Journées MAS 2010
Some recent statistical advances in Telecommunication

François Roueff http://www.tsi.enst.fr/~roueff

Outline

A quick overview

Layers

New directions

Network traffic modeling

Long memory

Semiparametric estimation of the Infinite source Poisson process

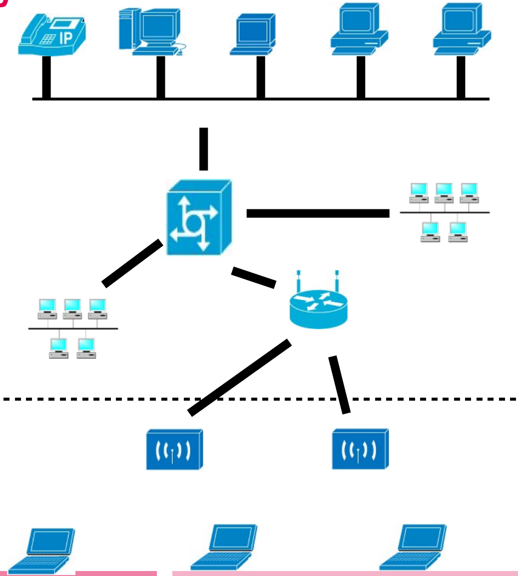
Layers

Statistical approaches such as stochastic modelling, statistical signal processing and time series analysis have been successfully applied in the last decade at several different layers of telecommunication systems.

The most significant advances are concerned with

1. the physical layer, motivated by the development of wireless network technologies.
2. the network layer, motivated by the development of the Internet.

Layers



Physical layer

Among many others, J.-F. Cardoso, G.B. Giannakis, P. Loubaton, E. Moulines, J. Najim, L. Tong and their co-authors contributed to the following topics :

1. Ill-posed inverse problem such as deconvolution and blind deconvolution.
2. Independent component analysis for blind source separation.
3. Asymptotic spectrum of Large random matrices for performance evaluation of MIMO systems.
4. Spectral estimation for cyclostationary processes for symbol timing recovery.

Network layer

Among many others, P. Abry, F. Baccelli, M. Crovella, C. Lévy-Leduc, V. Paxson, S. Resnick, P. Robert, G. Samorodnitsky, M.S. Taqqu, W. Willinger, and their co-authors contributed to

1. Network traffic modelling for network metrology and design based on simulated traffic data.
2. Inverse problems for network tomography.
3. Large data analysis and change-point detection for network monitoring and anomaly and/or attacks detection.
4. Stochastic geometry for performance evaluation of networks.

Some new directions

The recent development of overlay networks, wireless sensor networks, flexible networks ... have severely enlarged the scope of statistical approaches for communication systems, in particular the need of distributed statistical methods for solving the question: how to share disseminated information efficiently (energy / bandwidth) ?
See, among others, the recent works of P. Bianchi, M. Debbah, W. Hachem, J.M.F. Moura and their co-authors.
In addition to the preceding specific topics, we may cite

1. Error exponents for performance evaluation of decentralized detection in large sensor network.
2. Gossip algorithms for joint estimation in sensor networks.

A quick overview

Network traffic modeling

Stylized facts

Infinite source Poisson model

Applications

Long memory

Semiparametric estimation of the Infinite source Poisson process

Network flows are heavy tailed

Network traffic is an aggregation of flows with heavy tail characteristics:

1. a lot of small flows (mice)
2. and some very large flows (elephants),

see for instance the phd thesis by Y. Chabchoub and works by S. Resnick in the 2000's.

Interesting problems related to flows statistics :

1. Statistical analysis of sampled traffic data.
2. Heavy tail statistics of flows : durations VS workload.

Network flows are heavy tailed

Network traffic is an aggregation of flows with heavy tail characteristics:

1. a lot of small flows (mice)
2. and some very large flows (elephants),

see for instance the phd thesis by Y. Chabchoub and works by S. Resnick in the 2000's.

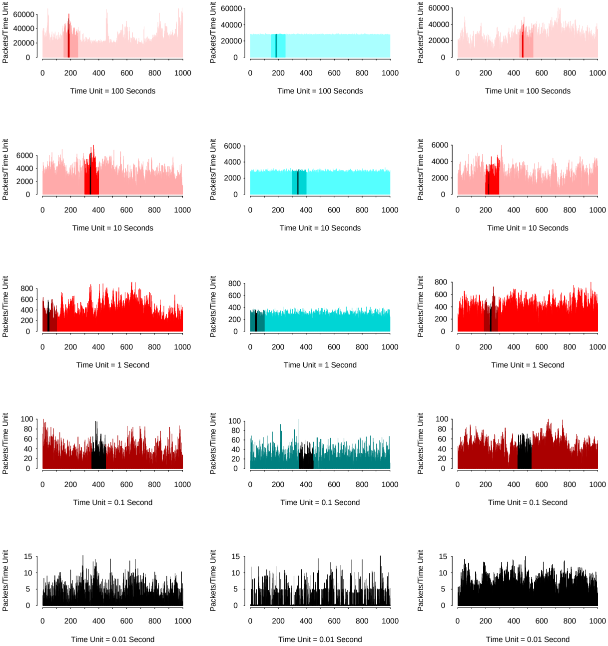
Interesting problems related to flows statistics :

1. Statistical analysis of sampled traffic data.
2. Heavy tail statistics of flows : durations VS workload.

... but requires a thorough analysis of complete teletraffic data !

Back to 1995: *scaling analysis* of LAN traffic

The following picture is borrowed from Willinger, Taqqu, Leland, Wilson (1995).


Packets/Time Unit
Time Unit = 100 Seconds
Time Unit = 100 Seconds
Time Unit = 100 Seconds
Time Unit = 10 Seconds
Time Unit = 10 Seconds
Time Unit = 10 Seconds
Time Unit = 1 Second
Time Unit = 1 Second
Time Unit = 1 Second
Time Unit = 0.1 Second
Time Unit = 0.1 Second
Time Unit = 0.1 Second
Time Unit = 0.01 Second
Time Unit = 0.01 Second
Time Unit = 0.01 Second

fBm and aggregation of On-Off models

Leland, Taqqu, Willinger and Wilson (1993) observed on Ethenert traffic data a self-similar behavior similar to that of a fractional Brownian motion.

fractional Brownian motion (fBm)

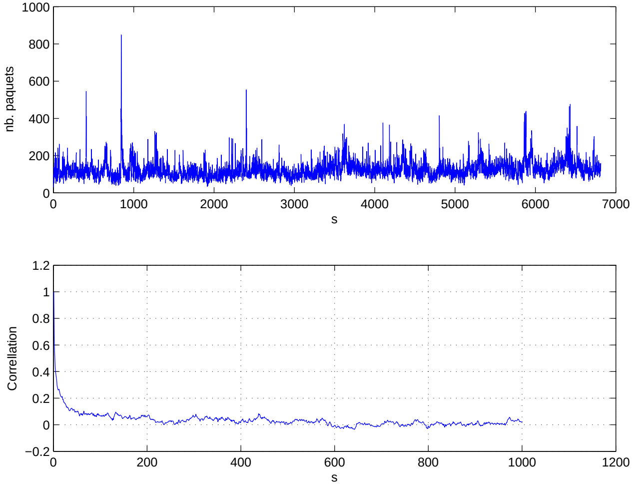
The cumulated throughput $\{X(t), t > 0\}$ can be modelled as a H-self-similar Gaussian process X_H with stationary increments :

1. $X_H(0) = 0$, $\mathbb{E}[X_H(t)] = 0$ for all $t > 0$,
2. $\mathbb{E}[X_H(s)X_H(t)] \propto s^{2H} + t^{2H} - |t - s|^{2H}$,

with $H \in (1/2, 1)$.

Implications on correlations: long range dependence

Around 2 hours IP traffic record aggregated every second.



Interpretation (Willinger *et al* (1995))

Let $\{S_i(t), t > 0\}$ be On-Off independent sources with heavy tailed sessions with index $\alpha \in (1, 2)$, and define

$$X_{N,T}(t) = \int_0^{tT} \sum_{i=1}^{N} S_i(s)\, \mathrm{d}s \,.$$

Then, if $N \to \infty$ and then $T \to \infty$,

$$T^{-H} N^{-1/2} (X_{N,T} - \mathbb{E}[X_{N,T}]) \Rightarrow X_H \,.$$

where X_H is a fBm with Hurst parameter $H = (3 - \alpha)/2$.
Many extensions to this result:

1. fast/slow/intermediate growth limits,
2. Infinite variance rewards (Telecom process).

Infinite source Poisson model

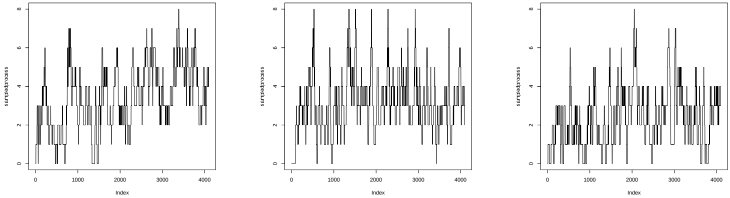
Consider a shot–noise process with random rectangle pulses

$$X_t = \sum_k U_k \mathbb{1}(T_k \leq t < T_k + \eta_k) ,$$

where $\{T_k\}$ are Poisson arrival times with intensity λ and $\{(U_k, \eta_k)\}$ are i.i.d. marks, independent of the arrival times. Assume that η is heavy tailed with index $\alpha \in (1, 2)$.

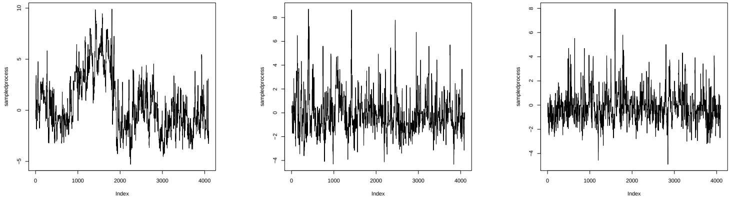
M/G/∞ queue

If $U_n = 1$, X is an M/G/∞ queue.
Here are three stationary sample paths for $\alpha = 1.1, 1.5$ and $\alpha = 1.8$.



Centered exponential rates

Here $1 + U_n \sim \exp(1)$.
Here are three sample paths for $\alpha = 1.1, 1.5$ and $\alpha = 1.8$.



Large scales approximations

Mikosch *et al* (2002) ($U \equiv 1$) and then Maulik *et al* (2002) studied the behavior of the cumulated workload

$$Z_T(t) = \int_0^{tT} X_s \, \mathrm{d}s \quad \text{centered and normalized}$$

as T (and λ) $\to \infty$. (The situation is similar to that of On-Off studied by Taqqu *et al*).
Resnick and van den Berg (2000) studied the convergence of similar workload processes in M_1-topology. This is necessary for providing asymptotic approximations of a fluid queue fed by such a workload:

$$Q_T(t) = \int_0^{tT} (X_s - \mathbb{E}[X_s]) \, \mathrm{d}s - \inf_{0 \le u \le tT} \int_0^u (X_s - \mathbb{E}[X_s]) \, \mathrm{d}s \, .$$

The bounded functional case

We recently extended the M_1-topology convergence to the process

$$Z_T(\phi, t) = \int_0^{tT} \phi(\{X_{s+u}, u \in [0, h]\})\, \mathrm{d}s \quad \text{centered and normalized}$$

under a general asymptotically independence assumption: there exists a measure ν on $(0, \infty] \times [0, \infty]$ such that $\nu((1, \infty] \times [0, \infty]) = 1$ and, as $n \to \infty$,

$$n\, \mathbb{P}\left(\left(\frac{\eta}{a(n)}, U\right) \in \cdot\right) \xrightarrow{v} \nu = \nu_\alpha \times G\,,$$

where $\nu_\alpha(x, \infty) = x^{-\alpha}$.

Lévy stable limit

We get

$$\frac{1}{a(T)}\left\{Z_T(\phi,u)-\mathbb{E}[Z_T(\phi,u)]\right\} \Rightarrow \int_0^u \int \left\{\mathcal{E}(w,\phi)-\mathcal{E}(0,\phi)\right\}\ M_\alpha(\mathrm{d}s,\mathrm{d}w)\ ,$$

where M_α is a totally skewed to the right α-stable random measure with control measure $\lambda c_\alpha \mathrm{Leb} \times G$ and

$$\mathcal{E}(w,\phi)=\mathbb{E}[\phi(\{w+X_{s+u}, u\in[0,h]\})], \quad \text{for all } w\ .$$

Application 1: fluid queue

The fluid queue fed by a truncated infinite source Poisson process is

$$Q_T(t) = \int_0^{tT} (\phi(X_s) - \mathbb{E}[\phi(X_s)])\, \mathrm{d}s - \inf_{0 \le u \le tT} \int_0^{u} (\phi(X_s) - \mathbb{E}[\phi(X_s)])\, \mathrm{d}s \,,$$

with $\phi(x) = x \wedge a$ defined on $x \in \mathbb{R}_+$ ($h = 0$).

Its large scale approximation is obtained by continuous mapping theorem: it has infinite variance !

Application 2: empirical process

Fix $u_0 = 0 < u_1 < \cdots < u_p$.
We obtain the asymptotic behavior of the multivariate empirical process. Set

$$\widehat{P}^{\mathbf{u}}_{tT}(-\infty, \mathbf{x}] = \frac{1}{T} \int_0^{tT} \prod_{i=0}^{p} \mathbb{1}_{X_{s+u_i} \leq x_i} \mathrm{d}s, \quad \mathbf{x} = (x_0, \ldots, x_p) \in \mathbb{R}^{p+1}, \quad t \geq 0 \; .$$

$(h = u_p)$ We get

$$\frac{T}{a(T)} \left\{ \widehat{P}^{\mathbf{u}}_{tT}(-\infty, \mathbf{x}] - \mathbb{P}(X_0 \leq x_0, X_{u_1} \leq x_1, \ldots, X_{u_p} \leq x_p) \right\} \Rightarrow S_\alpha(t, \mathbf{x}) \; .$$

A quick overview

Network traffic modeling

Long memory

- Semi-parametric models
- Standard examples
- A Network model example

Semiparametric estimation of the Infinite source Poisson process

The memory parameter d

Definition

The process X is said to have memory parameter d and short-range spectral density f_* if if $f_\star(\lambda)$ is non-zero and continuous at $\lambda = 0$ and X has a spectral density

$$f(\lambda) = |1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{-2d} f_\star(\lambda) . \quad (1)$$

Long range dependence

Observe that:

1. $f(\lambda) \asymp |\lambda|^{-2d}$ as $\lambda \to 0$;
2. for $d \in (0, 1/2)$, alternative (but not equivalent) definitions use the auto-covariance function, for instance,

$$\mathrm{cov}(X_0, X_t) \sim c\, t^{-1+2d} \quad \text{as} \quad t \to \infty\,. \tag{2}$$

3. for $d > 0$, (1) implies $\sum_t |\mathrm{cov}(X_0, X_t)| = \infty$ (LRD).

If X is weakly stationary then $d < 1/2$ but one can drop this assumption and take $d \in \mathbb{R}$ as follows ../...

The memory parameter d in $\mathbb{R}$

Let

$$V_d = \left\{ (\mathrm{h}_k) \in \ell_c \;:\; \int_{-\pi}^{\pi} |\mathrm{h}^*(\lambda)|^2 \, |\lambda|^{-2d} \mathrm{d}\lambda \right\} \;, \text{ where } \mathrm{h}^*(\lambda) = \sum_k \mathrm{h}_k \mathrm{e}^{-\mathrm{i}k\lambda} \;.$$

Definition

The process $\left\{ X(\mathrm{h}) := \sum_k \mathrm{h}_k X_k, \mathrm{h} \in V_d \right\}$ has memory parameter d if there exists a function f defined on $(-\pi, \pi]$, integrable away of 0 and such that

1. $f(\lambda) \asymp |\lambda|^{-2d}$ in a neighborhood of 0,
2. for all $(\mathrm{h}_k) \in V_d$,

$$\operatorname{var}(X(\mathrm{h})) = \int_{-\pi}^{\pi} |\mathrm{h}^*(\lambda)|^2 \; f(\lambda) \, \mathrm{d}\lambda \;. \tag{3}$$

Standard examples

The estimation of the memory parameter can be done using Fourier or wavelet analysis of the data sample. Standard examples are

1. Gaussian processes: fBm, fGn.
2. Linear processes: FARIMA(p, d, q).
3. Non Linear processes: Gaussian subordinator (in progress), Stochastic volatility models...

A Network model example

The Infinite source Poisson model

$$X_t = \sum_k U_k \mathbb{1}(T_k \leq t < T_k + \eta_k) \;,$$

where $\{T_k\}$ are Poisson arrival times with intensity λ and $\{(U_k, \eta_k)\}$ are i.i.d. marks, independent of the arrival times and satisfying

$$\mathbb{E}[U_0^2 \mathbb{1}_{\eta_0 > t}] = L_2(t)\, t^{-\alpha} \;,$$

with L_2 slowly varying as $t \to \infty$ and $\alpha \in (0, 2)$.

Stationarity issues

If $\mathbb{E}[\eta_0] < \infty$ ($\Leftarrow \alpha > 1$) then a stationary version of X can be defined by taking a uniform intensity on $\mathbb{R}$ for the arrivals $\{T_k\}$.
If $\mathbb{E}[\eta_0] = \infty$ ($\Leftarrow \alpha < 1$), this version is only defined for

$$X(\mathrm{h}) = \sum_k U_k \sum_{T_k \leq t < T_k + \eta_k} \mathrm{h}_t \left(= \sum_t \mathrm{h}_t X_t \right) ,$$

when $\sum_t \mathrm{h}_t = 0$, in which case

$$\mathrm{var}\left(X(\mathrm{h})\right) = \mathbb{E}\left[U_0^2 \sum_{t,t'} \mathrm{h}_t \mathrm{h}_{t'} \left(\eta_0 - |t - t'|\right)_+ \right] .$$

Observe that, for η_0 large enough,

$$\sum_{t,t'} \mathrm{h}_t \mathrm{h}_{t'} \left(\eta_0 - |t - t'|\right)_+ = -\sum_{t,t'} |t - t'| \mathrm{h}_t \mathrm{h}_{t'} .$$

Memory parameter of the infinite source Poisson model

One finds

$$\mathrm{var}\left(X(\mathrm{h})\right) = \int_{-\infty}^{\infty} |\mathrm{h}^*(\lambda)|^2 \, \frac{\mathbb{E}\left[U_0^2\{1-\cos(\lambda\eta_0)\}\right]}{\pi\lambda^2} \, \mathrm{d}\lambda \; .$$

Hence if, as $\lambda \to 0$,

$$\mathbb{E}\left[U_0^2\{1-\cos(\lambda\eta_0)\}\right] \asymp |\lambda|^{\alpha} \; ,$$

then (3) holds with

$$d = 1 - \alpha/2 \in (0,1) \; ,$$

and hence X has memory parameter d.
It makes sense to apply the wavelet estimator of the memory parameter.

A quick overview

Network traffic modeling

Long memory

Semiparametric estimation of the Infinite source Poisson process

- Heavy tails VS long memory
- Tail/LRD estimation
- Simulations

Heavy tails VS ...

α is the parameter of interest, since it determines the asymptotic behavior of the queue, its stability ....

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

Heavy tails VS ...

α is the parameter of interest, since it determines the asymptotic behavior of the queue, its stability ....
Natural approach : heavy tail estimator such as the Hill estimator.
Drawbacks :

Hidden information

Difficult/costly if not impossible to observe, say, $(U_n, \eta_n)_n$!

Heavy tails VS ...

α is the parameter of interest, since it determines the asymptotic behavior of the queue, its stability ....
Natural approach : heavy tail estimator such as the Hill estimator.
Drawbacks :

Hidden information

Difficult/costly if not impossible to observe, say, $(U_n, \eta_n)_n$!

Tails of X

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

Heavy tails VS ...

α is the parameter of interest, since it determines the asymptotic behavior of the queue, its stability ....
Natural approach : heavy tail estimator such as the Hill estimator.
Drawbacks :

Hidden information

Difficult/costly if not impossible to observe, say, $(U_n, \eta_n)_n$!

Tails of X

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

Return times to the empty state

But in real data, the empty state cannot be identify.

... long memory

Here we estimate α through the long memory parameter d 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, See the works by Robinson, Hurvich, Moulines, Soulier in the late 90's.

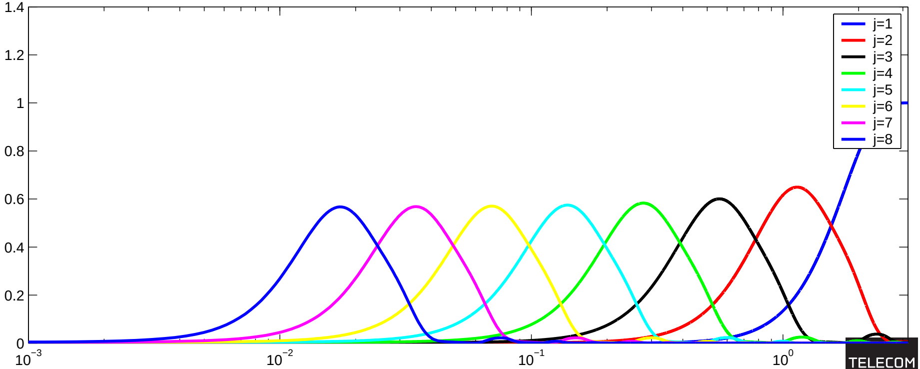
Wavelet methods

The Gaussian and linear cases have been treated in the 2000's.

Discrete wavelet transform (DWT)

Collection of sequences $(W_{j,k})_k$ = filtering by FIR h_j followed by 2^j decimation.

For Daubechies wavelets with 2 vanishing moments, $2^{-j}\left|\mathrm{h}_j^*(\lambda)\right|^2$ is



Bias: scale spectrum at large scales

For a well chosen wavelet with respect to d (frequency resolution+number of vanishing moments), we have

$$\mathrm{var}\left(W_{j,0}\right) \sim C\, 2^{2dj}, \quad \text{as} \quad j \to \infty\ ,$$

where C is a positive constant.
Under standard semi-parametric type of assumptions, the $\sim$ can be made more precise, e.g.

$$\left|\mathrm{var}\left(W_{j,0}\right) - \sigma^2 2^{2dj}\right| \leq C\, 2^{-\beta j}\, 2^{2dj}\ . \tag{4}$$

Estimation is based on the scalogram

$$\widehat{\sigma}_j^2 = \sum_k W_{j,k}^2\ .$$

Estimation result

The bias term behaves as in (4), if $\beta \leq 2 - \alpha$ and

$$\mathbb{E}[U_0^2 \, \mathbb{1}_{\eta_0 > t}] = c \, t^{-\alpha} \, (1 + O(|t|^{-\beta})) \, , \tag{5}$$

$$\text{or} \quad \mathbb{E}\left[U_0^2\{1 - \cos(\lambda \eta_0)\}\right] = c \, \lambda^{\alpha} \, (1 + O(|\lambda|^{\beta})) \, .$$

The fluctuation term behaves differently for $\alpha > 1$:

$$2^{-2dj}\left[\widehat{\sigma}_j^2 - \mathrm{var}\left(W_{j,k}\right)\right] = O_P\left(n_j^{-1/2} 2^{(\alpha-1)j/2}\right)$$

(instead of $n_j^{-1/2}$ in the linear case).

Rate of convergence

We obtain a rate of convergence slower than in the case of linear processes but:

If one observes the variables (U_k, η_k) directly, the best achievable rate under the condition (5) is precisely the rate obtained by the wavelet estimator.

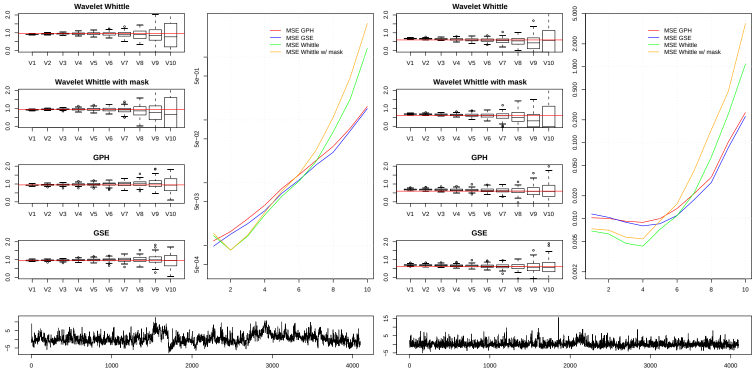
Hence we actually obtain the best achievable rate (recall that this is for $\beta \leq 2 - \alpha$).

Simulations

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

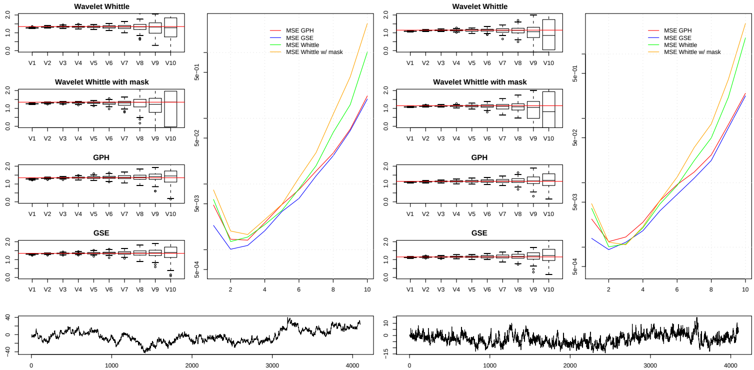
Stable case

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.



TELECOM ParisTech