} &\stackrel{d}{=} \bigvee_{k=0}^n \left(Y\left(\frac{k}{n}\right) - \frac{k}{n}Y(1)\right) - \bigwedge_{k=0}^n \left(Y\left(\frac{k}{n}\right) - \frac{k}{n}Y(1)\right)\\ &\to \left(Y^{(0)}\right)^{\vee}(1) - \left(Y^{(0)}\right)^{\wedge}(1) = R^*.\end{aligned}$$

If also $\{X_i\}$ is ergodic with $EX_1^2 < \infty$, so that

$$\mathcal{S}_n^2 \to \mathrm{Var}(X_1),$$

then

$$\frac{R_n/\mathcal{S}_n}{n^H} \Rightarrow R^*/\sqrt{\mathrm{Var}(X_1)}.$$

The conclusion is that if

(a) $Y(\cdot)$ is H-sssi with regular paths,

(b) $(Y(k) - Y(k-1), h \ge 1)$ is ergodic with $EY(1)^2 < \infty$, then

$J = H$, and the Hurst exponent equals the self-similarity parameter. Asymptotic versions of this result could be developed.

Recent discussions of ways to fool the $R/\mathcal{S}$ statistic are in [70, 69]. See also [13].

6.5.5. *Wavelet methods.* The recently developed wavelet estimation methodology works under a variety of assumptions. If the process under investigation is stationary with finite variance and long range dependence, the method yields an estimate of the long range dependence parameter β in (6.1). When applied to the increments of a self-similar process, the wavelet method yields an estimate of H.

The wavelet method has gained adherents because it provides an appealing compromise between low computational cost and good statistical performance. It does not require an exact parametric model and since it is based on identification of scaling in a log-log diagram, it is possible to judge the range of scales on which the model fits. It is also robust to smooth non-stationarities depending on the filtering capabilities of the wavelet chosen.

Important ideas are in [3, 1, 158, 2, 159]; here is the briefest outline. Let ψ be a reference or "mother wavelet"; that is, a smooth function, well-localised in both position and frequency which satisfies the admissibility and moment conditions $\int t^n \psi(t)dt = 0,\ n = 0, 1 \ldots N$. The phrase "well-localised" means the function has compact support or at least is rapidly decaying. The moment conditions allow the wavelet transform to filter means and smooth trends and to make the procedure robust against certain departures from stationarity. Define the location-scale family ψ_{ba} by

$$\psi_{ba}(t) := \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad b \in \mathbb{R},\ a > 0.$$

Define the wavelet transform $W_\psi X$ of a process X by

$$W_\psi X(b,a) := \int \psi_{ba}(t)X(t)dt = \int \sqrt{a}\psi(s)X(as+b)ds, \quad (b,a) \in \mathbb{R} \times \mathbb{R}^+. \tag{6.15}$$

If the process $X(\cdot)$ is stationary, or has stationary increments, then the process $W_\psi X(\cdot, a)$ is again stationary from (6.15). If a process Y is H-sssi with finite second moments, then

$$E|W_\psi Y(b,a)|^2 = a^{2H+1} E|W_\psi Y(b/a, 1)|^2.$$

Define $X(t) = Y(t+\Delta) - Y(t)$ for $\Delta > 0$. By a change of variables, a Taylor expansion and self-similarity,

$$W_\psi X(b,a) \sim -a^{H-1/2}\Delta \int \psi'(s)Y(s)ds, \quad a \to \infty, \tag{6.16}$$

and if second moments exist,

$$\mathrm{Cov}(X(0), X(a)) = E|W_\psi X(b,a)|^2 \sim K a^{2H-1}, \quad a \to \infty. \tag{6.17}$$

For comparison, note that if $X = B_H$, fractional Brownian motion, then for t fixed and $a \to \infty$ we have for a constant $\sigma > 0$

$$\mathrm{Cov}\big(X(t), X(t+a)\big) = \sigma\Big((t+a)^{2H} + t^{2H} - a^{2H}\Big) \sim 2\sigma H t a^{2H-1}.$$

Now suppose only that $X(\cdot)$ is a long range dependent stationary process satisfying the analogue of (6.1), then

$$\mathrm{Cov}(X(0), X(a)) = E|W_\psi X(b,a)|^2 \sim K a^{-\beta}, \quad a \to \infty, \tag{6.18}$$

which agrees with (6.17) for fractional Gaussian noise. The wavelet estimator is a regression method based on (6.17) or (6.18), which makes use of the fact that the wavelet transform is less correlated than the process.

In practice, wavelet coefficients are computed on a dyadic grid, so define

$$d(j,k) = W_\psi X(2^j k, 2^j), \quad j, k \in \mathbb{Z}.$$

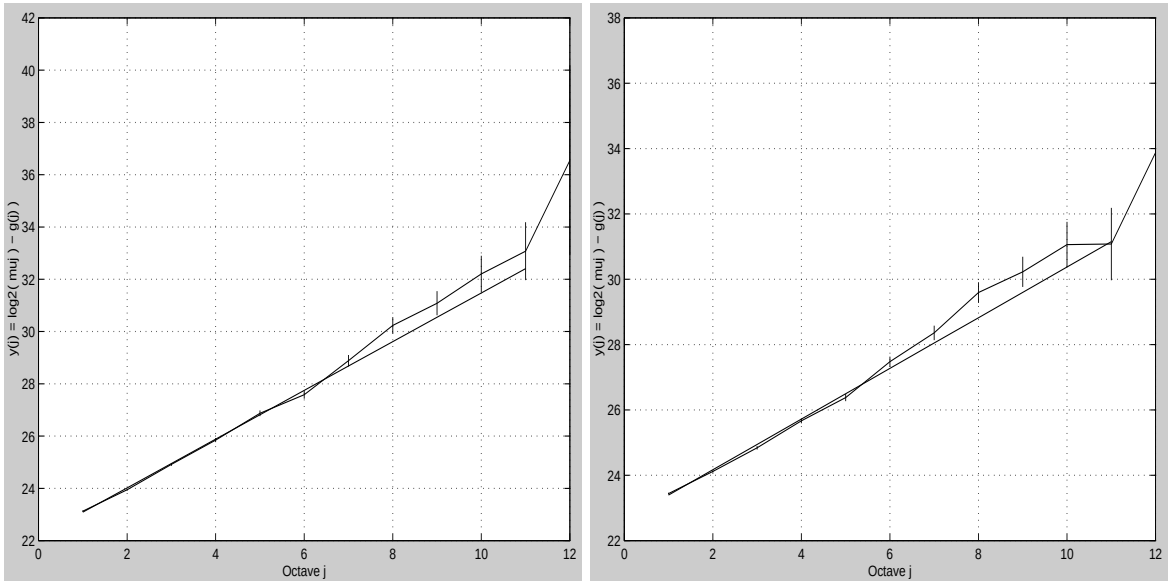

FIGURE 29. Wavelet regression plots for the Munich traces: Receiving (RX) on left and transmitting (TX) on right.

By (6.18),

$$\mathrm{Cov}(X(0), X(2^j)) = E|d(j,k)|^2 \sim K 2^{-j\beta}, \quad j \to \infty, \tag{6.19}$$

which suggests averaging coefficients at fixed scale,

$$\mathrm{Cov}(X(0), X(2^j)) = \frac{1}{n_j} \sum_k |d_{j,k}|^2,$$

where n_j is the number of available coefficients at scale 2^j. The parameter β in (6.18), or equivalently H, can then be estimated from a linear regression in the log-log diagram of $\{(j, \frac{1}{n_j}\sum_k |\hat{d}_{j,k}|^2)\}$, where $\hat{d}_{j,k}$ is the estimated wavelet coefficient obtained from the data.

There are refinements to reduce bias induced by the log-transform and to replace an ordinary least squares by weighted least squares. See [159, 3].

Example 7 (Munich Trace)**.** The *Munich low* data set contains measurements of cell rates for both the sending (TX) and receiving (RX) directions of an ATM link. The data was collected on Wednesday, November 12, 1997 (TX) and Wednesday, December 17, 1997 (RX) with a temporal resolution of 2 seconds, *i.e.* the total number of cells that passed the ATM link every 2 seconds was recorded. Recorded traffic was pure IP (mainly HTTP, FTP, and NNTP) data traffic, without any audio/video components. A shorter sample covering the period between 10 a.m. until 1 p.m. was selected for analysis. The shorter period was chosen to obtain a roughly stationary data set.

The estimated cumulative traffic looks neither Gaussian nor stable so the approximations in Theorems 1 and 2 do not seem in evidence. There is a clear and strong long range dependence, which persists over a wide range of scales as seen by autocorrelation functions and wavelet regression plots. In both traces, the estimates of H are close to 0.95. See Figure 29.

7. Does the model fit the data? Small time scales: Hölder exponents and multifractality.

It is now rather widely accepted (but not without some controversy) that network traces examined at time scales above a certain resolution appear statistically self-similar. The limiting processes discussed in Theorems 1 and 2 are self-similar. Lamperti's theorem [96] says to expect self-similarity whenever a process is obtained as a distributional limit by dilating time linearly and scaling space. (See the discussion following (6.8) and (6.9).)

The observed self-similarity seems to hold above a resolution of about 100 milliseconds and not below. At smaller thresholds, certain researchers fervently believe there is evidence for chaotic behavior or extreme intermittency, which may be typical of *multifractal* phenomena. See [119, 145, 62, 68, 144, 143, 163]. Current statistical goals are to verify if multifractality is really present and develop methods for dealing with it statistically. The probabilistic goal is to develop model based explanations for why the multifractality should be present. See [105, 23, 129, 107]. It is at the small time scales that one must seek the influence of protocols and network architecture.

There seem to be two literatures, which have not intersected much. One grows out of the Gaussian process tradition and defines Hölder exponents in a mean square sense. See [85, 84, 83, 86]. The other is path based. A nice summary is [143]. We give a brief summary of both approaches.

7.1. **Second order definition.** The semivariogram V of a second order process $\{X(t)\}$ is defined by

$$V(t,\tau) = \frac{1}{2}E(X(t+\tau) - X(t))^2,$$

and contains the same information as the covariance function. If $X(\cdot)$ is H-ss, then for $\theta > 0$, we have $\theta^{-2H}V(\theta t, \theta\tau) = V(t,\tau)$. If $X(\cdot)$ also has stationary increments, V does not depend on its first variable, $V(t,\tau) = V(\tau)$, and if it is also self-similar, then $V(\tau) = V(1)\tau^{2H} := c\tau^{2H}$. If the process is also a centered Gaussian process, this means that it is a fractional Brownian motion.

A stationary process $X(\cdot)$ has local Hölder index $H_o(t)$ at t (in the mean square sense) if the semivariogram satisfies

$$V(t,\tau) = c\tau^{2H_o(t)} + o(\tau^{2H_o(t)}), \quad \text{as } \tau \to 0, \tag{7.1}$$

for each t. For fBm, $H_o(\cdot)$ is constant and $H = H_o(t)$. This is significant because if empirical estimates of $H_o(t)$ for various values of t either reject the constancy of $H_o(\cdot)$ or indicate, for instance, that $H_o(0)$ is not equal to the estimate of H, then we have evidence against approximation to the cumulative traffic by fractional Brownian motion. For a Gaussian process, the Hölder index process gives precise information on sample paths ([4, 82]), on the rate of convergence of non-parametric estimates of the covariance function ([86]) and on the asymptotic behavior of wavelet coefficients of X ([83]). A final preliminary comment is that this approach is limited to L_2 processes and has no direct relevance to heavy tailed processes.

One method to estimate $H_o(t)$ can be based on simple empirical quadratic variation [91, 71, 26, 85]. Assuming the underlying process $X(\cdot)$ has stationary increments allows one to focus on $t = 0$. At scale $1/n$, define

$$V_n = \frac{1}{2}\sum_{k=1}^{n} \left(X(k/n) - X((k-1)/n)\right)^2.$$

Suitably normalized, this converges with probability 1,

$$\lim_{n\to\infty} n^{2H_o(0)-1}V_n = c,$$

where c appears in (7.1).

The quadratic variation estimator is not scale invariant and may converge slowly, so there are proposed refinements which weight the squared terms and then take ratios and which also filter certain polynomial trends. In [72], the quadratic variation based estimator of $H_o(0)$ discussed in [85, page 437] was used to study various traces. Empirical evidence suggested that $H_o(t)$ was not constant and in no case was it found that $H = H_o(0)$. The estimator of $H_o(0)$ is designed for Gaussian processes and more experience in traffic contexts is required to assess its reliability.

Table 7.1, from [72] compares estimated values of H and $H_o(0)$ for various traces. The agreement is not good.

Data set	$\hat{H}$	$\hat{H}_o(0)$
simM/G/∞	$.90 \pm .01$	.88
BUburst 10s	$.89 \pm .02$	.73
BUburst 1s	$.81 \pm .01$	.87
UCB 10s	$.58 \pm .03$	.65
UCB syn 10s	$.95 \pm .07$	1.36
Munich lo TX	$.89 \pm .01$	.86
Munich lo RX	$.97 \pm .01$	.85
Munich hi .1s	$1.02 \pm .03$	.66
Munich hi .01s	$.1.03 \pm .04$	.56
Ericsson	$.88 \pm .02$	1.21
Eri syn 1s	$1.48 \pm .02$	1.51

TABLE 2. Comparison of estimated values for H and $H_o(0)$ for various data sets.

7.2. **Pathwise definition.** Since self-similarity is a distributional property, it seems like mixing apples and oranges to consider small time scale properties in the pathwise sense. In the literature, two types of Hölder exponents have been considered, namely, one based on exponential growth rate and one based on polynomial approximation. We consider the first.

The *Hölder exponent based on exponential growth rate* of the function $x(\cdot)$ at t is defined as

$$h_x(t) := \liminf_{\epsilon \downarrow 0} \frac{\log \sup_{u:|u-t| \le \epsilon} |x(u) - x(t)|}{\log \epsilon}. \tag{7.2}$$

For non-decreasing functions, the definition of $h_x(t)$ in (7.2) simplifies to

$$h_x(t) = \liminf_{\epsilon \downarrow 0} \frac{\log[(x(t+\epsilon) - x(t)) \wedge (x(t) - x(t-\epsilon))]}{\log \epsilon} = \liminf_{\epsilon \downarrow 0} \frac{\log(x(t+\epsilon) - x(t-\epsilon))}{\log \epsilon}. \tag{7.3}$$

The Hölder exponent of the sum of two functions satisfies the following inequality. For two functions x and y, we have

$$h_{x+y}(t) \ge h_x(t) \wedge h_y(t). \tag{7.4}$$

Furthermore, equality holds if $h_x(t) \neq h_y(t)$ or if x and y are monotone. We are, of course interested in Hölder exponents of sums of functions because of the need to superimpose traffic streams.

How do we decide which values of the Hölder exponent are common? One way to do this is with the *multifractal spectrum*. The multifractal spectrum of the Hölder exponent of the function x for the Hölder exponent based on exponential growth rate is

$$d_x(a) = \dim(\{t > 0 : h_x(t) = a\}), \quad a \in [0, \infty),$$

where for a set Λ, $\dim(\Lambda)$ is the Hausdorff dimension (cf. Chapter 2 of [57]) of Λ.

An increasing Lévy process $X(\cdot)$ has the interesting property that, restricted to any interval, it has a non-random multifractal spectrum for the Hölder exponent based on exponential growth rate [87]. In fact the spectrum is linear and is thus non-trivial. Empirically, spectra that have been estimated from traffic, have never been observed to be linear. Neither do observed spectra correspond to the monofractal structure of fbm.

Note that if we require the increasing Lévy process $X(\cdot)$ to also be an α-stable Lévy motion, then $X(\cdot)$ is $1/\alpha$-ss (a distributional property) as well as a random multifractal with a non-random multifractal spectrum.

8. A model for large and small time scales.

The good news about the infinite source Poisson model is that it provides a simple, compelling reason why long range dependence exists in the traffic data. The not so good news is:

(1) The model does not fit the data all that well.
   (a) The constant transmission rate assumption is clearly wrong.
   (b) Not all times of transmission initiation can be modeled as Poisson. Presumably humans acting independently can have the times of their acts modeled as a Poisson process but machine triggered events, in say web browsing, will never be Poisson.
   (c) There is no hope the simple model can successfully match fine time scale behavior observed below, say, 100 milliseconds which is speculated to be multifractal.
(2) The model predicts that cumulative traffic can be approximated by either FBM or stable Lévy motion. In practice, traffic is not observed to be heavy tailed, and while some report traffic at heavily loaded links is Gaussian, there is the disturbing evidence that Hölder exponents are not constant and not equal to the self-similarity parameters, as would be the case if the FBM approximation were valid.
(3) The traffic data is often not *stationary*. There are always time of day effects. This has, so far, been ignored.

This is quite a laundry list of defects needing repair. Let's make a start and at least try to address 1(c).

8.1. **A more general model appropriate for the study of small and large time scales.**

Here are the ingredients of a more general model.

(1) Denote the time when the k-th transmission begins by Γ_k. The sequence $\{\Gamma_k\}$ is strictly increasing to ∞.
(2) The size of the file required to be transmitted at Γ_k is J_k and we assume $J_k > 0$.
(3) There is a transmission schedule, denoted by $A_k(\cdot)$, where $A_k(t)$ denotes the amount of data transmitted in time t after the kth transmission has begun. It is a non-decreasing càdlàg function starting at 0 and increasing to ∞, which vanishes on the negative real axis.

With these assumptions, the accumulated traffic in $[0,T]$ aggregated over the different users is

$$A(t) = \sum_{k=1}^{\infty} A_k(t - \Gamma_k) \wedge J_k. \tag{8.1}$$

Because the assumption of unit transmission rate is highly arbitrary and restrictive, there have been efforts to remove this deficiency. A non-constant but deterministic transmission schedule has been considered in [93, 137] and with a constant input rate it was shown that a Lévy stable motion approximation could be fashioned. A transmission schedule which was linear with randomly chosen slope (chosen not necessarily independent of the file size) is considered in [108]. Related results about renewal reward processes are given in [156, 103, 102, 123]. However, none of these efforts captures the small time scale properties. Our model allows a random, time-dependent transmission schedule $A_k(\cdot)$ and random file size J_k chosen at the beginning of each transmission. The length of transmission L_k is obtained as a function of these two random quantities and is readily defined as

$$L_k = A_k^{\leftarrow}(J_k) := \inf\{s : A_k(s) \geq J_k\}.$$

This model suggests that the fine time scale behavior of network traffic results from individual transmission schedules exhibiting multifractality. This results in multifractal behavior for the cumulative taffic process at the microscopic level, and still gives a stable Lévy motion as the macroscopic approximation. This may not be good news if you believe this approximation has never been seen in practice but at least this set-up allows multifractal behavior on fine time scales and self-similar scaling at large time scales. The suggestion that the multifractal behavior of the cumulative traffic process results from similar behavior of the individual input processes, presumably due to extreme intermittency and blocking caused by congestion at various nodes, needs to be verified by empirical study of individual, user level input processes.

The following summarizes findings in [107].

8.1.1. *Small time scale behavior.* The small time scale behavior of the cumulative traffic process $A(\cdot)$ requires the following further minimal assumptions on the transmission schedule $\{A_k\}$:

(4) Suppose $\{A_k\}$ are identically distributed with stationary increments.
(5) The multifractal spectrum of $A_k(\cdot)$ is not degenerate at a single point. This ensures that we consider processes with paths that show real multifractal behavior.
(6) The multifractal spectrum of $A_k(\cdot)$ restricted to any (non-random) interval is non-random.

If A_k is, for example, an increasing Lévy process, then, restricted to any interval, it has a non-random multifractal spectrum for the Hölder exponent based on exponential growth rate [87].

If the assumptions (1) – (6) hold, then with probability 1,

$$d_A = d_{A_1}. \tag{8.2}$$

So the aggregate traffic process $A(\cdot)$ inherits the multifractal structure of the individual transmission schedules.

8.1.2. *Large time scale behavior.* Small time scale behavior was a path by path analysis. Large time scale analysis requires distributional assumptions.

(7) Suppose $\{\Gamma_k\}$ forms a homogeneous Poisson process with intensity parameter λ.
(8) Assume $\{(A_k, J_k) : k \geq 1\}$ are iid and independent of $\{\Gamma_k\}$.
(9) Assumptions about the joint distribution of $(A_1(\cdot), J_1)$: Suppose
  (a) there exists a regularly varying function $\sigma(\cdot)$ with index $H > 0$;

(b) there exists a proper random process χ (that is, $\chi(t)$ is non-degenerate for each t) with stationary increments, taking values in $\mathbb{D}[0,\infty)$.

Then suppose that for each fixed $\epsilon > 0$,

$$\frac{1}{\bar{F}_J(\sigma(T))} P\left[\frac{J_1}{\sigma(T)} > \epsilon, \frac{A_1(T\cdot)}{\sigma(T)} \in \cdot\right] \xrightarrow{w} \epsilon^{-\alpha_J} P[\chi \in \cdot]$$

on $\mathbb{D}[0,\infty)$ where "$\xrightarrow{w}$" denotes weak convergence and we assume $1 < \alpha_J < 2$.

(10) Technical: For all $\gamma > 0$, assume

$$\lim_{\epsilon \downarrow 0} \limsup_{T\to\infty} \frac{1}{\bar{F}_J(\sigma(T))} P\left[\frac{J_1}{\sigma(T)} \le \epsilon, \frac{L_1}{T} > \gamma\right] = 0.$$

(11) Technical: Assume

$$E\left[\chi(1)^{-\alpha_J}\right] < \infty.$$

These assumptions lead to certain consequences and there are several simplifying conditions which make the assumptions easier to check. Here is extra information which should make the framework clearer.

(A) The limit χ is H-self-similar. To see this, one must elaborate the methods used to prove Lamperti's theorem [96, 55, 161, 160, 115, 114].

(B) The limit χ being càdlàg, self-similar and non-degenerate implies that $H > 0$.

(C) The limit χ being non-decreasing implies $H \ge 1$.

(D) If $H = 1$, then $\chi(t) = t\chi(1)$ is linear and no interesting structure exists so we exclude this possibiliity.

(F) The distribution tail of J_1 is regularly varying with index $-\alpha_J$:

$$\bar{F}_J(x) = 1 - F_J(x) = x^{-\alpha_J} L(x), \quad x > 0.$$

(G) Cumulative traffic on large time scales looks like stable Lévy motion. We describe this in more detail next.

The precise statement for (G) is as follows: Suppose (1)–(3) and (7)–(11) hold with $1 < \alpha_J < 2$ and define

$$Y_T(t) = \frac{A(Tt) - \lambda T t E(J_1)}{b_J(T)},$$

where

$$b_J(T) = \left(\frac{1}{1 - F_J}\right)^{\leftarrow}(T)$$

is the usual quantile function. Then we have

$$Y_T \xrightarrow{fidi} Z_{\alpha_J}, \tag{8.3}$$

where Z_{α_J} is a mean 0, skewness 1, α_J-stable Lévy motion.

Here are some simpler sufficient conditions for (8.3) to hold.

(i) J_1 and A_1 are independent.

(ii) J_1 has a tail of index α_J.

(iii) A_1 is itself a proper H-ss process.

(iv) $E[A_1(1)^{-\rho}] < \infty$ for some $\rho > \alpha_J$.

Note that if $A_1(\cdot)$ is a $\frac{1}{H}$-stable Lévy motion, then (iii) and (iv) are automatically satisfied.

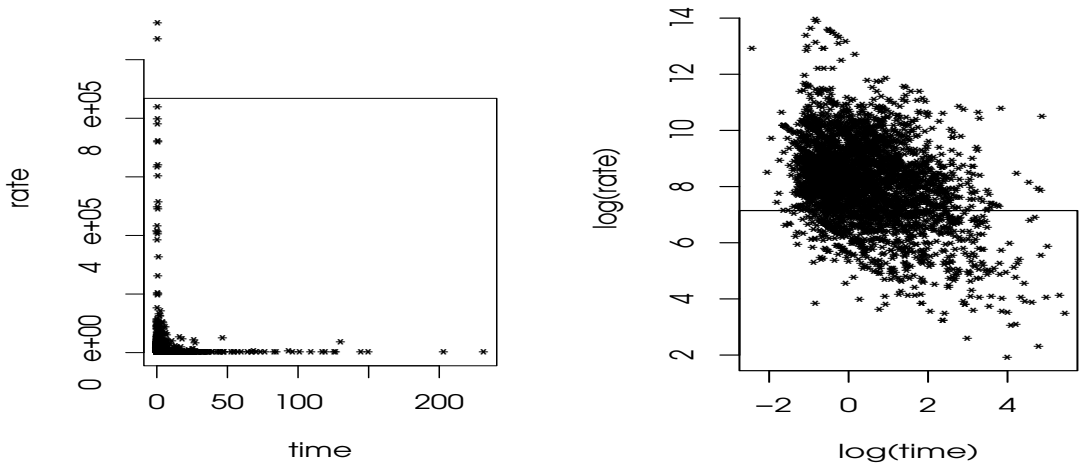

FIGURE 30. Plot of the length of transmission against the rate of the transmission of the BUburst data in (left) natural scale and in (right) log-log scale

8.2. **A model more amenable to statistics.** Section 8 gives a model which successfully exhibits both small scale multifractality and large scale self similarity. However, it is not easy to do statistics with this model. Here we will see what can be achieved by dropping the assumption of fixed unit transmission rate and allowing the rate to be random. We will examine whether the rate is independent of the transmission duration. Some of this material is adapted from [108].

It is difficult to conclude from evidence in measured data that the rate and the length of the transmission are always independent, but in certain cases we may reasonably assume that the rate and the length of the transmission are at least asymptotically independent. As an example, we consider the BUburst dataset considered by [72]. This is data processed from the original 1995 Boston University data described in the report by [36] and catalogued at the Internet Traffic Archive (ITA) web site `www.acm.org/sigcomm/ITA/`. A plot of the transmission length against the transmission rate (Figure 30) shows that most of the data pairs hug the axes, which suggests the variables are at least asymptotically independent. However, if we plot the data in the log scale on both the axes, then a weak linear dependence is observable and the correlation coefficient between the two variables after log transform is approximately -0.379, which argues against an independence assumption. We consider the log transform to make the variables have finite second moment, so that correlation coefficient becomes meaningful.

The Hill estimates obtained for the transmission length, the transmission rate and the size of the transmitted file are 1.407, 1.138 and 1.157 respectively. The corresponding Hill plots are given in Figure 31. For each of the variables, the first column gives the Hill plots, the second column gives the AltHill plots (see Subsubsection 5.1.3) and the third plot, named the Stărică plot, is an exploratory device suggested in [155, Section 7] to help decide on the number of upper order statistics to be used. (It was designed for higher dimensional problems but is basically a one-dimensional technique.) It uses the fact that for a random variable X with Pareto tail of parameter

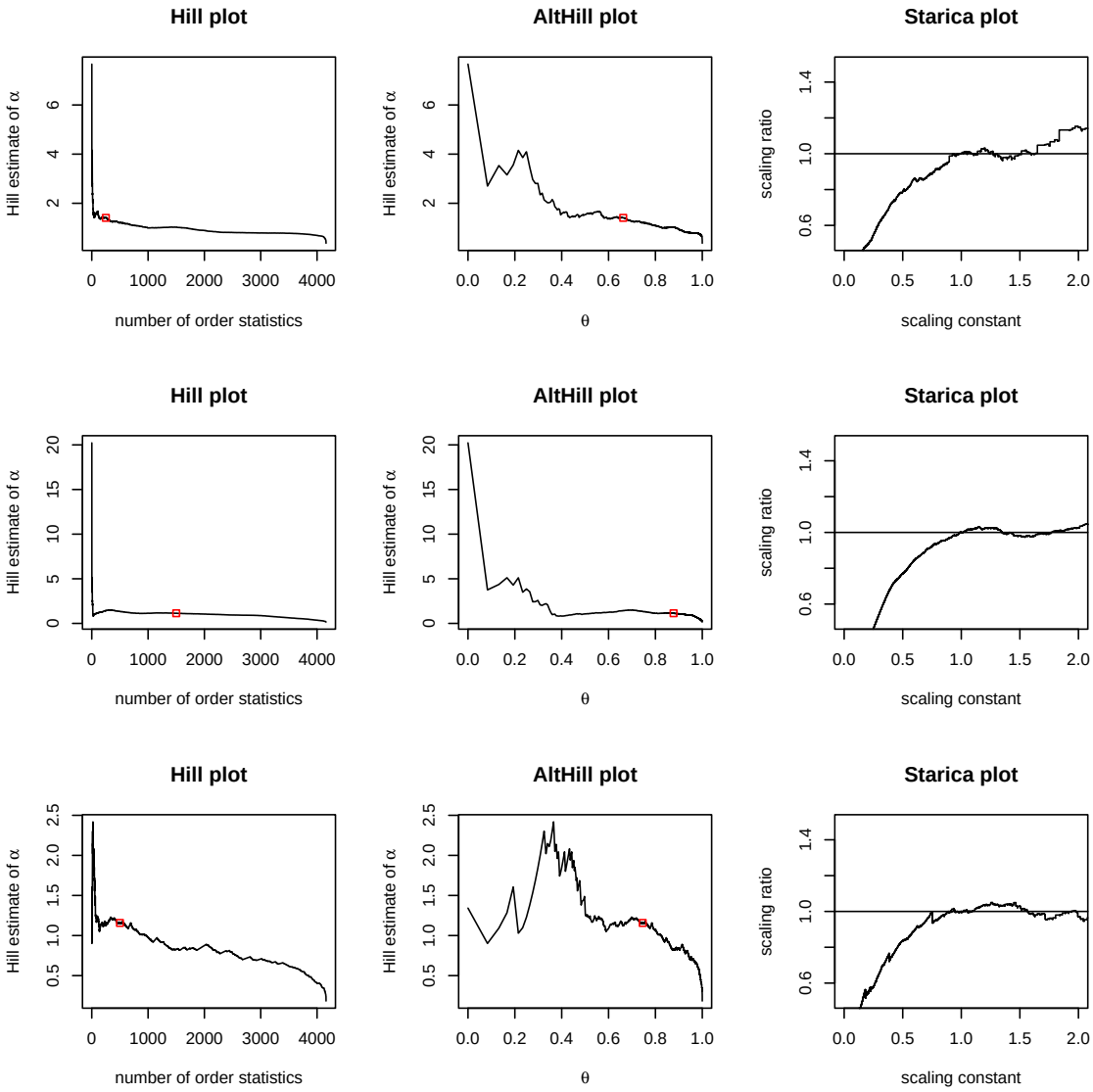

FIGURE 31. Hill plots of transmission length (top), transmission rate (middle), and transmitted file size (bottom).

α, we have

$$\lim_{T\to\infty} TP\left[\frac{X}{T^{\frac{1}{\alpha}}} > r\right] = r^{-\alpha}.$$

For every k, we estimate the left hand side by

$$\hat{\nu}_{n,k}\big((r,\infty]\big) = \frac{1}{k}\sum_{i=1}^{n} \mathbf{I}\left[\frac{X_i}{(n/k)^{1/\hat{\alpha}_{n,k}}} > r\right].$$

We expect the ratio of $\hat{\nu}_{n,k}\big((r,\infty]\big)$ and $r^{-\hat{\alpha}_{n,k}}$, called the scaling ratio, to be approximately 1, at least for values of r in a neighborhood of 1, if we have made the correct choice of k. In the Stărică plot, we plot the above scaling ratio against the scaling constant r, and choose k so that the graph hugs the horizontal line of height 1. The interesting point to be noted is the fact that the rate of the transmission has a much heavier tail than the length of transmission. So a model with a random transmission rate with heavy tails seems appropriate.

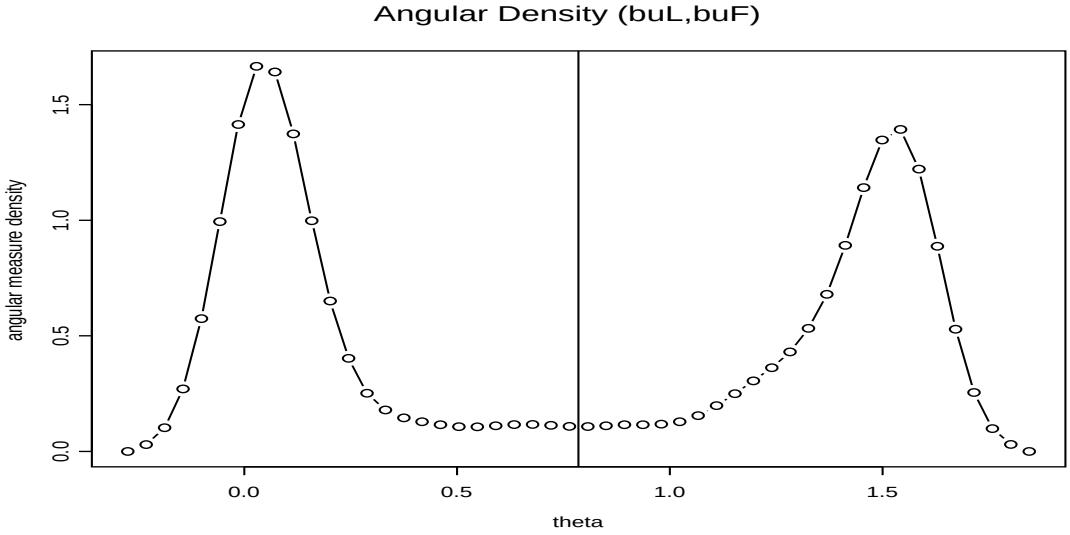

FIGURE 32. The data (buL,buF): Density estimate for the angular measure.

When two random variables are heavy tailed, a modest assumption on the joint distribution $F(\boldsymbol{x})$ of the pair is that the distribution be in the domain of attraction of an extreme value distribution. When both marginal distributions of $F(\boldsymbol{x})$ are the same. the extreme value distribution will have an exponent measure $\nu_*(\cdot)$ in *standard* form. See ([45, 139, 43, 46]. However, when the two marginal distributions are not the same (which is always the case except in simulations), we must first rescale each component of the bivariate vector by its quantile function (or its estimate, the k-th largest marginal order statistic) and then adjust for the difference in tail exponents by powering each component of the random vector by the tail exponent (or its estimate). The transformation to polar coordinates

$$T : \boldsymbol{x} \mapsto (r, \theta)$$

induces a product form on ν_*,

$$\nu_* \circ T^{-1}(dr, d\theta) = cr^{-1}S(d\theta),$$

for some constant $c > 0$ and $S(\cdot)$ a probabilty measure. When the bivariate vector $\boldsymbol{X}$ concentrates on $\mathbb{R}^2_+$, $S(\cdot)$ concentrates on $[0, \pi/2]$. Asymptotic independence corresponds to S concentrating on $\{0\}$ and $\{\pi/2\}$. If the bivariate observations are $\boldsymbol{X} = \{(X_{1,i}, X_{2,i}), 1 \le i \le n\}$, then we transform

$$\boldsymbol{X} \to \Big\{ \Big((\frac{X_{i1}}{X_{1,(k)}})^{\alpha_1}, (\frac{X_{i2}}{X_{2,(k)}})^{\alpha_2} \Big); i = 1, \ldots, n \Big\} \stackrel{T}{\mapsto} \{(R_i^{(n)}, \Theta_i^{(n)}),\ 1 \le i \le n\}$$

and we need to estimate S from

$$\{\Theta_i^{(n)} : R_i^{(n)} > 1\}.$$

If we make a density plot of these Θ's, we should be able to get an impression of the tendency toward asymptotic dependence. Modes at 0 and $\pi/2$ indicate a tendency toward asymptotic independence while a mode at $\pi/4$ indicates a tendency towards asymptotic dependence.

For the (buL,buF) data, that is, the length of download against the file size downloaded, the estimated angular measure density, given in Figure 32, shows 2 modes at 0 and $\pi/2$ indicating a tendency toward independence. The angular measure estimation was made using $k = 100$. This is based on the Stărică plot given in Figure 33 which shows the superiority of the choice $k = 100$.

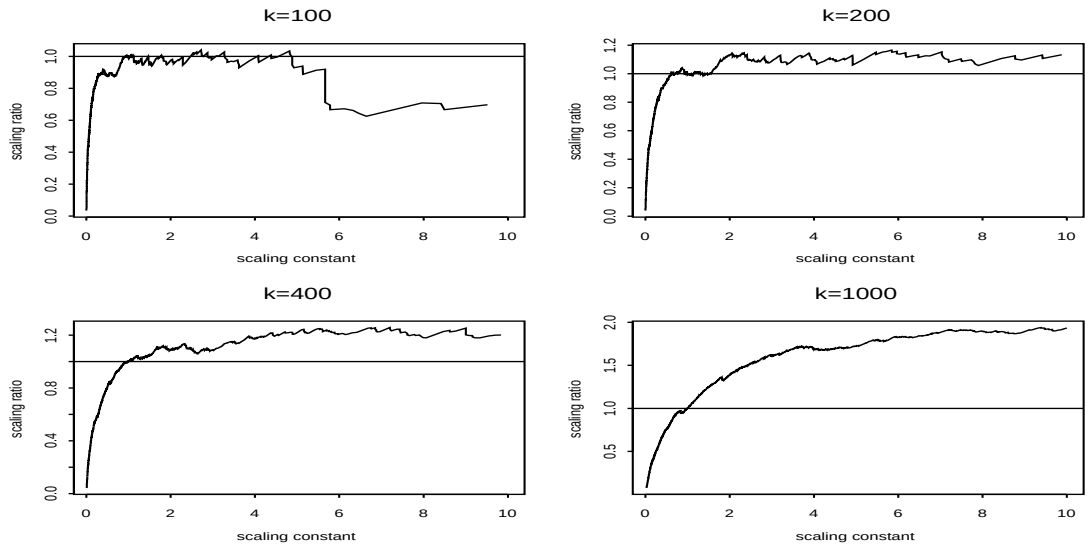

FIGURE 33. The data (buL,buF): Stărică plot showing preference for choice $k = 100$.

Interestingly, the analysis of (buL,buR), the download length vs the inferred transmission rate, shows much less tendency toward asymptotic independence (see [108]) and the whole issue of the dependence between download time, download rate and downloaded file size deserves more careful study. See [20] for more comments on this point.

8.2.1. *A model for asymptotic independence.* An obvious generalization of the infinite source Poisson model is to retain Poisson connection times but generalize the transmission mechanism. We assume that there are iid pairs $\{(L_k, R_k), k \geq 1\}$ giving length of a download and rate of transmission so that when a download is initiated, a random rate is chosen for the transmission. Empirically it is often the case that each marginal distribution of L_k and R_k is heavy tailed. The file transmitted will, of course, be $L_k R_k$ and analyzing the infinite source Poisson model will require knowledge of how the product $L_k R_k$ behaves. The classical extreme value notion of asymptotic independence is not strong enough to generate useful information about the product and a strengthening is required. We briefly discuss the approach assuming that L has the heavier tail; a minor modification is necessary if R has the heavier tail.

Define the measure ν_α on $(0, \infty]$ by

$$\nu_\alpha(dx) = \alpha x^{-\alpha-1} dx, \quad x > 0, \alpha > 0.$$

We assume the following.

(1) Asymptotic independence. There exists a probability distribution G on $\mathbb{R}_+$ and a sequence $b_n \to \infty$ such that

$$nP[(\frac{L}{b_n}, R) \in \cdot] \xrightarrow{v} \nu_\alpha \times G,$$

vaguely on $\mathbb{E} := (0,\infty] \times [0,\infty]$. This implies that L has a regularly varying tail with parameter α and that for large n, R and large values of L tend to be independent.

(2) Tightness. For some $\eta > 0$,

$$\lim_{\epsilon \downarrow 0} \limsup_{n\to\infty} nE[\Big(\frac{L}{b_n}R\Big)^{\alpha+\eta} 1_{[L \le b_n \epsilon]} = 0.$$

Then provided

$$c := \int_0^\infty u^\alpha G(du) < \infty$$

we have

$$P[LR > x] \sim cP[L > x], \quad x \to \infty.$$

It turns out that (1) holds iff multivariate regular variation ([139, 9, 8]) holds and suggests that (L, R) is multivariate regularly varying but in the wrong coordinate system:

$$nP[b_n^{-1}(LR, R) \in \cdot] \xrightarrow{v} \mu$$

on $\mathbb{E}$. It is not hard to see that μ and $\nu_\alpha \times G$ are related.

The implication for the modified infinite source Poisson model is that cumulative traffic at large time scales is still approximated by a stable Lévy motion. It is undoubtedly true, in addition, that under the right asymptotic regime, FBM limits are possible. This is either good or bad news depending on whether you believe cumulative traffic is or is not either FBM or stable Lévy motion. And keep in mind that in the data example given, it was (buL,buF) that seemed to be asymptotically independent (in the classical sense) and not the pair (buL,buR). When L and R both have regularly varying tails with parameters α_L and α_R in the range $(1, 2)$ but L and R are not asymptotically independent, $F = LR$ has tail index $\alpha_F < 1$ [108]. Ouch!

9. A model with a control.

One of the obvious problems with the previous models is that they do not incorporate controls and protocols which have been designed to make the real internet robust and scalable. One reason why it may be unlikely to observe stable Lévy motion as an approximation to cumulative traffic is that controls prevent the big buildup of traffic that a heavy tailed model would predict. Protocols, such as tcp (transport control protocol) are designed to keep congestion at the end user when packets are being lost by the simple mechanism that the sending machine is signaled to slow transmission. Roughly, how it works is as follows [92]: files are broken into packets and the first packet is sent. If an acknowledgement from the destination is received within a specified time window, then the sender sends two packets. As long as acknowledgements are received from the destination machine, transmission continues to increase linearly. However, if no acknowledgement is received, the packet is presumed lost and resent but the transmission rate is halved. The rate continues to be halved until an acknowledgement is received at which point linear build-up of the rate resumes.

It is a curious fact that the transmission rate is halved when packets are lost. Why not divide the rate by 3? We sought a model which would help us study this question as well as the delay the

protocol imposes on the sender. What price in terms of delay does a sender pay for the socially or globally optimal behavior of allowing the transmission rate to be slowed?

Because it is very difficult to model the complex interactions in the internet in a realistic way, we have studied these questions in a restricted, simplified context of a single channel on/off souce destination pair with a buffer in between. See [17].

The model consists of iid jobs or files of size $\{L_n\}$, which have common distribution F_L as well as "off" periods denoted by the iid random variables $\{Y_n\}$, which have common distribution F_Y and which are independent of the file sizes. There is a buffer between the source and destination and the buffer empties at constant rate r, assuming the buffer is non-empty. We assume $r \in (0,1)$ to ensure that buffer content can exit from 0. In place of the system of acknowledgements, we assume that there is a critical buffer level $\gamma > 0$. During a transmission in this single channel model, if

- buffer content $< \gamma$, then the transmission *rate* increases linearly like $1 + at$; and if
- buffer content $\geq \gamma$, then the transmission *rate* decreases geometrically like e^{-dt}.

We suppose X_n is the buffer content at time of the beginning of the n-th transmission. The process $\{X_n\}$ will have a Markov chain structure. Here is a more formal outline of the construction of the process.

9.1. **Construction.** Step 1. Cumulative input function. Begin by defining the cumulative input function $I_x(t)$, which is the amount of work injected into the system in $[0,t]$ starting from initial content x, assuming the job size is ∞. There are two cases: $x < \gamma$ and $x \geq \gamma$.

Case $x \geq \gamma$. Define

$$t_0(x) = \inf\{u > 0 : x + \int_0^u e^{-dw} dw - ru = \gamma\},$$

which is the time necessary for buffer content to move from level x down to γ. Provided $t \in (0, t_0(x))$, we have

$$\frac{d}{dt} I_x(t) = e^{-dt}.$$

At time $t_0(x)$, we reset the input rate to 1, and then let it decrease exponentially again. Repeat this at times $t_0(x) + nt_0(\gamma), n \geq 1$.

Case $x < \gamma$. Define

$$t(x) = \inf\{u > 0 : x + \int_0^u (1 + as) ds - ru = \gamma\}$$

as the time required to move from x up to level γ. At time $t(x)$, the input rate is $1 + at(x)$ and the input rate begins to decrease exponentially until time $t_a(x) + t(x)$, where

$$t_a(x) = \inf\{u : \gamma + \int_0^u (1 + at(x)) e^{-dw} dw - ru = \gamma,$$

at which time, buffer content is again γ. Then at $t(x) + t_a(x)$,

- reset the input rate to 1, and let the rate decrease exponentially until the content level is again γ. Then
- reset the input rate to 1, let the rate decrease exponentially until content is again γ. Then
- $\cdots$.

Step 2. The transmission time $\tau(x,l)$. We have defined $I_x(t)$ assuming there was an infinite file to send. What if the file is of size l? The time necessary for transmitting a file of size l when content is initially x is

$$\tau(x,l) = I_x^{\leftarrow}(l) = \inf\{t > 0 : I_x(t) = l\}.$$

STEP 3. THE EMBEDDED MARKOV CHAIN. Define the embedded Markov chain $\{X_n\}$. With $\{L_n\}, \{Y_n\}$ representing iid file sizes and off periods, define

$$X_n = \Big(X_{n-1} + L_n - r\tau(X_{n-1}, L_n) - rY_n\Big)_+$$

to be the buffer content at the beginning of transmissions of files. This says the buffer content when a transmission begins is the content when the previous file was sent, plus the file sent minus what left the buffer during the transmission of the file minus what left the buffer during an off period.

STEP 4. A TIME SCALE. If X_n is the buffer content when the n-th transmission commences, let S_n be the time and we define

$$S_0 = 0, \quad S_n = S_{n-1} + \tau(X_{n-1}, L_n) + Y_n.$$

We think of $[S_n, S_{n+1})$ as the nth transmission cycle and it consists of an on period of length $\tau(X_{n-1}, L_n)$ plus an off period. Then we can define a continuous time process between the times $\{S_n\}$: for $t \in [S_{n-1}, S_{n-1} + \tau(X_{n-1}, L_n))$

$$X(t) = X_{n-1} + I_{X_{n-1}}(t - S_{n-1}) - r(t - S_{n-1})$$

and for $t \in [S_{n-1} + \tau(X_{n-1}, S_n)$,

$$X(t) = \Big(X\big(S_{n-1} + \tau(X_{n-1}, L_n)\big) - r(t - [S_{n-1} + \tau(X_{n-1}, L_n)]\Big)_+.$$

9.2. **Results.** Here is an outline of features of this model.

1. CONTROLLED GROWTH. The Markov chain can increase only in controlled way due to the effect of the control. If $X_n \geq \gamma$ then

$$X_{n+1} - X_n \leq \sup_{t \geq 0}\Big(\int_0^t e^{-dw} dw - rt\Big) = \frac{1 - r + r\log r}{d} =: \delta.$$

2. THIN TAILS. Since the service rate satisfies $r < 1$, if L is too small and $Y = 0$, the content would always increase. To prevent circumstances like this, it is desirable to assume

$$P[L > \frac{1}{d}] > 0.$$

In this case, $\{X_n\}$ possesses a stationary distribution π and

$$\pi(x, \infty) \leq c_1 e^{-c_2 x},$$

for sufficiently large x. Thus, even if we assume heavy tailed file sizes, the control keeps buffer content moderate and the stationary distribution tail is exponentially bounded. Since buffer content is thin tailed, we have a right to worry about the effect on user transmission times.

3. ERGODICITY. Under mild additional conditions, $\{0\}$ is an atom and $\{X_n\}$ is geometrically ergodic, strongly mixing and Harris recurrent. There exists $0 < \rho < 1$ such that

$$\rho^{-n} \| P_x[X_n \in \cdot] - \pi(\cdot) \|_{\text{TV}} \to 0,$$

where $\| \cdot \|_{\text{TV}}$ denotes the total variation metric.

4. STATIONARY TRANSMISSION TIMES. For any $y \geq 0$

$$P_y[\tau(X_{n-1}, L_n) \in A] \to P_\pi[\tau(X, L) \in A],$$

where X and L are independent random variables with $L \stackrel{d}{=} L_1$ and X has the stationary distribution π.

5. Transmission time tail behavior. If L has a subexponential distribution,

$$P_\pi[\tau(X,L) > x] \sim \bar{F}_L(rx).$$

Thus, the transmission tail is asymptotically equivalent to the file size tail, modulo the factor r for the work rate. Note this expression is **independent of** d**,** the exponential rate parameter. So the exponential decrease parameter does not affect the asymptotic form of the tail. (Whether you halve the rate or divide by 17 does not influence the asymptotic tail of the transmission time distribution.)

Where is the effect of d felt? Based on simulations, it appears to influence $E(\tau(X,L))$ mildly, but we only have bounds on $E(\tau(X,L))$ and the influence of d is difficult to quantify.

10. Conclusion.

Data networks are fascinating to the curious and broadly trained investigator. Their thorough study benefits from techniques taken from probability, statistics, applied mathematics, computer science, electrical and computer engineering. Many studies today have isolated one small part of the network for study and ignored many interactions and dependencies. The challenge for the future is to increase model realism by accounting for interaction among users competing for network resources. Most of the current attempts to realistically take into account competition have been empirical or simulation based.

Heavy tail and extreme value analyses are frought with imprecisions. Modeling assumptions based on asymptotic properties are inherently controversial. The assumption to model with a heavy tail or to incorporate long range dependence into the model may be viewed in the same light as conviction about religion: it works for the user but there may be other models that work for other users. There are usually several models consistent with collected data measurements. Imagine the difficulty of trying to statistically distinguish between a Pareto distribution and a truncated Pareto. It is useful to maintain a focus which suggests that the key question is not *does a particular model fit the data*? but rather *what is the class of models consistent with the data*?

References

[1] P. Abry. *Ondelettes et Turbulences - Multirésolutions, Algorithmes de Décompositions, Invariance d'Echelle et Signaux de Pression*. Diderot, Editeur des Sciences et des Arts, Paris, 1997.

[2] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. In *Self-Similar Network Traffic and Performance Evaluation*, pages 39–88. Wiley, 2000.

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

[4] R.J. Adler. *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*, volume Vol. 12 of *IMS Lectures Notes*. Institute of Mathematical Statistics, Hayward, 1990.

[5] R.J. Adler, R.E. Feldman, and M.S. Taqqu (editors). *A Practical Guide to Heavy Tails. Statistical Techniques and Applications*. Birkhauser, Boston, 1998. xvi, 533 pages.

[6] E. Altman, K. Avratchenkov, and C. Barakat. A stochastic model of TCP/IP with stationary random losses. In *Proceedings of the ACM SIGCOMM*, pages 231–242, 2000.

[7] M. Arlitt and C. Williamson. Web server workload characterization: the search for invariants. Master's thesis, University of Saskatchewan, 1996.

[8] B. Basrak. *The sample autocorrelation function of non-linear time series*. PhD thesis, Rijksuniversiteit Groningen, Groningen, Netherlands, 2000.

[9] B. Basrak, R. Davis, and T. Mikosch. A characterization of multivariate regular variation. Preprint, 2000.

[10] J. Beirlant, P. Vynckier, and J. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, 91(436):1659–1667, 1996.

[11] J. Beran. Statistical methods for data with long-range dependence. *Statistical Science*, 7(4):404–416, 1992. With discussions and rejoinder, pages 404-427.

[12] J. Beran. *Statistical Methods for Long Memory Processes.* Chapman and Hall, London, 1994.
[13] R.N. Bhattacharya, V.K. Gupta, and E. Waymire. The Hurst effect under trends. *J. Applied Probability*, 20:649–662, 1983.
[14] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Cambridge University Press, 1987.
[15] D.C. Boes. Schemes exhibiting Hurst behavior. In J. Srivastava, editor, *Essays in Honor of Franklin Graybill*, pages 21–42, Amsterdam, 1988. North Holland.
[16] D.C. Boes and J. D. Salas-La Cruz. On the expected range and expected adjusted range of partial sums of exchangeable random variables. *J. Applied Probability*, 10:671–677, 1973.
[17] M. Borkovec, A. Dasgupta, S. Resnick, and G. Samorodnitsky. A single channel on/off model with tcp-like control. web available at http://critical.orie.cornell.edu/trlist/trlist.html; to appear: J. Applied Probability;, 2000.
[18] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, second edition, 1991.
[19] P.J. Brockwell, N. Pacheco-Santiago, and S. Resnick. Weak convergence and range analysis for dams with Markovian input rate. *J. Applied Probability*, 19:272–289, 1982.
[20] F. H. Campos, J.S. Marron, S. Resnick, and K. Jaffay. Extremal dependence: Internet traffic applications. Web available at http://www.orie.cornell.edu/∼sid, 2002.
[21] J. Cao, W. Cleveland, D. Lin, and Don X. Sun. The effect of statistical multiplexing on internet packet traffic: Theory and empirical study. http://cm.bell-labs. com/cm/ms/departments /sia/InternetTraffic/ webpapers.html, 2001.
[22] J. Cao, William Cleveland, D. Lin, and Don X. Sun. On the nonstationarity of internet traffic. *Proceedings of the ACM SIGMETRICS '01*, pages 102–112, 2001. http://cm.bell-labs.com /cm/ms/departments/sia/ InternetTraffic/webpapers.html.
[23] P. Carlsson and M. Fiedler. Multifractal products of stochastic processes: fluid flow analysis. In *Proceedings of 15th Nordic Teletraffic Seminar NTS-15, August 22-24, 2000; Lund, Sweden*, pages 173–184, 2000.
[24] E. Castillo. *Extreme Value Theory in Engineering.* Academic Press, San Diego, California, 1988.
[25] J. Cohen, S. Resnick, and G. Samorodnitsky. Sample correlations of infinite variance time series models: an empirical and theoretical study. *J. Appl. Math. Stochastic Anal.*, 11(3):255–282, 1998.
[26] A.G. Constantine and P. Hall. Characterizing surface smoothness via estimation of effective fractal dimension. *J.R. Statist. Soc. B*, 56:97–113, 1994.
[27] D.R. Cox. Long-range dependence: a review. In H.A. David and H.T. David, editors, *Statistics: An Appraisal*, pages 55–74. Iowa State University Press, 1984.
[28] M. Crovella and A. Bestavros. Explaining world wide web traffic self–similarity. Preprint available as TR-95-015 from {crovella,best}@cs.bu.edu, 1995.
[29] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. In *Proceedings of the ACM SIGMETRICS '96 International Conference on Measurement and Modeling of Computer Systems*, volume 24, pages 160–169, 1996.
[30] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *Performance Evaluation Review*, 24:160–169, 1996.
[31] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.
[32] M. Crovella, A. Bestavros, and M.S. Taqqu. Heavy-tailed probability distributions in the world wide web. In M.S. Taqqu R. Adler, R. Feldman, editor, *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions.* Birkhäuser, Boston, 1999.
[33] M. Crovella, G. Kim, and K. Park. On the relationship between file sizes, transport protocols, and self-similar network traffic. In *Proceedings of the Fourth International Conference on Network Protocols (ICNP'96)*, pages 171–180, 1996.
[34] S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates for the tail index of a distribution. *Ann. Statist.*, 13:1050–1077, 1985.
[35] S. Csörgő and D. Mason. Central limit theorems for sums of extreme values. *Math. Proc. Camb. Phil. Soc.*, 98:547–558, 1985.
[36] C. Cunha, A. Bestavros, and M. Crovella. Characteristics of www client–based traces. Preprint available as BU-CS-95-010 from {crovella,best}@cs.bu.edu, 1995.
[37] R.A. Davis and T. Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statistics.*, 26(5):2049–2080, 1998.

[38] R.A. Davis and S.I. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probability*, 13(1):179–195, 1985.
[39] R.A. Davis and S.I. Resnick. More limit theory for the sample correlation function of moving averages. *Stochastic Processes and their Applications*, 20:257–279, 1985.
[40] R.A. Davis and S.I. Resnick. Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statistics*, 14(2):533–558, 1986.
[41] R.A. Davis and S.I. Resnick. Limit theory for bilinear processes with heavy tailed noise. *Ann. Applied Probability*, 6:1191–1210, 1996.
[42] L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematisch Centrum Amsterdam, 1970.
[43] L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–46, 1998.
[44] L. de Haan and L. Peng. Comparison of tail index estimators. *Statist. Neerlandica*, 52(1):60–70, 1998.
[45] L. de Haan and S.I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie*, 40:317–337, 1977.
[46] L. de Haan and S.I. Resnick. Estimating the limit distribution of multivariate extremes. *Stochastic Models*, 9(2):275–309, 1993.
[47] L. de Haan and S.I. Resnick. On asymptotic normality of the Hill estimator. *Stochastic Models*, 14:849–867, 1998.
[48] L. de Haan and H. Rootzén. On the estimation of high quantiles. *J. Statist. Plann. Inference*, 35(1):1–13, 1993.
[49] A.L.M. Dekkers and L. de Haan. On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, 17:1795–1832, 1989.
[50] A.L.M. Dekkers and L. de Haan. Optimal choice of sample fraction in extreme-value estimation. *Journal of Mult. Anal.*, 1993.
[51] A.L.M. Dekkers, J.H.J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *Ann. Stat.*, 17:1833–1855, 1989.
[52] H. Drees, L. de Haan, and S. Resnick. How to make a Hill plot. *Ann. Statistics*, 28(1):254–274, 2000.
[53] N.G. Duffield, J.T. Lewis, N. O'Connell, R. Russell, and F. Toomey. Statistical issues raised by the bellcore data. In *Proceedings of 11th IEE UK Teletraffic Symposium, Cambridge UK*, pages 23–25, London, 1994. IEE.
[54] D.E. Duffy, A.A. McIntosh, M. Rosenstein, and W. Willinger. Analyzing telecommunications traffic data from working common channel signaling subnetworks. In M.E. Tarter and M.D. Lock, editors, *Computing Science and Statistics Interface, Proceedings of the 25th Symposium on the Interface*, volume 25, pages 156– 165, San Diego, California, 1993.
[55] R. Durrett and S.I. Resnick. Weak convergence with random indices. *Stochastic Processes and Their Applications*, 5:213–220, 1977.
[56] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extreme Events for Insurance and Finance.* Springer-Verlag, Berlin, 1997.
[57] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* John Wiley & Sons Ltd., Chichester, 1990.
[58] P. Feigin and S. Resnick. Linear programming estimators and bootstrapping for heavy tailed phenomena. *Advances in Appl. Probability*, 29:759–805, 1997.
[59] P. Feigin and S.I. Resnick. Estimation for autoregressive processes with positive innovations. *Stochastic Models*, 8:479–498, 1992.
[60] P. Feigin and S.I. Resnick. Limit distributions for linear programming time series estimators. *Stochastic Processes and their Applications*, 51:135–165, 1994.
[61] P. Feigin and S.I. Resnick. Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes*, 1:391–422, 1999.
[62] A. Feldmann, A.C. Gilbert, and W. Willinger. Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic. In *Proceedings of the ACM Sigcomm '98*, pages 25–38, Vancouver, B.C., 1998.
[63] W. Feller. The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Statistics*, 22:427–432, 1951.
[64] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.
[65] M.W. Garrett and W. Willinger. Analysis, modeling and generation of self similar vbr video traffic. In *Proceedings of the ACM SigComm*, London, 1994.

[66] J. Geluk, L. de Haan, S. Resnick, and C. Stărică. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, 69(2):139–159, 1997.
[67] A. Gilbert, Y. Joo, and N. McKeown. Congestion control and periodic behavior. Preprint: AT&T Labs–Research, 180 Park Ave, Florham Park, NJ 07932-0971, 2000.
[68] A.C. Gilbert, W. Willinger, and A. Feldmann. Scaling analysis of conservative cascades, with applications to network traffic. *IEEE Transactions on Information Theory*, 45(3):971–991, 1999.
[69] L. Giraitis, P. Kokoszka, and R. Leipus. Testing for long memory in the presence of a general trend. *J. Applied Probability*, 38:1033–1054, 2001.
[70] L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssiere. Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity. *Statistical Inference for Stochastic Processes*, 3:113–128, 2000.
[71] E.G. Gladyshev. A new limit theorem for processes with Gaussian increments. *Theory Probab. Appl.*, 6:52–61, 1961.
[72] C.A. Guerin, H. Nyberg, O. Perrin, S. Resnick, H. Rootzen, and C. Stărică. Empirical testing of the infinite source poisson data traffic model. Technical Report 1257, School of ORIE, Cornell University, Ithaca NY 14853; available at www.orie.cornell.edu/∼sid; to appear: Stochastic Models, 1999.
[73] P. Hall. On some simple estimates of an exponent of regular variation. *J. Roy. Stat. Assoc.*, 44:37–42, 1982. Series B.
[74] D. Heath, S. Resnick, and G. Samorodnitsky. Patterns of buffer overflow in a class of queues with long memory in the input stream. *Ann. Applied Probability*, 7(4):1021–1057, 1997.
[75] D. Heath, S. Resnick, and G. Samorodnitsky. Heavy tails and long range dependence in on/off processes and associated fluid models. *Math. Oper. Res.*, 23(1):145–165, 1998.
[76] D. Heath, S. Resnick, and G. Samorodnitsky. How system performance is affected by the interplay of averages in a fluid queue with long range dependence induced by heavy tails. *Ann. Appl. Probab.*, 9:352–375, 1999.
[77] D. Heyman and T.V. Lakshman. What are the implications of long-range dependence for vbr-video traffic engineering? *IEEE-ACM Transactions on Networking*, 4(3):301–317, 1996.
[78] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.
[79] T. Hsing. On tail estimation using dependent data. *Ann. Statist.*, 19:1547–1569, 1991.
[80] H.E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808, 1951.
[81] H.E. Hurst. Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers, Part I*, pages 519–577, 1955.
[82] I.A. Ibragimov and Y.A. Rozanov. *Gaussian Random Processes*. Springer-Verlag, New York, 1978. Translated from the Russian by A. B. Aries.
[83] J. Istas. Wavelet coefficients of a Gaussian process and applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(4):537–556, 1992.
[84] J. Istas and G. Lang. Variations quadratiques et estimation de l'exposant de Hölder local d'un processus gaussien. *C. R. Acad. Sci. Paris Sér. I Math.*, 319(2):201–206, 1994.
[85] J. Istas and G. Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(4):407–436, 1997.
[86] J. Istas and C. Laredo. Estimation de la fonction de covariance d'un processus gaussien stationnaire par méthodes d'échelles. *C. R. Acad. Sci. Paris Sér. I Math.*, 316(5):495–498, 1993.
[87] S. Jaffard. The multifractal nature of Lévy processes. *Probab. Theory Related Fields*, 114(2):207–227, 1999.
[88] P. Jelenković and A. Lazar. Subexponential asymptotics of a Markov-modulated random walk with queueing applications. *J. Applied Probability*, 35(2):325–347, 1998.
[89] P. Jelenković and A. Lazar. Asymptotic results for multiplexing subexponential on-off processes. *Advances in Applied Probability*, 31:394–421, 1999.
[90] P. Jelenković and A.A. Lazar. A network multiplexer with multiple time scale and subexponential arrivals. In Glasserman P., K. Sigman, and D.D. Yao, editors, *Stochastic Networks*, volume 117 of *Lecture Notes in Statistics*, pages 215–235. Springer, New York, 1996.
[91] J. Kent and A.T.A. Wood. Estimating the fractal dimension of a locally self-similar gaussian process by using increments. *J.R. Statist. Soc. B*, 59(3):679–699, 1997.
[92] S. Keshav. *An Engineering Approach to Computer Networking; ATM Networks, the Internet, and the Telephone network*. Addison-Wesley, Reading, Mass., 1997.

[93] T. Konstantopoulos and S.J. Lin. Macroscopic models for long-range dependent network traffic. *Queueing Systems Theory Appl.*, 28(1-3):215–243, 1998.
[94] M. Kratz and S. Resnick. The qq–estimator and heavy tails. *Stochastic Models*, 12:699–724, 1996.
[95] H. Künsch. Discrimination between monotonic trends and long range dependence. *J. Applied Probability*, 23:1025–1030, 1986.
[96] J.W. Lamperti. Semi-stable stochastic processes. *Transaction of the American Mathematical Society*, 104:62–78, 1962.
[97] J.W. Lamperti. Semi-stable Markov processes i. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 22:205–225, 1972.
[98] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, New York, 1983.
[99] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic. *ACM/SIGCOMM Computer Communications Review*, pages 183–193, 1993.
[100] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. Statistical analysis of high time-resolution ethernet Lan traffic measurements. In *Proceedings of the 25th Symposium on the Interface between Statistics and Computer Science*, pages 146– 155, 1993.
[101] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1–15, 1994.
[102] J. Levy and M. Taqqu. Renewal reward processes with heavy-tailed interrenewal times and heavy-tailed rewards. *Bernoulli*, 6:23–44, 2000.
[103] J. Levy and M.S. Taqqu. On renewal processes having stable inter-renewal intervals and stable rewards. *Les Annales des Sciences Mathematiques du Quebec*, 11:95–110, 1987.
[104] B.B. Mandelbrot and M.S. Taqqu. Robust R/S analysis of long-run serial correlation. In *Proceedings of the 42nd Session of the International Statistical Institute*, Manila, 1979. Bulletin of the I.S.I. Vol. 48, Book 2, pp. 69-104.
[105] P. Mannersalo, I. Norros, and R. Riedi. Multifractal products of stochastic processes: a preview. Technical Document COST257TD(99)31, September, 1999; available at http://www.vtt.fi/tte/staff2/petteri/research.html, 1999.
[106] D. Mason. Laws of large numbers for sums of extreme values. *Ann. Probability*, 10:754–764, 1982.
[107] K. Maulik and S. Resnick. Small and large time scale analysis of a network traffic model. Technical report, available at www.orie.cornell.edu/trlist/trlist.html, 2001.
[108] K. Maulik, S. Resnick, and H. Rootzén . A network traffic model with random transmission rate. Technical report, web available at www.orie.cornell.edu/∼sid; To appear: J. Applied Probability, 2003.
[109] T. Mikosch, S. Resnick, H. Rootzén, and A.W. Stegeman. Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Applied Probability*, 12(1):23–68, 2002.
[110] T. Mikosch and A. Stegeman. The interplay between heavy tails and rates in self-similar network traffic. Technical Report. University of Groningen, Department of Mathematics, Available via www.cs.rug.nl/ eke/iwi/preprints/, 1999.
[111] T. Mikosch and C. Stărică. Change of structure in financial data. Available at: http://www.math.ku.dk/ mikosch/preprint.html, 1999.
[112] T. Mikosch and C. Stărică. Long range dependence effects and ARCH modeling. Available at: http://www.math.ku.dk/ mikosch/preprint.html, 2000.
[113] P.A.P. Moran. On the range of cumulative sums. *Ann. Inst. Stat. Math.*, 16:109–112, 1964.
[114] G.L. O'Brien and W. Vervaat. Marginal distributions of self-similar processes with stationary increments. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 64:129–138, 1983.
[115] G.L. O'Brien and W. Vervaat. Self-similar processes with stationary increments generated by point processes. *Ann. Probility*, 13:28–52, 1985.
[116] T.J. Ott, J.H.B. Kemperman, and M. Mathis. The stationary behavior of ideal tcp congestion avoidance. In *Proceedings of IEEE INFOCOM'99*, New York, 1999.
[117] T.J. Ott and A. Misra. The window distribution of idealized tcp congestion avoidance with variable packet loss. In *INFOCOM (3)*, pages 1564–1572, 2000.
[118] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP throughput: a simple model and its empirical validation. In *ACM SIGCOMM '98 Conference on Applications, Technologies, Achitectures, and Protocols for Computer Communication*, pages 303–314, Vancouver, Candad, 1998. citeseer.nj.nec.com/padhye98modeling.html.

[119] K. Park and W. Willinger. Self-similar network traffic: An overview. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*. Wiley-Interscience, New York, 2000.
[120] M. Parulekar and A.M. Makowski. Tail probabilities for a multiplexer with a self-similar traffic. *Proceedings of the 15th Annual IEEE INFOCOM*, pages 1452–1459, 1996.
[121] M. Parulekar and A.M. Makowski. Tail probabilities for M/G/∞ input process (i): Preliminary asymptotics. University of Maryland Institute for Systems Research Technical Report TR 96-41, 1996.
[122] L. Peng. *Second Order Condition and Extreme Value Theory*. PhD thesis, Tinbergen Institute, Erasmus University, Rotterdam, 1998.
[123] V. Pipiras and M. Taqqu. The limit of a renewal-reward process with heavy-tailed rewards is not a linear fractional stable motion. *Bernoulli*, 6:607–614, 2000.
[124] R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel, second edition, 2001.
[125] S. Resnick. Heavy tail modeling and teletraffic data. *Ann. Statist.*, 25:1805–1869, 1997.
[126] S. Resnick. Why non-linearities can ruin the heavy tailed modeler's day. In M.S. Taqqu R. Adler, R. Feldman, editor, *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*, pages 219–240. Birkhäuser, Boston, 1998.
[127] S. Resnick and P. Greenwood. A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis*, 9:206–221, 1979.
[128] S. Resnick and G. Samorodnitsky. A heavy traffic approximation for workload processes with heavy tailed service requirements. *Management Science*, 46:1236–1248, 2000.
[129] S. Resnick and G. Samorodnitsky. Limits of on/off hierarchical product models for data transmission. Technical Report 1281 available at www.orie.cornell.edu/trlist/trlist.html, 2001.
[130] S. Resnick, G. Samorodnitsky, and F. Xue. How misleading can sample acf's of stable ma's be? (Very!). *Ann.Applied Probability*, 9(3):797–817, 1999.
[131] S. Resnick, G. Samorodnitsky, and F. Xue. Growth rates of sample covariances of stationary symmetric α-stable processes associated with null recurrent markov chains. *Stochastic Processes and Their Applications*, 85:321–339, 2000.
[132] S. Resnick and C. Stărică. Consistency of Hill's estimator for dependent data. *J. Applied Probability*, 32(1):139–167, 1995.
[133] S. Resnick and C. Stărică. Asymptotic behavior of Hill's estimator for autoregressive data. *Stochastic Models*, 13:703–723, 1997.
[134] S. Resnick and C. Stărică. Smoothing the Hill estimator. *Adv. Applied Probability*, 29:271–293, 1997.
[135] S. Resnick and E. van den Berg. Sample correlation behavior for the heavy tailed general bilinear process. *Stochastic Models*, 16(2):233–258, 2000.
[136] S. Resnick and E. van den Berg. A test for nonlinearity of time series with infinite variance. *Extremes*, 3(4):145–172, 2000.
[137] S. Resnick and E. van den Berg. Weak convergence of high-speed network traffic models. *J. Applied Probability*, 37(2):575–597, 2000.
[138] S.I. Resnick. Point processes, regular variation and weak convergence. *Adv. Applied Probability*, 18:66–138, 1986.
[139] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.
[140] S.I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.
[141] S.I. Resnick. *A Probability Path*. Birkhäuser, Boston, 1998.
[142] S.I. Resnick and H. Rootzén. Self-similar communication models and very heavy tails. *Ann. Applied Probability*, 10:753–778, 2000.
[143] R. H. Riedi. Multifractal processes. Technical Report, ECE Dept., Rice University, TR 99-06, 1999.
[144] R. H. Riedi and J. Levy-Vehel. TCP traffic is multifractal: A numerical study. Preprint, 1997.
[145] R. H. Riedi and W. Willinger. Toward an improved understanding of network traffic dynamics. In *Self-Similar Network Traffic and Performance Evaluation*. Wiley, 2000.
[146] H. Rootzén, M.R. Leadbetter, and L. de Haan. Tail and quantile estimation for strongly mixing stationary sequences. Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, 1990.
[147] H. Rootzén, M.R. Leadbetter, and L. de Haan. On the distribution of tail array sums for strongly mixing stationary sequences. *Ann. Applied Probability*, 8(3):868–885, 1998.
[148] J.D. Salas and D. Boes. Expected range and adjusted range of hydrologic sequences. *Water Resoures Research*, 10(3):457–463, 1974.

[149] J.D. Salas and D. Boes. Nonstationarity of mean and Hurst phenomenon. *Water Resources Research*, 14((1)):135–143, 1978.

[150] J.D. Salas, D. Boes, and V. Yevjevich. Hurst phenomenon as a pre-asymptotic behavior. *J. Hydrology*, 44((1-2)):1–15, 1979.

[151] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling. New York, NY: Chapman & Hall. xviii, 632, 1994.

[152] R. Srikant. Control of communication networks. In Tariq Samad, editor, *Perspectives in Control Engineering: Technologies, Applications, New Directions*, pages 462–488. IEEE Press, 2000.

[153] A. Stegeman. Modeling traffic in high-speed networks by on/off models. Master's thesis, University of Groningen, Department of Mathematics, 1998.

[154] A. Stegeman. Non-stationarity versus long-range dependence in computer network traffic measurements. Technical report, University of Groningen, Department of Mathematics, P.O. Box 800, 9700 AV Groningen, The Netherlands, 2002. available at http://www.math.rug.nl/ stegeman/.

[155] C. Stărică. Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance*, 6:515–553, 1999.

[156] M.S. Taqqu and J. Levy. Using renewal processes to generate long-range dependence and high variability. In E. Eberlein and M.S. Taqqu, editors, *Dependence in Probability and Statistics*, pages 73–89, Boston, 1986. Birkhäuser.

[157] M.S. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. *Computer Communications Review*, 27:5–23, 1997.

[158] D. Veitch and P. Abry. Estimation conjointe en ondelettes des paramètres du phénomène de dépendance longue. In *Proc. 16ième Colloque GRETSI, Grenoble, France*, pages 1451–1454, 1997.

[159] D. Veitch and P. Abry. A wavelet based joint estimator for the parameters of LRD. *IEEE Trans. Info. Th.*, 1998. special issue "Multiscale Statistical Signal Analysis and Its Application".

[160] W. Vervaat. Sample paths of self-similar processes with stationary increments. *Ann. Probabability*, 13:1–27, 1985.

[161] W. Vervaat. Properties of general self-similar processes. *Bulletin of the International Statistical Institute*, 52(Book 4):199–216, 1987.

[162] W. Whitt. *Stochastic Processs Limits: An Introduction to Stochastic-Process Limits And their Application to Queues*. Springer-Verlag, New York, 2002.

[163] W. Willinger and V. Paxson. Where mathematics meets the Internet. *Notices of the American Mathematical Society*, 45(8):961–970, 1998.

[164] W. Willinger, M.S. Taqqu, M. Leland, and D. Wilson. Self–similarity through high variability: statistical analysis of ethernet lan traffic at the source level. *Computer Communications Review*, 25:100–113, 1995. Proceedings of the ACM/SIGCOMM'95, Cambridge, MA.

[165] W. Willinger, M.S. Taqqu, M. Leland, and D. Wilson. Self–similarity through high variability: statistical analysis of ethernet lan traffic at the source level (extended version). *IEEE/ACM Transactions on Networking*, 5(1):71–96, 1997.

Sidney Resnick, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853 USA

E-mail address: sid@orie.cornell.edu, http://www.orie.cornell.edu/∼sid