Estimation of the long Memory parameter using an Infinite Source Poisson model applied to transmission rate measurements

François Roueff

Ecole Nat. Sup. des Télécommunications
46 rue Barrault, 75634 Paris cedex 13, France
http://www.tsi.enst.fr/~roueff/

August 13, 2005

Joint work with
Gilles Faÿ (UST de Lille) and Philippe Soulier (Univ. Paris X)

Outline

Second order long memory

A second-order stationary process $X(t)$ has long memory parameter (or Hurst index) $H \in (1/2, 1)$ if

$$\mathrm{cov}(X(0), X(t)) = \ell(t)\, t^{2H-2}$$

for some ℓ slowly varying at infinity. A possible extension to $H \geq 1$ (X is non-stationary), is

$$\mathrm{var}\left(\int_0^t X(s)\, \mathrm{d}s\right) = L(t)\, t^{2H}$$

for some L slowly varying at infinity.

Examples

1. Linear processes :
   1.1 FGN (Gaussian, [Mandelbrot and Van Ness(1968)]),
   1.2 ARFIMA ([Granger and Joyeux(1980)]);
2. Non-linear processes :
   2.1 Shot noise ([Giraitis et al.(1993)]),
   2.2 Renewal-reward ([Taqqu and Levy(1986)]),
   2.3 On-Off sources ([Taqqu et al.(1997)]),
   2.4 Infinite Source Poisson ([Mikosch et al.(2002)], [Maulik et al.(2002)]).

Linear VS Non-linear models

Remark 1

Non-linear Long memory models are often derived from Point processes : appropriate for traffic models.

Linear VS Non-linear models

Remark 1

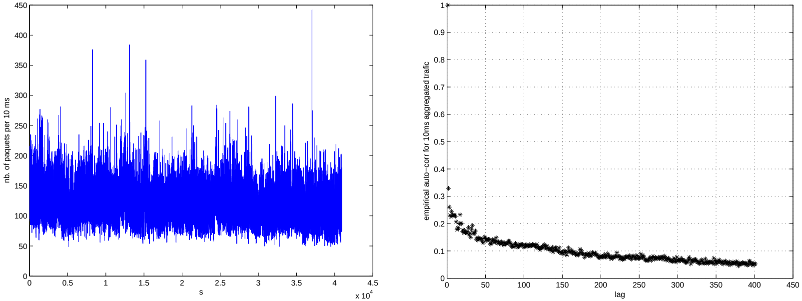
Non-linear Long memory models are often derived from Point processes : appropriate for traffic models.

Remark 2

Little is known about the estimation of H in the non-linear case. Here we investigate the non-linear Poisson case.

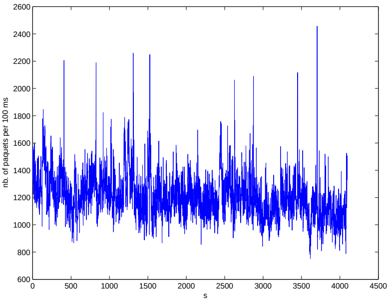
An example : traffic measurements

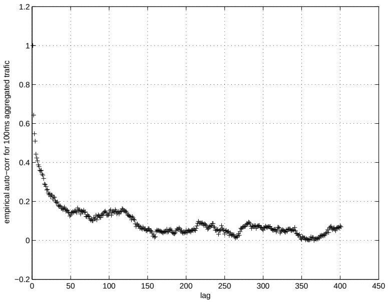
Here is a 10ms aggregated traffic (obtained form packet counts through some point of the Internet network)



Same example (aggregated again)

The same 100ms aggregated traffic (obtained form packet counts through some point of the Internet network)





Infinite Source Poisson model

1. $\{S_n\}_{n\in\mathbb{N}}$: points of a unit rate homogeneous Poisson process;
2. $\{(U_n, \eta_n)\}$: i.i.d. independent of $\{S_n\}$;

U_n are referred to as transmission rates,
η_n are the flows durations,
U_n and η_n may be dependent.
The n^{th} flow starts at time S_n, has rate U_n and is transmitting for a duration η_n. We observe the cumulative rate

$$\boxed{X(t) := \sum_{n\in\mathbb{N}} U_n \, \mathbf{1}_{[S_n, S_n+\eta_n)}(t). \text{ for } t \in [0, T]}$$

If $\mathrm{E}\eta < \infty$, this process has a stationary version defined by

$$\boxed{X_S(t) := \sum_{n\in\mathbb{Z}} U_n \, \mathbf{1}_{[S_n, S_n+\eta_n)}(t).}$$

Second order properties

If $\boxed{\mathrm{E}[U^2] < \infty,}$ then, for all $s \geq 0$, $\mathsf{E}[X^2(s)] < \infty$ and for $s \leq t$,

$$\mathrm{cov}(X(s), X(t)) = \mathrm{E}[U^2\{s - (t - \eta)_+\}_+] .$$

The stationary version is weakly stationary if

$$\boxed{\mathsf{E}[U^2\,\eta] < \infty.}$$

Then

$$\mathrm{cov}(X_S(0), X_S(t)) = \mathrm{E}[U^2\,(\eta - t)_+] .$$

Long memory is a consequence of heavy tails

Assume $\mathrm{E}[U^2] < \infty$ and define

$$\mathcal{H}(t) = \mathrm{E}\left[U^2 \mathbf{1}_{\{\eta > t\}}\right] .$$

Assumption
$\mathcal{H}$ is regularly varying with index $\alpha \in (0,2)$:

$$\mathcal{H}(t) = \ell(t)\ t^{-\alpha}\ ,\text{where } \ell \text{ is slowly varying.}$$

In words, durations are **heavy tailed** with non-necessarily independent rates.

Examples

Example 1

U and η are independent and η is heavy tailed, e.g. the **M/G/**∞ queue, when $U = 1$.

Examples

Example 1

U and η are independent and η is heavy tailed, e.g. the **M/G/**∞ queue, when $U = 1$.

Example 2

In a more general Internet traffic framework : the workload $U \times \eta$ of the flow is heavy tailed and independent of the transmission rate U (and the latter stays away from zero).

Consequence on the weak dependence of X

Non-stationary case, $\alpha \in (0,2)$

$$\mathrm{var}\left(\int_0^T X(s)\,\mathrm{d}s\right) \sim c_\alpha\ \ell(T)\,T^{2H} \text{ as } T \to \infty$$

Stationary case, $\mathrm{E}[\eta] < \infty$

$$\mathrm{cov}(X_S(0), X_S(t)) \sim \frac{1}{\alpha - 1}\ \ell(t)\ t^{1-\alpha} = \frac{1}{2-2H}\ \ell(t)\ t^{2H-2}.$$

Consequence on the dependence structure of X

The processes X and X_S have Hurst index

$$\boxed{H = (3 - \alpha)/2}$$

Remark

$$H < 1 \Leftrightarrow \alpha \in (1, 2)\ ; \quad H \geq 1 \Leftrightarrow \alpha \in (0, 1].$$

Remark

$$\alpha \in (1, 2) \Rightarrow \mathrm{E}[\eta] < \infty \Rightarrow \alpha \in [1, 2)$$

Asymptotic properties

The asymptotic behavior of

$$Y(t) := \int_0^t (X(s) - \mathrm{E}X(s))\, \mathrm{d}s$$

conveniently renormalized has been studied in various situations by [Mikosch et al.(2002)], [Maulik et al.(2002)] or [Mikosch and Resnick(2004)].

Two very different cases

Let $T \to \infty$.

Stable case : $1 < \alpha < 2$, *i.e.* $1/2 < H < 1$

$T^{-1/\alpha} Y(Tt)$ converges weakly to an α-stable Levy process.

Two very different cases

Let $T \to \infty$.

Stable case : $1 < \alpha < 2$, *i.e.* $1/2 < H < 1$

$T^{-1/\alpha} Y(Tt)$ converges weakly to an α-stable Levy process.

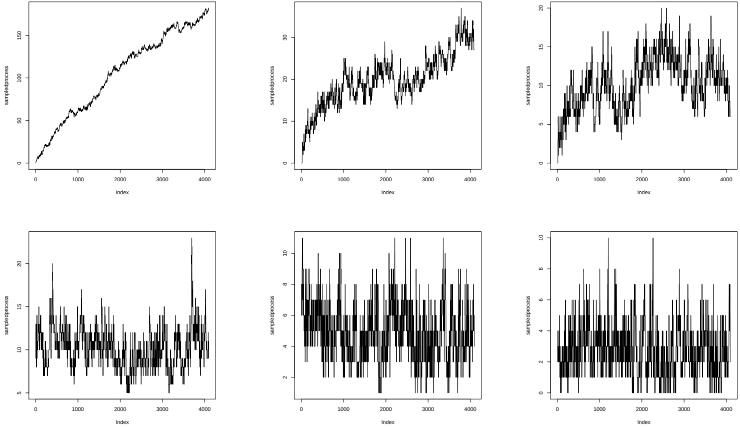
Unstable case : $0 < \alpha < 1$, *i.e.* $1 < H < 3/2$

$\mathcal{H}^{-1/2}(T) T^{-H} Y(Tt)$ converges weakly to the Gaussian process W with auto-covariance function

$$\mathrm{cov}(W(s), W(t)) = \frac{1}{1-\alpha} \int_0^t \int_0^s \{(u \vee v)^{1-\alpha} - |u-v|^{1-\alpha}\} \, \mathrm{d}u \, \mathrm{d}v \, .$$

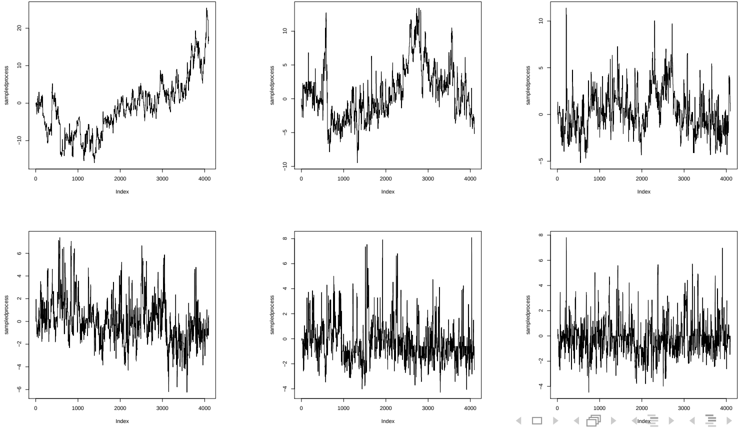
M/G/∞ queue

$\alpha = 0.3,\ 0.7,\ 0.9,\ 1.1,\ 1.4$ and 1.7.



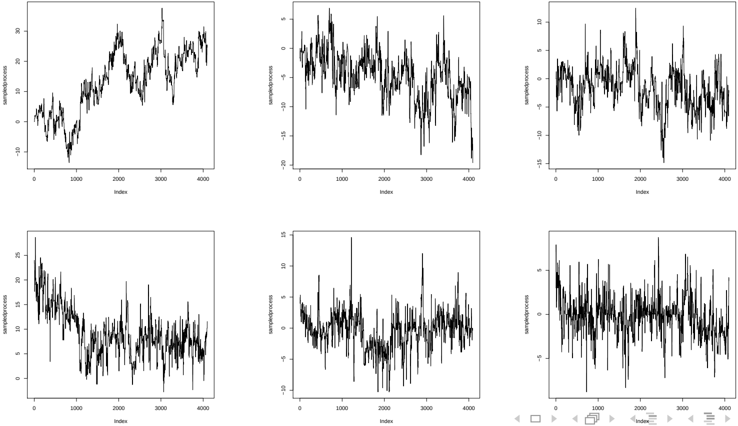
Centered exponential rewards

Here are sample paths for $1 + U_n \sim \exp(1)$ and $\eta_n \sim \mathrm{Pareto}(\alpha)$ with $\alpha = 0.3$, 0.7, 0.9, 1.1, 1.4 and 1.7.



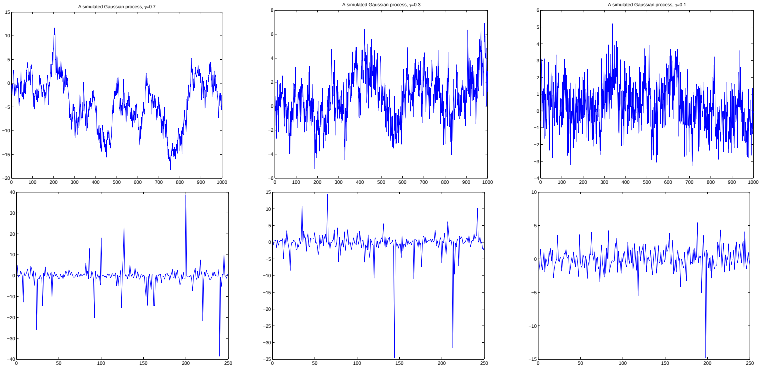
Symmeterized exponential rewards

Here are sample paths for $U_n \sim \exp(1) - \exp(1)'$ and $\eta_n \sim \mathrm{Pareto}(\alpha)$ with $\alpha = 0.3,\ 0.7,\ 0.9,\ 1.1,\ 1.4$ and 1.7.



Differentiated asymptotic sample paths

Gaussian (top) and Lévy-stable (bottom) increments corresponding to $\alpha = 0.3, 0.7, 0.9, 1.1, 1.4$ and 1.7.



Three possible observation schemes

1. We observe the **continuous** path $X(t)$ for all $t \in [0, T]$.
2. We observe the **discrete** sample path $X(t)$ for all $t = 1, 2, \dots, T$.
3. We observe **discrete** local averages

$$Y_k = \int_k^{k+1} X(t)\,\mathrm{d}t$$

for all $k = 0, 1, \dots, T$.

In fact the two last cases can be treated similarly. Hence we only consider the continuous case and the discrete sample case.

α is the parameter of interest

$\Rightarrow$ asymptotic behavior, stability, (queuing performances ?) ....

Natural approach : heavy tail estimator such as the Hill estimator.

α is the parameter of interest

$\Rightarrow$ asymptotic behavior, stability, (queuing performances ?) ....

Natural approach : heavy tail estimator such as the Hill estimator.

Drawbacks :

Hidden information

Difficult and costly to observe, say, $(U_n, \eta_n)_n$ (one needs to reconstruct the flows),

α is the parameter of interest

$\Rightarrow$ asymptotic behavior, stability, (queuing performances ?) ....

Natural approach : heavy tail estimator such as the Hill estimator.

Drawbacks :

Hidden information

Difficult and costly to observe, say, $(U_n, \eta_n)_n$ (one needs to reconstruct the flows),

Tails of X

In general the marginals of $X(t)$ are not even heavy tailed,

α is the parameter of interest

$\Rightarrow$ asymptotic behavior, stability, (queuing performances ?) ....

Natural approach : heavy tail estimator such as the Hill estimator.

Drawbacks :

Hidden information

Difficult and costly to observe, say, $(U_n, \eta_n)_n$ (one needs to reconstruct the flows),

Tails of X

In general the marginals of $X(t)$ are not even heavy tailed,

Return times to the empty state

In the stable case **only**, there are iid heavy tailed active and exponential idle periods. But real data indicate that the model is not suitable at fine scales : the empty state is hard to identify.

Here we estimate α through the long memory parameter H based on empirical second-order properties of the path. Standard approaches are

Fourier methods (GPH, GSE)

Efficient in practice and in theory for standard time series. Non-linear case is open.

Wavelet methods

Promising results in the context of self-similar Gaussian and stable processes. Easier to adapt in the non-linear case ?

Mother and father wavelets

Take a function ψ with compact support and such that $\int \psi = 0$.
Take a function ϕ with compact support and satisfying

$$\sum_{k \in \mathbb{Z}} \phi(t-k) = 1, \quad t \in \mathbb{R} .$$

Denote

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} t - k) .$$

In the context of a multi-resolution analysis, one calls ψ a **wavelet** and ϕ is the corresponding scaling function (or *father wavelet*). Additional assumptions are possibly required on ϕ and ψ but will not be mentioned in the following.

Continuous or discrete time Wavelet coefficients

In continuous time, one computes

$$d_{j,k} = \int X(t)\psi_{j,k}(t)\,\mathrm{d}t$$

for any (j, k) such that the support of $\psi_{j,k}$ falls into $[0, T]$.
In discrete time, one first compute a continuous interpolation

$$\mathrm{I}_\phi[X](t) := \sum_{k\in\mathbb{Z}} X(k)\phi(t-k)$$

and then (this setting includes the usual (decimated) **DWT**)

$$d^D_{j,k} = \int \mathrm{I}_\phi[X](t)\psi_{j,k}(t)\,\mathrm{d}t\ .$$

The set of indices

Let J denote the largest j such that $d_{j,k}$ or $d^D_{j,k}$ can be computed from $\{X(t), t \in [0, T]\}$ for all $j = 0, \ldots, J$ and $k = 0, \ldots, 2^{J-j}$; One has

$$\boxed{J = \log_2(T) + O(1)} .$$

For positive integers $J_0 < J_1 \leq J$, define

$$\Delta := \{(j, k) \;, J_0 < j \leq J_1 \;, \quad 0 \leq k \leq 2^{J-j} - 1\} \;.$$

Definition of the wavelet Whittle estimator

Adapting the Whittle estimator (Fourier) to the wavelet domain, define

$$\hat{\alpha} := \arg\min_{\alpha'} \log\left(\sum_{(j,k)\in\Delta} \frac{d_{j,k}^2}{2^{(2-\alpha')j}}\right) - \delta \log(2)(2-\alpha') ,$$

where

$$\delta = \frac{1}{\#\Delta} \sum_{(j,k)\in\Delta} j .$$

Consistency

Let $T \to \infty$ with $J_0 \to \infty$ and $J_1 - J_0 \to \infty$.

Theorem (Stable case)

If $\alpha > 1$ and $\limsup J_0/J < 1/\alpha$ then $\hat{\alpha} \xrightarrow{P} \alpha$.

Theorem (Unstable case)

If $\alpha \leq 1$ and $\limsup J_1/J < 1/(2-\alpha)$ then $\hat{\alpha} \xrightarrow{P} \alpha$.

In particular we have consistency for all $\alpha \in (0, 2)$ for

$$\boxed{\limsup J_1/J < 1/2}$$

Proof of consistency

The two main points are to show that

$$\mathrm{E} d_{j,k}^2 \approx c 2^{(2-\alpha)j} \tag{1}$$

for $J_0 < j \leq J_1$ and

$$2^{j-J} \sum_{k=1}^{2^{J-j}} \left(\frac{d_{j,k}^2}{c 2^{(2-\alpha)j}} - 1 \right) \xrightarrow{P} 0 \ , \tag{2}$$

for $j \approx J_0$.
In the case $\alpha > 1$, (2) is true only if $\limsup J_0/J < 1/\alpha$.
In the case $\alpha \leq 1$, (1) is true only if $\limsup J_1/J < 1/(2-\alpha)$.

Rates (Stable case)

Assume that $\alpha > 1$. Rates can be established by assuming that

$$\mathrm{E}[U^2 \cos(u\eta)] = 1 + |u|^\alpha \left(1 + O(|u|^\beta)\right) .$$

Theorem

$\hat{\alpha}$ achieves the rate $T^{-\gamma/(2\gamma+\alpha)}$ for $J_0 = J/(2\gamma + \alpha)$, with :

$\gamma = \beta$ in **continuous time**

and

$\gamma = (2 - \alpha) \wedge \beta$ in **discrete time**.

Comments

Discrete observations

We estimate the zero frequency behavior of the spectral density f of $X_t, t = 1, \ldots, T$, $f(\lambda) \sim |\lambda|^{\alpha-2}$. The limitation in discrete time is due to **aliasing** (can be made low in simulations but presumably not negligible in practice).

Comments

Discrete observations

We estimate the zero frequency behavior of the spectral density f of $X_t, t = 1, \ldots, T$, $f(\lambda) \sim |\lambda|^{\alpha-2}$. The limitation in discrete time is due to **aliasing** (can be made low in simulations but presumably not negligible in practice).

The Pareto case

If η has Pareto distribution, we find $\beta = 2 - \alpha$: same rate in continuous and discrete time.

Comments

Discrete observations

We estimate the zero frequency behavior of the spectral density f of $X_t, t = 1, \ldots, T$, $f(\lambda) \sim |\lambda|^{\alpha-2}$. The limitation in discrete time is due to **aliasing** (can be made low in simulations but presumably not negligible in practice).

The Pareto case

If η has Pareto distribution, we find $\beta = 2 - \alpha$: same rate in continuous and discrete time.

Further work

Other non-linear models;
Alternative estimator, e.g. based on tails of busy periods;
Semiparametric standards: adaptive estimators, Minimax rates ...

Simulations

The estimator is computed on simulated discrete observations $X_1, \ldots, X_n$.

The wavelet Whittle estimator is compared to the usual Fourier domain GPH and GSE estimators (whose theoretical properties are only available for Gaussian and linear processes).

Fourier methods

The periodogram ordinates:

$$I_{n,k} = (2\pi n)^{-1} \left| \sum_{t=1}^{n} X_t \mathrm{e}^{\mathrm{i} t x_k} \right|^2 ,$$

$$\mathbb{E}[I_{n,k}] = c x_k^{\alpha - 2}(1 + o(1)) .$$

The GPH and local Whittle (GSE) estimators.

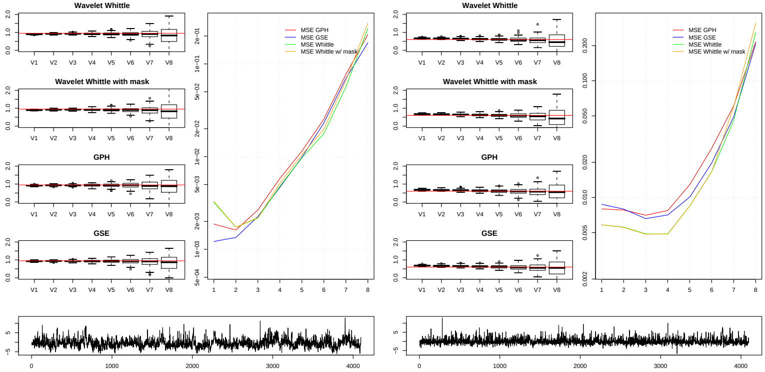
$$\hat{\alpha}_{GPH} = \arg\min_{c',\alpha'} \sum_{j=1}^{m} \left\{ \log(I_{n,k}) - c' - (2-\alpha') \log(k) \right\}^2 ,$$

$$\hat{\alpha}_{GSE} = \arg\min_{c',\alpha'} \log \left(\sum_{j=1}^{m} k^{2-\alpha'} I_{n,k} \right) - (2-\alpha') \frac{1}{m} \sum_{k=1}^{m} \log(k) .$$

m is the bandwidth parameter; plays a role similar to J_0 for the path-wise estimator.

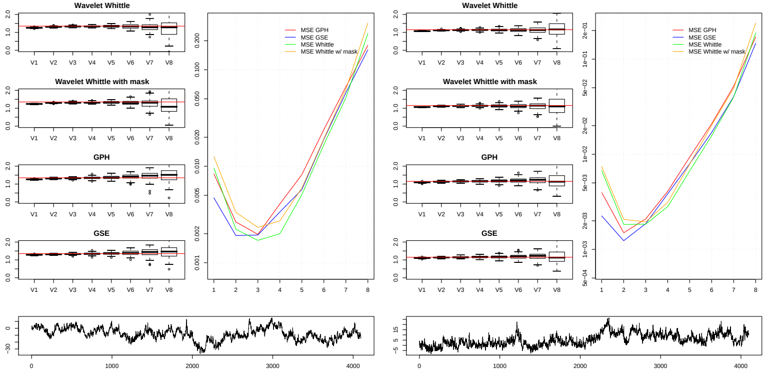
Stable case

100 Monte Carlo simulations with $\alpha = 1.1$ and $\alpha = 1.8$, centered exponential rates.



Unstable case

100 Monte Carlo simulations with $\alpha = 0.3$ and $\alpha = 0.7$, centered exponential rewards.



Further readings I

L. Giraitis, S. A. Molchanov, and D. Surgailis.
Long memory shot noises and limit theorems with application to Burgers' equation.
In *New directions in time series analysis, Part II*, volume 46 of *IMA Vol. Math. Appl.*, pages 153–176. Springer, New York, 1993.

C. W. J. Granger and R. Joyeux.
An introduction to long memory time series and fractional differencing.
Journal of Time Series Analysis, 1:15–30, 1980.

Further readings II

Benoit B. Mandelbrot and John W. Van Ness.
Fractional Brownian motions, fractional noises and applications.
SIAM Rev., 10:422–437, 1968.
ISSN 1095-7200.

Krishanu Maulik, Sidney Resnick, and Holger Rootzén.
Asymptotic independence and a network traffic model.
Journal of Applied Probability, 39(4):671–699, 2002.

Thomas Mikosch and Sidney Resnick.
Activity rates with very heavy tails.
Technical Report 1411, Cornell University, 2004.

Further readings III

Thomas Mikosch, Sidney Resnick, Holger Rootzén, and Alwin Stegeman.
Is network traffic approximated by stable Levy motion or fractional Brownian motion?
Annals of Applied Probability, 12:23–68, 2002.

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

M.S. Taxqqu and J.M. Levy.
Using renewal processes to generate long range dependence and high variability.
In E. Eberlein and M.S. Taqqu (eds), *Dependence in Probability and Statistics*. Boston, Birkhäuser, 1986.