A bird-eye view of fluid queues in communication network models: heavy tails and long memory. Introduction

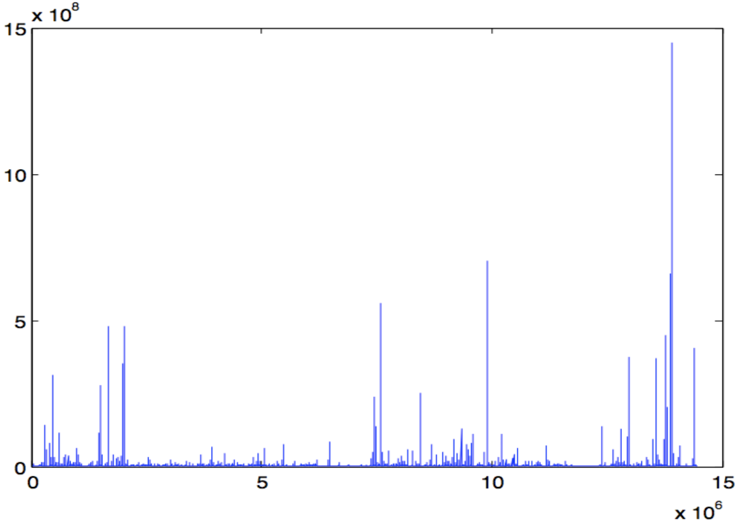
January 2009

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 technical term for such high variability is "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.



In this context, when one says that a random variable X is heavy tailed, one usually means that X has a regularly varying (right) tail: for some $\alpha > 0$

$$P(X > x) \sim x^{-\alpha} L(x), \; x \to \infty,$$

where L is a slowly varying function.

Notation and terminology:

- X (or its distribution F_X) is regularly varying with exponent α;
- $X \in \text{Reg}(\alpha)$ (or $F_X \in \text{Reg}(\alpha)$).

Important:

- One or more of random quantities involved in the input to a queuing system has heavy tails;
- if $X \in$Reg(α) and $\alpha < 2$, then X does not have a finite variance;
- if $X \in$Reg(α) and $\alpha < 1$, then X does not have a finite mean;
- highly irregular input processes tend to cause delays in queues.

Therefore: it is important to understand how irregular the input process to a queue be because of the heavy tails.

Slightly more technically: it is important to understand the deviations of the input process to a queue from its average behaviour.

A very influential step in this direction was taken by Mikosch, Resnick, Rootzén and Stegeman (2002), who considered two of the best known models for fluid input processes: the ON-OFF model and the Infinite Source Poisson 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 $F_{\mathrm{ON}} \in$Reg(α_{on}) and a finite mean μ_{on};
- the lengths of OFF-periods (Y_i) are iid with a common distribution $F_{\mathrm{OFF}} \in$Reg(α_{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 $\alpha_{\rm on} \in (1,2)$, $\alpha_{\rm off} > \alpha_{\rm on}$, and the cycle-length distribution is spread-out. Then

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

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

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. Assume the process is ergodic.
- The total amount of work brought into the system by the time $t \geq 0$ is

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

- Note: both $EI(t) = tEN(0)$ for all $t \geq 0$ and $I(t)/t \to EN(0)$ with probability 1 as $t \to \infty$.
- We think of $(I(t),\ t \in \mathbb{R})$ as the cumulative input process caused by a single "user".
- Typically, one assumes that there are many "users" contributing to the input.

- Let n denote the number of users (assumed to be large).
- Let $(I_j(t),\ t \in \mathbb{R}),\ j = 1, \dots, n$ be the cumulative input processes corresponding to different users.
- We assume that the input process $(I_j(t),\ t \in \mathbb{R}),\ j = 1, \dots, n$ are iid.
- The total input by the time t is $\sum_{j=1}^{n} I_j(t)$ and the average input is $n\, t\, EN(0),\ t \geq 0$.
- **Question**: what is the "bird-eye" behaviour of the deviation of the total input process from its average?

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

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

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.

Case 1 Suppose that the number of "users" n grows much faster than the time scale T.

Viewing T as, approximately, fixed, we can view the deviation

$$D_{n,T}(t) = \sum_{j=1}^{n} \int_0^{tT} \big(N_j(s) - EN(0)\big)\, ds, \ t \geq 0,$$

as the sum of, approximately, iid random functions.

For most reasonable models these functions have a finite variance at every point.

Therefore, it is reasonable to expect that the properly normalized deviation process $(D_{n,T}(t),\ t \geq 0)$ will converge to a Gaussian limit.

The covariance function of the limiting Gaussian process will be determined by the fact that the time scale T grows to zero as well.

Recall: the covariance function of the process $(N(t),\ t \in \mathbb{R})$, describing, for each "user", the number of sessions running at time t, is often regularly varying at infnity.

As a consequence, the scaling by the growing time scale T can be expected to cause a "power-like" behaviour of the covaraince of the limiting Gaussian process.

Therefore, it is reasonable to expect a Fractional Brownian limit of the properly normalized deviation process $(D_{n,T}(t),\ t \geq 0)$.

When the number of "users" n grows much faster than the time scale T, the term "the fast growth regime" was introduced by Mikosch, Resnick, Rootzén and Stegeman (2002).

Case 2 Suppose that the time scale T grows much faster than the number of "users" n.

Viewing now n as, approximately, fixed, we can view the deviation

$$D_{n,T}(t) = \sum_{j=1}^{n} \int_0^{tT} \big(N_j(s) - EN(0)\big)\, ds, \;\; t \geq 0$$

by looking, separately, at each term, which contains an individual input process at a very fast time scale.

At very fast time scales the heavy tailed sessions bring their work in the system almost immediately.

This means that each individual input process can be viewed as consisting of a large number of independent work requirements.

Therefore: it is reasonable to expect that each individual process, when properly normalized, will converge to a process with stationary and independent increments, i.e. to a Lévy process.

If the session durations have regularly varying tails in Reg(α) with $1 < \alpha < 2$, then this Lévy process will be an α-stable Lévy process.

The growing number of "users" n will only change the scale of the α-stable Lévy process.

Therefore, it is reasonable to expect an α-stable Lévy limit of the properly normalized deviation process $(D_{n,T}(t),\ t \geq 0)$.

When the time scale T grows much faster than the number of "users" n, the term "the slow growth regime" was introduced by Mikosch, Resnick, Rootzén and Stegeman (2002).

The main result of Mikosch, Resnick, Rootzén and Stegeman (2002) was that this intuition was valid for both the ON-OFF input model and the Infinite Source Poisson input model.

The boundary between 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.

In the fast growth regime:

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

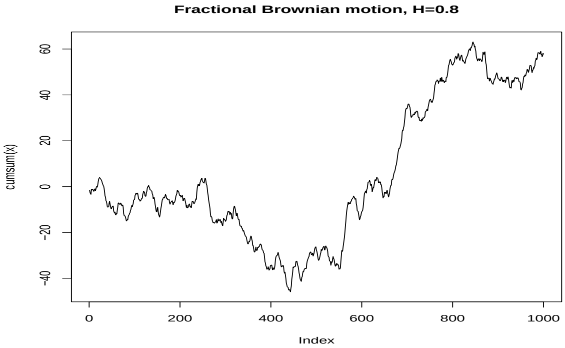
weakly in $\left(D[0,\infty),\ J_1\right)$. Here B_H is a Fractional Brownian motion with $H = (3-\alpha)/2$.

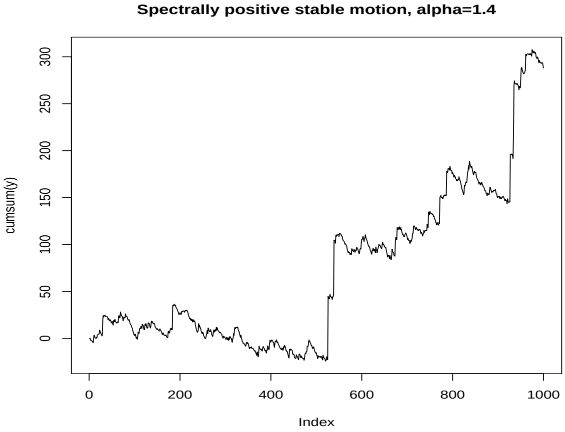
In the slow growth regime:

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

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

The accepted wisdom, therefore, has become that, from the bird-eye point of view, the deviations of the cumulative input to a communication network network from the average look either like a Fractional Brownian motion, or a Lévy stable motion.




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

A Fractional Brownian motion, and a Lévy stable motion are two very different processes.

- A Fractional Brownian motion with $1/2 < H < 1$ has long range dependent increments, but light tails.
- A Lévy stable motion has independent increments but heavy tails.

What behaviour is more widespread?