What self-similar processes best describe the input to communication network models?

Vicky Fasen and Gennady Samorodnitsky

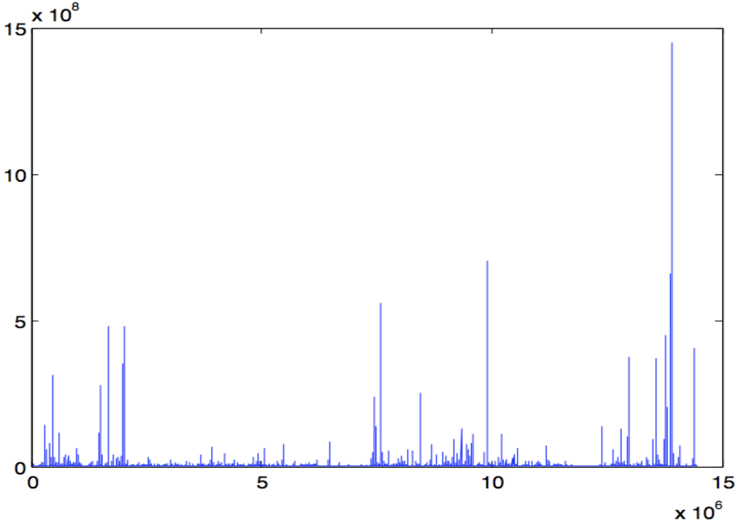
July 2009

It is widely believed that various important objects associated with modern communication networks feature extreme oscillations and irregularity. For example:

- Sizes of files oscillate between very small and huge;
- throughput rates oscillate between very high and almost zero;
- "think times" can be very short and very long.

The terms often mentioned in conjunction with such high variability are "fractals", "long memory" and "heavy tails"; see e.g. Crovella and Bestavros (1996), Willinger et al. (1995), Park and Willinger (2000), Barabasi (2005).

The file sizes downloaded from a server at UNC Chapel Hill, November 2008.



It is widely believed that *the input to a communication network can be well approximated by a self-similar process*.

Question: which self-similar processes best approximate the input to a communication network?

The first step towards answering this question was taken by Mikosch, Resnick, Rootzén and Stegeman (2002), who considered two specific models of the input to a communication system:

- the ON-OFF input model,
- the Infinite Source Poisson input model.

The ON-OFF model

- A cycle consists of an ON-period and an OFF-period;
- the lengths of ON-periods (Z_i) are iid with a common distribution $F_{\mathrm{ON}} \in \mathrm{Reg}(\alpha_{\mathrm{on}})$ and a finite mean μ_{on};
- the lengths of OFF-periods (Y_i) are iid with a common distribution $F_{\mathrm{OFF}} \in \mathrm{Reg}(\alpha_{\mathrm{off}})$ and a finite mean μ_{off};
- the two sequences are independent;
- the work arrives at the unit rate during an ON-period (and at rate 0 during an OFF-period);

Heavy tails in the ON-OFF model cause long range dependence in the sense of slowly decaying correlations.

Consider the stationary input process

$$N(t) = \mathbf{1}\big(t \in \text{ an ON-period}\big),\ \ t \in \mathbb{R}.$$

Assume that $\alpha_{\mathrm{on}} \in (1, 2)$, $\alpha_{\mathrm{off}} > \alpha_{\mathrm{on}}$, and the cycle-length distribution is spread-out. Then

$$R_N(t) := \mathrm{Cov}\big(N(0), N(t)\big)$$

$$\sim \frac{\mu_{\mathrm{off}}^2}{(\alpha_{\mathrm{on}} - 1)\,(\mu_{\mathrm{on}} + \mu_{\mathrm{off}})^3}\, t\,\overline{F}_{\mathrm{ON}}(t)$$

as $t \to \infty$ (Heath, Resnick, Samorodnitsky (1998)).

The Infinite Source Poisson model

- Beginnings of sessions arrive according to a time homogeneous Poisson process with rate λ;
- session durations are iid random variables (X_i) independent of the arrival Poisson process, with a common distribution $F \in \text{Reg}(\alpha)$ and a finite mean μ;
- during the duration of each session, work is generated at the unit rate.

Heavy tails in the Infinite Source Poisson model cause long range dependence in the sense of slowly decaying correlations.

Consider the stationary input process

$$N(t) = \text{ number of sessions running at time } t, \; t \in \mathbb{R}.$$

Assume that $\alpha > 1$. Then

$$R_N(t) := \mathrm{Cov}\big(N(0), N(t)\big)$$

$$\sim \frac{\lambda}{\alpha - 1} \, t \, \overline{F}(t)$$

as $t \to \infty$.

- Let $(N(t),\ t \in \mathbb{R})$ be a stationary process describing the number of sessions running at time t.
- The total amount of work brought into the system by the time $t \geq 0$ is

$$I(t) = \int_0^t N(s)\, ds,\ \ t \geq 0\,.$$

- We think of $(I(t),\ t \in \mathbb{R})$ as the cumulative input process caused by a single "user"; a large number n of "users" contribute to the input.
- Let $(I_j(t),\ t \in \mathbb{R}),\ j = 1, \ldots, n$ be the cumulative input processes corresponding to different users, assumed i.i.d.

Let $T > 0$ be the time scale (large). Consider the deviation from the mean input process at the scale T defined by

$$D_{n,T}(t) = \sum_{j=1}^{n} (I_j(tT) - tT\, EN(0)) , \quad t \geq 0.$$

How does the (properly normalized) deviation process $(D_{n,T}(t),\ t \geq 0)$ **behave as** n, T **grow to infnity**?

Intuitively, the answer to this question depends on the relative rate of growth of the number of "users" n and the time scale T.

Define the fast growth regime and the slow growth regime:

$$\begin{cases} nT\bar{F}_{\mathrm{ON}}(T) \to \infty & \text{the fast growth regime} \\ nT\bar{F}_{\mathrm{ON}}(T) \to 0 & \text{the slow growth regime} \end{cases} .$$

A boundary regime exists as well.

The behaviour of the rescaled cumulative input to the system under the two models has been decsribed by Mikosch, Resnick, Rootzén and Stegeman (2002) for each one of the two regimes.

In the fast growth regime:

$$\left(\frac{1}{(n\bar{F}_{\mathrm{ON}}(T))^{1/2}}D_{n,T},\ t \geq 0\right) \to \left(B_H(t),\ t \geq 0\right)$$

weakly in $\left(D[0,\infty),\ J_1\right)$. Here B_H is a Fractional Brownian motion with $H = (3-\alpha)/2$.

In the slow growth regime:

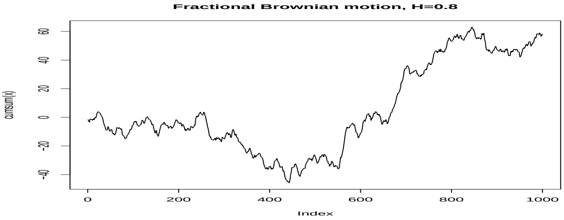
$$\left(\frac{1}{b(nT)}D_{n,T},\ t \geq 0\right) \to \left(L_\alpha(t),\ t \geq 0\right)$$

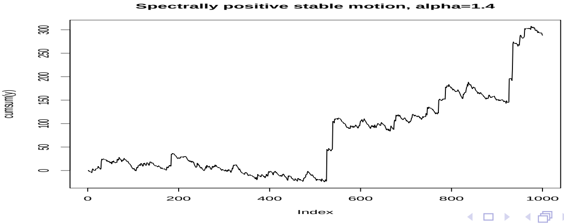
in finite-dimensional distributions. Here $b(t) = \left(1/\bar{F}_{\mathrm{ON}}\right)^{\leftarrow}(t)$ is the left continuous tail inverse.

Two questions:

1. Does the cumulative input to a system usually looks like a Fractional Brownian motion or a Lévy stable motion, or can other self-similar pictures appear?
2. Does a Fractional Brownian motion or a Lévy stable motion appear more frequently as good descriptor of the cumulative input to a system?

In other words: how universal and robust are the Fractional Brownian motion and the Lévy stable motion as limits of the rescaled deviations of the cumulative input to a system from its mean?


Fractional Brownian motion, H=0.8
cumsum(x)
-40
-20
0
20
40
60
0
200
400
600
800
1000
Index


Spectrally positive stable motion, alpha=1.4
cumsum(y)
0
50
100
150
200
250
300
0
200
400
600
800
1000
Index

Suppose that the input to the queue caused a single "user" follows a general stationary marked point process

$$((T_n, Z_n))_{n\in\mathbb{Z}},$$

where

- $\cdots \le T_{-1} \le T_0 \le 0 \le T_1 \le T_2 \le \cdots$ are the starting times for each session;
- the mark $Z_n \in \mathbb{Z}$ is the duration of the session starting at time T_n;
- while each session lasts, the work is added to the queue at the unit rate.

The goal: to understand whether the Fractional Brownian limit, a stable Lévy limit or a different self-similar limit occur, and how "often" under a "relatively general" stationary marked point process input model.

Many of the properties of the input process generated by a stationary marked point process are effectively described by **the Palm measure** of the process and its various **moment measures**.

Consider the function

$$g(x) = \int_{\mathbb{R}^2 \times \mathbb{R}_+^2} I\left\{s_1 \le 0 < s_1 + u_1, s_2 \le x < s_2 + u_2\right\} \gamma_2(ds_1, ds_2, du_1, du_2)\,,$$

where γ_2 is the so-called *covariance measure* of the point process.

It turns out that it is the function g that, mostly, determines, from the bird-eye view, the deviations of the cumulative input from the average in the fast growth situation.

Fast Growth Regime

Theorem 1 (Mikosch and Samorodnitsky (2007))
Assume that the function g is regularly varying at infinity with an exponent $\beta \leq 0$, and that the stationary number of open sessions satisfies $E[|N(0)|^{2+\delta}] < \infty$ for some $\delta > 0$.

- If $\beta \in (-1, 0]$ or $\beta = -1$ and $\int_0^\infty g(x)dx = \infty$ and for some $\delta' < \delta$

$$n \gg T^{|\beta|(1+2/\delta')}$$

then, as $n, T \to \infty$,

$$[n \operatorname{Var}(I(T)]^{-1/2} D_{n,T} \Rightarrow B_H$$

in finite-dimensional distributions as $n \to \infty$, with $H = 1 + \beta/2$.

- Suppose that $\beta < -1$ or $\beta = -1$ and $\int_0^\infty |g(x)|\, dx < \infty$. Let $\int_0^\infty g(x)\, dx \neq 0$. If

$$n \gg T^{1+2/\delta},$$

then the above convergence holds with $H = 0.5$.

Under certain conditions on λ_n, the convergence can be strengthen to the weak convergence in $C[0, \infty)$.

Therefore: a Fractional Brownian limit of the deviation of the cumulative input from its mean is very robust. A regular variation assumption on a function related to the covariance of the input process is sufficient for a Fractional Brownian limit in a fast growth regime.

Slow Growth Regime

Under slow growth regimes, two competing factors can affect the limiting behaviour of the deviation $\left(D_{n,T}(t)\right)$ of the cumulative input process from its mean:

- the deviations of the partial sums of the marks (session durations) from their mean, $\sum_{k=1}^{m} S_k - m\, E_0 S_0$ under the Palm probability;
- the deviations of the arrival numbers from their mean, number of $\left\{k : 0 < T_k \leq T\right\} - \lambda T$, under the stationary probability.

Note: Under the Palm probability the marks form an arbitrary stationary process with a finite mean.

The available limit theorems for the partial sums of stationary stochastic processes show that the deviations of these partial sums from their mean can converge to a large variety of different self-similar processes with stationary increments.

Only in very rare circumstances will the limit be a stable motion!

For example: if (X_n) is a centered unit variance stationary Gaussian process with covariance function that is regularly varying with exponent $0 < \alpha < 1$, and H is a function of Hermit rank k satisfying $k\alpha < 1$, then, after an appropriate normalization,

$$\left(\frac{1}{A_n}\sum_{j=1}^{[nt]}\big(H(X_j) - EH(X_1)\big),\, 0 \le t \le 1\right) \Rightarrow \Big(Y_k(t),\, 0 \le t \le 1\Big),$$

where Y_k is a self-similar process with stationary increments which is in the kth Gaussian chaos (Dobrushin and Major (1979)).

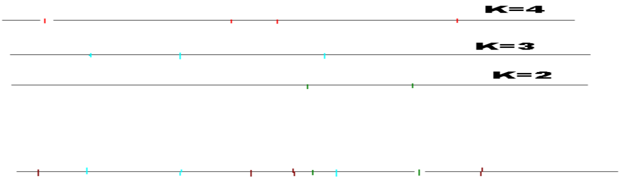
Many other self-similar process with stationary increments are possible, including processes in the stable chaos of an arbitrary order.

It was an open question whether the limit can be different from the stable motion in the case when the deviations of the arrival numbers from their mean, number of $\{k : 0 < T_k \le T\} - \lambda T$*, under the stationary probability is the dominating factor*. The answer is YES.

Example The cluster Poisson model

Basic ingredients:

- A rate λ_0 Poisson process of arrivals of clusters.
- A generic cluster: a point process N_c stopped at a random renewal K.
- At each point of a Poisson process an independent cluster (stopped point process) starts.
- Clusters are independent of the Poisson process.



Theorem (Fasen and Samorodnitsky (2008)) Suppose that for some $\beta > 1$, the within-cluster interarrival time law satisfies $F_X \in \mathrm{Reg}(1/\beta)$. Let the cluster size law satisfy $F_K \in \mathrm{Reg}(\alpha)$ for some $\alpha \in \big(1, \min(2, \beta)\big)$. Then in the slow growth regime

$$n \ll \Big(\big(\bar{F}_X(T)\big)^{-2} \bar{F}_K\big(\bar{F}_X(T)^{-1}\big) \Big)^{-\big((\alpha-1)/(\alpha+\beta-1)\big)+\epsilon}$$

for some $\epsilon > 0$, under a technical assumption,

$$[n \,\mathrm{Var}(I(T))]^{-1/2} D_{n,T} \Rightarrow B_H$$

in finite-dimensional distributions as $n \to \infty$, with $H = (2 + \beta - \alpha)/2\beta$.

Conclusions

The Fractional Browniam limit is very robust, and occurs under very weak assumptions in the fast growth regime, and may occur in the slow growth regime as well.

In the absence of additional information, Fractional Brownian motion should be used to approximate the deviations of the cumulative input from its mean.

The Lévy stable motion is very non-robust, and should be used as an approximation only with great care.