Recent Results for Stochastic Models with Long-Range Dependence and Heavy Tails

Venkat Anantharam

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, California

February 23, 2011

Stochastic Approximation

Given a function h, we seek to find a solution to $h(x^*) = 0$. However, we only observe $h(x_n)$ in noise. Use the following recursion.

Algorithm

$$x[n+1] = a_n \left[h(x_n) + M_{n+1}\right]$$

where originally, M is mean zero, uncorrelated, bounded variance noise.

Stochastic Approximation

Under suitable stability conditions (e.g. $\sup|x_n| < K$), the recursion can be approximated by ODE

$$\dot{x}(t) = h(x(t))$$

Which can be shown to converge if

- $\sum a_n = \infty$
- $\sum a_n^2 < \infty$

ROBBINS-MONRO SCHEME (1951)

to solve $h(x) = 0$ given noisy measurements of $h(\cdot) : \mathcal{R}^d \to \mathcal{R}^d$:

$$x_{n+1} = x_n + a(n)[h(x_n) + M_{n+1}], \; n \geq 0.$$

- $\|h(x) - h(y)\| \leq L\|x - y\| \; \forall \; x, y$
- $E[M_{n+1}|x_i, M_i, i \leq n] = 0 \; \forall n$. ('martingale difference')
- $E[\|M_{n+1}\|^2|x_i, M_i, i \leq n] \leq K(1 + \|x_n\|^2) \; \forall \; n$.
- $a(n) > 0, \; \sum_n a(n) = \infty, \; \sum_n a(n)^2 < \infty$.

The 'o.d.e.' approach:

(Derevitskii – Fradkov – Ljung)

Consider the iteration as a 'noisy discretization' of the o.d.e. (*ordinary differential equation*)

$$\dot{x}(t) = h(x(t))$$

with step-sizes $\{a(n)\}$. If

- x_ns track $x(t)$ in a suitable sense, and
- $x(t) \to H := \{x : h(x) = 0\}$,

then we can expect $x_n \to H$ a.s.

EXAMPLES

1. Gradient schemes

Here $h(x) = -\nabla F(x)$. As an example, consider N users share an ergodic Markov channel with stationary distribution ν.

Aim: Minimize average power subject to a minimum rate constraint.

- $A = \{$ unit coordinate vectors in $\mathcal{R}^N\}$ ($i-$th vector $\approx$ choice of $i-$th user for the slot).

- $p_2(y|x) =$ conditional distribution of the user given channel state,

- $p_1(q|y,x) =$ conditional distribution of this user's power consumption.

Problem:

$$\min \int \nu(dx) \sum_{y\in A} \int_0^\infty p_1(dq|y,x)p_2(y|x)q \quad \text{subject to}$$

$$\int \nu(dx) \sum_{y\in A} \int_0^\infty p_1(dq|y,x) \log(1+qy_i x_i) \geq C_i \ \forall i.$$

Central idea: Use the Lagrange multiplier formulation in order to cast the constrained optimization problem as an unconstrained min-max (= max-min) problem, do the minimization over both the users and power explicitly as above, and the maximization over Lagrange multipliers by stochastic approximation. The foregoing theory ensures desired asymptotics (also verified by simulation experiments).

The optimal solution is to select user

$$k = \text{argmin}_i \left((\lambda_i - \frac{1}{x_i})^+ - \lambda_i [\log(1 + (\lambda_i - \frac{1}{x_i})^+ x_i) - C_i] \right),$$

who will transmit power

$$q^* = (\lambda_k - \frac{1}{x_k})^+,$$

λ_i being the Lagrange multiplier associated with the $i-$th constraint.

$\{\lambda_i\}$ can be learnt adaptively by the stochastic gradient scheme

$$\lambda_i(n+1) = \Gamma(\lambda_i(n) - a(n) y_i(n) [\log(1 + (\lambda_i - \frac{1}{x_i})^+ x_i(n)) - C_i]), \ \forall i.$$

Here $y_i(n) = I\{\alpha_i \leq \alpha_j, j \neq i\}$ for

$$\alpha_i = q_i^* - \lambda_i(n) [\log(1 + (\lambda_i - \frac{1}{x_i(n)})^+ x_i(n)) - C_i], \ 1 \leq i \leq N,$$

and Γ is projection to $[0, L]$ for a large L.

2. Fixed point iterations

Here, $h(x) = F(x) - x$, F a contraction w.r.t. a suitable norm.

Aim: Find its unique fixed point x^* given by $x^* = F(x^*)$. ($\approx$ globally asymptotically stable equilibrium for the o.d.e. $\dot{x}(t) = F(x(t)) - x(t)$. This can be extended to nonexpansive maps in some cases.)

Application to Dynamic Programming: Queue process $\{X_n\}$ given by $X_{n+1} = X_n - u_n + W_{n+1}$, where $\{W_n\} \approx$ i.i.d. packet arrival process with law μ and $u_n \in [0, x_n] \approx$ the number of packets transmitted at time n.

Constrained Markov decision process: Minimize

$$\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1} c(X_m, u_m) \text{ s.t. } \overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1} c_i(X_m, u_m) \leq C_i, \ \forall i.$$

Lagrange multiplier formulation $\Rightarrow$ *unconstrained* MDP with running cost $c + \sum_i \lambda_i c_i$, λ_i's $\approx$ the Lagrange multipliers. The corresponding dynamic programming equation is

$$\tilde{V}(x) = \min_u [c(x,u) + \sum_{i=1}^{N} \lambda_i c_i(x,u) - \beta + \sum_w \mu(w)\tilde{V}(x - u + w)].$$

Here $\beta =$ the optimal cost.

View the transition $X_n \to X_{n+1}$ as a composition of

$X_n \to X_n^+ = X_n - u_n$ (the 'post-state')

and

$X_n^+ \to X_{n+1} = X_n^+ + W_{n+1}$.

In terms of $\{X_n^+\}$, the dynamic programming equation becomes

$$V(x) = \sum_w \mu(w) \min_u [c(x+w, u) + \sum_{i=1}^{N} \lambda_i c_i(x+w, u) - \beta + V(x-u+w)].$$

Note: The minimization is now inside the expectation. This allows us to write the stochastic approximation version of the corresponding 'relative value iteration':

Let $\nu_n(i) := \sum_{m=0}^{n} I\{X_m = i\}$ ('local clock') and i_0 a prescribed state.

$$\begin{aligned} V_{n+1}(i) \;=\; & V_n(i) + a(\nu(i,n)) I\{X_n^+ = i\}[\min_u [c(X_{n+1}, u) + \\ & \sum_{i=1}^{N} \lambda_i(n) c_i(X_{n+1}, u) - V_n(i_0) + V(X_{n+1}^+)]. \end{aligned}$$

The Lagrange multipliers are updated on a slower timescale by the stochastic ascent:

$$\lambda_i(n+1) = \lambda_i(n) + b(n)[c_i(X_n, u_n) - C_i] \ \forall i.$$

The convergence can be proved by using the two timescale analysis above.

That the slow component performs the correct gradient ascent is a consequence of the generalized ‘envelope theorem’ from mathematical economics.

Back to Robbins-Monro: Idea of proof:

1. Let $t(0) = 0, t(n) = \sum_{m=0} a(m)$. Then $t(m) \uparrow \infty$.

2. Let $\bar{x}(t(n)) = x_n$ with linear interpolation on $[t(n), t(n+1)]$ (piecewise linear interpolation).

3. For $s \geq 0$, let $\dot{x}^s(t) = h(x^s(t)), x^s(s) = \bar{x}(s)$.

Then if $P(\sup_n \|x_n\| < \infty) = 1$ (i.e., iterates remain bounded with probability one), then for $T > 0$,

$$\lim_{s \uparrow \infty} \max_{t \in [s, s+T]} \|\bar{x}(t) - x^s(t)\| = 0 \text{ w.p. } 1.$$

To prove this, use **Gronwall inequality** to obtain:

$\max_{t \in [s, s+T]} \|\bar{x}(t) - x^s(t)\| \leq$

(error due to discretization) + (error due to noise).

$a(n) \to 0 \implies$ error due to discretization $\to 0$.

The martingale $\sum_n a(n) M_{n+1}$ converges with prob. 1

$\implies$ the 'tail' $\sum_{m \geq n} a(m) M_{m+1}$ goes to zero w. p. 1

$\implies$ error due to noise $\to 0$.

Need 'stability': $\sup_n \|x_n\| < \infty$ with prob. 1.

Test for stability:

Let $h_\infty(x) := \lim_{0<a\uparrow\infty} \frac{h(ax)}{a}$.

If the origin is the globally asymptotically stable equilibrium for $\dot{x}(t) = h_\infty(x(t))$, then $\sup_n \|x_n\| < \infty$ with prob. 1.

‘martingale difference noise’ $\{M_n\}$:

- $E[M_{n+1}|x_m, M_m, m \leq n] = 0 \Longrightarrow$ ‘uncorrelated’
- $E[\|M_{n+1}\|^2|x_m, M_m, m \leq n] \leq K(1 + \|x_n\|^2) \Longrightarrow$ ‘light (conditional) tails’

$\Longrightarrow$ *‘GOOD’ NOISE*. In practice, noise can get *BAD* (long range correlations) or even outright *UGLY* (heavy tails).

MIKOSCH, RESNICK, ROOTZEN, AND STEGEMAN characterize the regimes when one can expect these (Annals of Applied Probability, 2002)

Applications

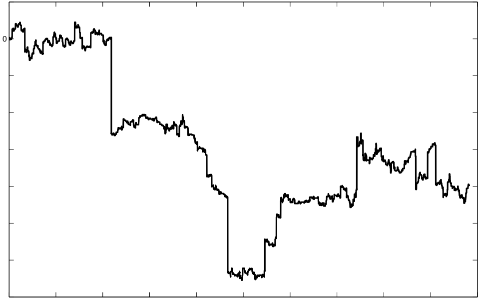
- Many DSP applications, including adaptive filtering
- Network control
- Adaptive routing
- Service time control in queuing networks

In network applications, we wish to run control algorithms based on the values of the flows. However, these might not be directly observed, might be available as noisy estimates.
It has been observed empirically that queues and flows in large computer networks exhibit heavy tailed distributions or long range dependence.

Alpha stable Levy motion

- Take X_i i.i.d. symmetric, $P(|X_1| > x) = x^{-\alpha}L(x)$ then
- $$\frac{S_{nt}}{(nL(n))^{\frac{1}{\alpha}}} \to^{d} S\alpha S$$

  (symmetric α-stable Levy motion)



Alpha-stable Levy Motion properties

- stationary, α-stable, i.i.d. increments.
- Distribution of $\frac{S_{nt}}{\sqrt{n}} \to \infty$ (long range dependence)
- $\text{Var}(S_t) = \infty$
- Self-similarity: $S_{nt} =^d n^{\frac{1}{\alpha}} S_t$

Samorodnitsky, Taqqu. "Stable Non-Gaussian Random Processes: Stochastic models with infinite variance"

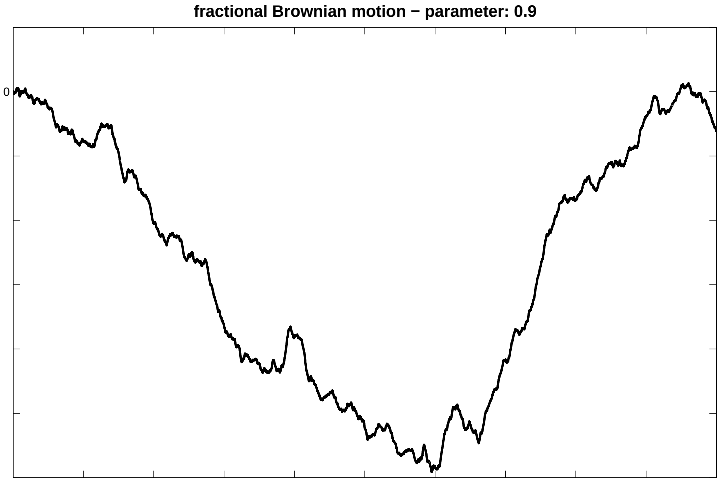
Fractional Brownian Motion

- Fractional Brownian Motion is the unique Gaussian H-sssi process.
- $\mathrm{Cov}(B_H(t_1), B_H(t_2)) = \frac{1}{2}\left\{|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}\right\}\mathrm{Var}(B_H(1))$
- H-sssi

fBM limit

Let $Cov(X_1, X_n) = n^{-\alpha}L(n)$ regularly varying. And $\{X_i\}$ zero-mean Gaussian.
Then, $\frac{S_{nt}}{n^H} \to^d B_H(t)$, where $H = (1 - \frac{\alpha}{2})$.

Fractional Brownian Motion



Consider a stochastic approximation scheme in $\mathcal{R}^d$ of the type

$$x_{n+1} = x_n + a(n)[h(x_n) + M_{n+1} + R(n)B_{n+1} + D(n)S_{n+1} + \zeta_{n+1}],$$

where

- M_{n+1} for $n \geq 0$ is the martingale difference noise as before,
- $B_{n+1} := \tilde{B}(n+1) - \tilde{B}(n)$, where $\tilde{B}(t), t \geq 0$, is a d-dimensional fractional Brownian motion with Hurst parameter $\nu \in (0, 1)$,
- $S_{n+1} := \tilde{S}(n+1) - \tilde{S}(n)$, where $\tilde{S}(t), t \geq 0$, is a symmetric α-stable process with $1 < \alpha < 2$,
- $\{\zeta_n\}$ is an 'error' process satisfying $\sup_n \|\zeta_n\| \leq K_0 < \infty$ a.s. and $\zeta_n \to 0$ a.s.,

- $\{R(n)\}$ is a bounded deterministic sequence of $d \times d$ matrices,
- $\{D(n)\}$ is a bounded sequence of $d \times d$ random matrices adapted to $\mathcal{F}_n := \sigma(x_i, B_i, M_i, S_i, \zeta_i, i \leq n)$,
- $\{a(n)\}$ as before with $a(n) = \Theta(n^{-\kappa})$ for $\kappa \in (\frac{1}{2}, 1]$.

Main theorem:

Suppose that the o.d.e. $\dot{x}(t) = h(x(t))$ has x^* as the unique globally asymptotically stable equilibrium and in addition, the following 'stability condition' holds:

for some $\xi, 1 \leq \xi < \alpha$,

$$\sup_n E[\|x_n\|^{\xi}] < \infty.$$

Then for $1 < \xi' < \xi$

$$E[\|x_n - x^*\|^{\xi'}] \to 0.$$

(Key steps of the proof follow.)

Recall that

$$E[\|\tilde{B}(t) - \tilde{B}(s)\|^2] = C|t-s|^{2\nu}, t \geq s \geq 0,$$

and for $I :=$ the identity matrix,

$$E[(\tilde{B}(t) - \tilde{B}(s))(\tilde{B}(u) - \tilde{B}(v))]$$

$$= \frac{C}{2}\left(|t-v|^{2\nu} + |s-u|^{2\nu} - |t-u|^{2\nu} - |s-v|^{2\nu}\right) I,$$

Then for $m_r(n) := \min\{n' \geq n : \sum_{i=n}^{n'} a(i) \geq r\}$ and $\gamma := 2\kappa(1-\nu)$ for $\nu < \frac{1}{2}$, $:= \kappa$ for $\nu \geq \frac{1}{2}$,

$$E\left[\| \sum_{m_s(n)}^{m_t(n)} a(i)R(i)B(i+1)\|^2\right] \leq \frac{C}{n^\gamma}.$$

Fernique's inequality for Gaussian processes:

For $p \geq 2, K := \frac{5\sqrt{2\pi}}{2}p^2, \gamma := \sqrt{1 + 4\log p}$ and

$$\varphi(h) := \max_{s,t \in [0,1], |s-t| \leq h} E\left[(X_t - X_s)^2\right]^{\frac{1}{2}},$$

the following holds:

$$P\left(\max_{t \in [0,1]} |X_t| \geq \left[\max_{x \in [0,1]} E[X_t^2]^{\frac{1}{2}} + (2+\sqrt{2}) \int_1^\infty \varphi(hp^{-y^2}) dy\right] x\right)$$

$$\leq \quad K\Psi(x).$$

Combining, these lead to: for prescribed $T > 0$ and

$$m(n) \geq \min\{m \geq n : \sum_{j=n}^{m} a(j) \geq T\},$$

we have,

$$E\left[\sum_{n \leq N \leq m(n)} \| \sum_{i=n}^{N} a(i)R(i)B_{i+1}\|^2\right] \to 0.$$

'Joulin's inequality' $\Longrightarrow$

$$P\left(\sup_{n\le j\le m(n)} \|\sum_{i=n}^{j} a(i)D(i)S_{i+1}\| \ge x\right)$$

$$\le \frac{C(\sum_{i=n}^{m(n)} a(i)^{\frac{\alpha^2-1}{\alpha}+1})^{\frac{\alpha}{\alpha+1}}}{x^{\alpha}}$$

for

$$x > C(\sum_{i=n}^{m(n)} a(i)^{\frac{\alpha^2-1}{\alpha}+1})^{\frac{1}{\alpha+1}}.$$

(A. Joulin, 'On maximal inequalities for stable stochastic integrals', Potential Analysis 26 (2007), pp. 57-78.)

$\Longrightarrow$ for $0 < \xi' < \xi$,

$$E\left[\sum_{n\le N\le m(n)} \|\sum_{i=n}^{N} a(i)D(i)S_{i+1}\|^{\xi'}\right] \to 0.$$

As before, Gronwall inequality $\Longrightarrow$

(Deviation from o.d.e. in ξ'th mean on an interval of length T) $\leq$

(discretization error)

+ (error due to martingale difference noise)

+ (error due to long range dependent noise)

+ (error due to heavy-tailed noise)

+ (error due to $\{\zeta_n\}$)

$\Longrightarrow$ convergence in ξ'th mean

Other results:

1. The stability test applies!

2. Concentration result for constant stepsize algorithms

3. Extension to general attractors, Markov noise,

   asynchronous schemes.

Entropy rate

- Let (X_n) be a discrete-time $\mathcal{X}$-valued ergodic process, $\mathcal{X}$ finite.
- $p(x_1, \ldots, x_n)$ denotes $P(X_1 = x_1, \ldots, X_n = x_n)$.
- $$\lim_n \frac{1}{n} E[-\log p(X_1, \ldots, X_n)] =: \eta \quad \text{exists,}$$
  and is called the entropy rate of the process.
  The logarithm is to base 2.
- Write X_1^n for $(X_1, \ldots, X_n)$.
- In fact:
  $$\eta = E[-\log p(X_1 | X_{-\infty}^0)] \ .$$
- The ergodic theorem implies:
  $$\frac{1}{n} \sum_{k=1}^{n} -\log p(X_k | X_{-\infty}^{k-1}) \to \eta \quad \text{a.s.}$$

Data compression

- Let $\{0,1\}^*$ denote the set of binary strings of finite length.
- Consider a prefix-free mapping:

$$\mathcal{X}^n \to \{0,1\}^* \ .$$

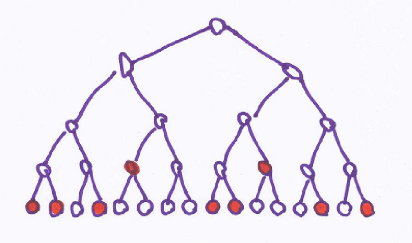
This means that no image is the prefix of any other image.

- The image of x_1^n is called the codeword for x_1^n.
- Let $L_n(x_1^n)$ denote the length of the codeword for x_1^n.
- We have Kraft's inequality:

$$E[2^{-L_n(X_1^n)}] \leq 1 \ .$$

Kraft's inequality

Proof of Kraft's inequality, $E[2^{-L_n(X_1^n)}] \leq 1$.



Barron's lemma

- Let $\{c(n)\}$ be positive constants with $\sum 2^{-c(n)} < \infty$.
- Barron's lemma says we have:

$$L_n(X_1^n) \geq -\log p(X_1^n | X_{-\infty}^0) - c(n), \text{ eventually, a.s.}$$

- A consequence, from the ergodic theorem, is that:

$$\liminf_n \frac{L_n(X_1^n)}{n} \geq \eta \quad \text{a.s.}$$

- This may be called a first order converse source coding theorem.

Proof of Barron's lemma

$$P(L_n(X_1^n) < -\log p(X_1^n|X_{-\infty}^0) - c(n)|X_{-\infty}^0)$$

$$= \sum_{x_1^n} p(x_1^n|X_{-\infty}^0) 1(p(x_1^n|X_{-\infty}^0) < 2^{-L_n(x_1^n)-c(n)})$$

$$\leq \sum_{x_1^n} 2^{-L_n(x_1^n)-c(n)} 1(p(x_1^n|X_{-\infty}^0) < 2^{-L_n(x_1^n)-c(n)})$$

$$\leq 2^{-c(n)} \sum_{x_1^n} 2^{-L_n(x_1^n)}$$

$$\leq 2^{-c(n)} \ .$$

Second order source coding theorems

- By Barron's lemma, we have:

$$L_n(X_1^n) - n\,\eta \geq \left[\sum_{k=1}^{n} -\log p(X_k | X_{-\infty}^{k-1}) - n\,\eta \right] - c(n)$$

  eventually a.s. for any sequence of positive constants with $\sum 2^{-c(n)} < \infty$.
- Thus, if (X_n) is a sufficiently fast mixing process (e.g. a finite-order Markov chain), then

$$\liminf_n \frac{L_{\lfloor nt \rfloor} - nt\eta}{n^{1/2}} \geq W_t$$

  where W_t is a scaled Brownian motion.
- This is the second order converse source coding theorem of Kontoyiannis, 1997.
- For finite-order Markov chains, or if one has sufficiently strong mixing for

$$\max_{x_1} E |\log p(x_1 | X_{-n+1}^0) - \log p(x_1 | X_{-\infty}^0)|\ ,$$

  a matching second order direct source coding theorem holds for most reasonable codes, e.g. Shannon codes, Huffman codes or Lempel-Ziv codes.

Aim of this talk

- We are motivated by the empirical observation of long-range dependence in variable bit rate video traffic, starting with Garrett and Willinger, 1994 and Beran, Sherman, Taqqu, and Willinger, 1995.
- We ask:
  What happens when (X_n) is a long-range dependent process?
  Specifically:
  Can we find a codec to make (L_n) short-range dependent?
- Loosely speaking, we propose the answer: No.
- More precisely, we prove a theorem about long-range-dependent renewal processes that says: No.

Long-range-dependent renewal processes

Known facts from Daley, 1999

Let (X_n) be a renewal process with interarrival times having the distribution of T. Then, for $1 < p < 2$, the following statements are equivalent:

i T has moment index p.

ii (X_n) has Hurst index $H = \frac{1}{2}(3 - p)$.

Here, the moment index of T is defined by:

$$p = \sup\{\kappa \geq 1 \ : \ E[T^\kappa] < \infty\} \ ,$$

and a stationary ergodic process (Z_n) with $E[Z_0^2] < \infty$ is said to have Hurst index H if:

$$H = \inf\{h \ : \ limsup_n \frac{\text{var}(\sum_{k=1}^n Z_k)}{n^{2h}} < \infty\} \ .$$

Main theorem

- Let ρ_n denote $-\log P(X_n | X_{-\infty}^{n-1})$.
- We show that $E[\rho_n^2] < \infty$ and that (ρ_n) has the same Hurst index as (X_n).
- By Barron's lemma, this, in principle, gives second-order converse source coding theorems for long-range-dependent renewal processes.

Main theorem

Let (X_n) be a renewal process with interarrival times having the distribution of T. Then, for $1 < p < 2$, the following statements are equivalent:

i T has moment index p.
ii (X_n) has Hurst index $H = \frac{1}{2}(3 - p)$.
iii (ρ_n) has Hurst index $H = \frac{1}{2}(3 - p)$.

The question of interest for Markov chains

- For a countable state stationary ergodic Markov chain, if any return time has infinite variance, then all return times have the same moment index (Carpio & Daley, 2007).
- For such a chain, we say it has Hurst index H if a (any) return time has moment index corresponding to Hurst index H.
- Our question is of the type:
  When does an instantaneous function of such a Markov chain have the same Hurst index as the chain?
- An example for which this fails (Carpio & Daley 2007):
  Let $M_n^3 = (M_n^1, M_n^2) \in \mathbb{S}_1 x \mathbb{S}_2$. where (M_n^1) has Hurst index $\frac{1}{2} < H < 1$, while (M_n^2) has return times with finite variance. Then (M_n^3) will inherit the Hurst index of (M_n^1). However, instantaneous functions of M_3^n that only depend on M_n^2 will produce processes with Hurst index $\frac{1}{2}$.

The relevant facts from Carpio & Daley

Carpio & Daley, 2007

$$\frac{\mathrm{var}(\rho_0 + \ldots + \rho_n) - (n+1)\mathrm{var}(\rho_0)}{2R_{11}^{(n)}/\pi_1} = \sum_i \sum_j \rho(i)\rho(j)\pi_i\pi_j \frac{R_{ij}^{(n)}/\pi_j}{R_{11}^{(n)}/\pi_1}$$

- π_i denotes the stationary distribution of state i
- $Q_{ij}^{(n)} := \sum_{r=1}^{n}(p_{ij}^{(r)} - \pi_j)$
- $R_{ij}^{(n)} := \sum_{r=1}^{n} Q_{ij}^{(r)}$
- $Q_{ij}^{(n)} \to \infty$
- $\frac{R_{ij}^{(n)}}{n} \to \infty$
- $\frac{R_{ij}^{(n)}/\pi_j}{R_{11}^{(n)}/\pi_1} \to 1$
- $\lim_n \frac{\mathrm{var}(\rho_0+\ldots+\rho_n)}{2R_{11}^{(n)}/\pi_1} \stackrel{?}{=} \sum_i \sum_j \rho(i)\rho(j)\pi_i\pi_j$

A general theorem for instantaneous functions

Theorem

Let (M_n) be a stationary ergodic Markov chain, taking values in $\mathbb{N}$, with interarrival times having infinite variance. Let $\rho : \mathbb{N} \to \mathbb{R}$ is such that $\sum_{i \in \mathbb{N}} \pi_i \rho(i)^2 < \infty$. Write ρ_n for $\rho(M_n)$. If

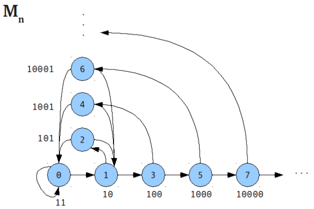
$$\lim_{i \to \infty} \rho(i) = \mu \ ,$$

with

$$\mu \neq \sum_i \pi_i \rho(i) \ ,$$

then (ρ_n) has the same Hurst index as (M_n).

Applying the theorem to a renewal process

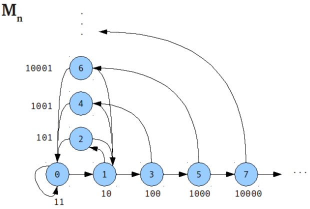


X_n	1	1	0	0	1	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0
M_n	-	0	1	3	4	1	3	5	7	8	0	0	1	3	5	6	1	3	5	7	9

- $\rho(0) = -\log P(T = 1)$.
- $\rho(2k-1) = -\log P(T > k | T \geq k)$.
- $\rho(2k) = -\log P(T = k+1 | T \geq k+1)$.
- $\pi(2k) = \frac{P(T=k+1)}{E[T]}$, $k = 0, 1, 2, \ldots$.
- $\pi(2k-1) = \frac{P(T \geq k+1)}{E[T]}$, $k = 1, 2, \ldots$.

- We can show that $\sum_i \pi(i)\rho(i)^2 < \infty$.

The theorem is not good enough



X_n	1	1	0	0	1	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0
M_n	-	0	1	3	4	1	3	5	7	8	0	0	1	3	5	6	1	3	5	7	9

- Consider the basic renewal process (X_n) with $P(T > k) = k^{-\alpha}L(k)$. $1 < \alpha < 2$, where $L(k)$ is slowly varying.
- Then $\rho(i) \to 0$ when moving through the odd states
- However, $\rho(i) \to \infty$ when moving through the even states.
- The theorem is not directly applicable, even to this basic example.

A souped-up theorem

Theorem

With (M_n) and assumptions as before, let $A \subseteq \mathbb{N}$ be subset of states such that the functions

$$i \to \frac{1}{Q_{11}^{(n)}} 1(i \in A)|\rho(i) - \mu| \sum_{r=1}^{n} \sum_{j \in A} {}_1p_{ij}^{(r)} |\rho(j) - \mu|$$

are uniformly integrable with respect to the probability distribution $i \mapsto \pi_i$ and that $\frac{1}{Q_{11}^{(n)}} \sum_{i \in A} \pi_i (\rho(i) - \mu) \sum_{r=1}^{n} \sum_{j \in A} {}_1p_{ij}^{(r)} (\rho(j) - \mu)$ converges to 0 as $n \to \infty$. If

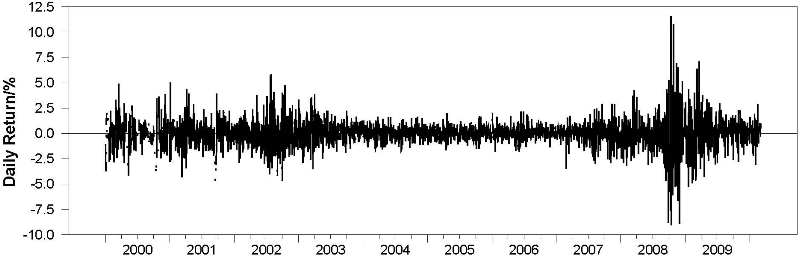
$$\lim_{K \to \infty} \sup_{i \in A^c \cap \{i > K\}} |\rho(i) - \mu| = 0$$

and

$$\mu \neq \sum_i \pi_i \rho(i) ,$$

then (ρ_n) has the same Hurst index as (M_n).

Financial time series

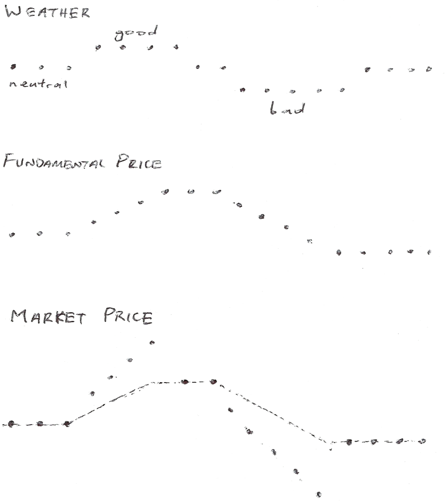


- $r_n = \log \frac{P_n}{P_{n-1}}$, is called the log returns, where P_n is the price of some asset.
- r_n is well modeled by a Martingale difference process, due to the efficient market hypothesis.
- The absolute returns $|r_n|^d$ have been empirically shown to exhibit long memory.

Mandelbrot's model for wheat prices

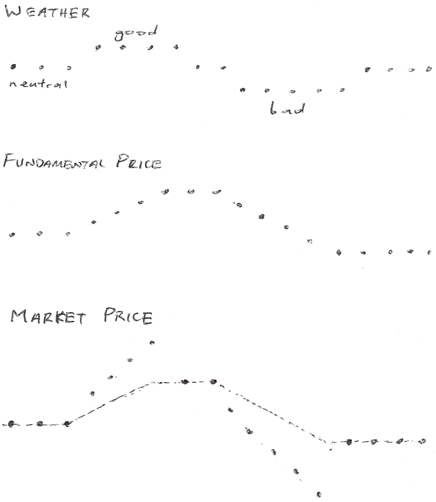
- Can a simple model account for this observation?
- Mandelbrot's model for wheat prices
- Weather has runs of good/neutral/bad days.
- Good/bad period is followed by neutral period (and visa versa)
- The 'fundamental price', $\hat{X}_n$, increases by 1 on good days, decreases by 1 on bad days, unaltered on neutral days.

Mandelbrot's model for wheat prices



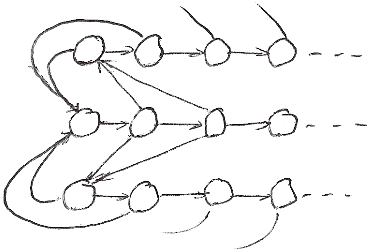
- Distribution of length of each period $f(T > t) = t^{-\alpha}$
- Market calculates $X_n = \lim_{t\to\infty} E[\hat{X}_{n+t}|\hat{X}_{-\infty}^{n-1}]$.
- By construction X_n is a Martingale.

Mandelbrot's model for wheat prices



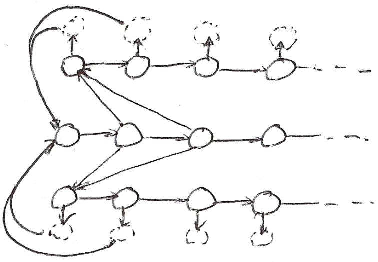
- X_n changes as follows: increases by $\frac{\alpha}{\alpha-1}$ for every good day,
- Decreases by $\frac{\alpha}{\alpha-1}$ for every bad day.
- The first neutral following t good days decreases X_n by $\frac{t}{\alpha-1}$.
- The first neutral following t bad days increases X_n by $\frac{t}{\alpha-1}$.
- The price is unchanged for the following neutral days.

The Markov chain



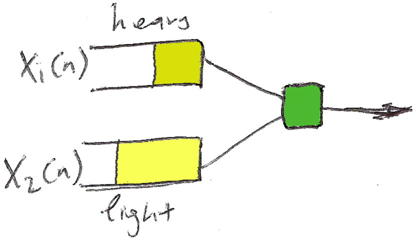
- The weather can be modeled as a MC.
- The differences in fundamental price is a function of this chain (Good = '1', Bad='-1', Neutral='0')

The Markov chain



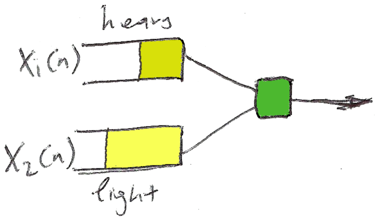
- For the Market price returns (r_n), we need to also know where we jumped from.
- r_n takes values $\pm\frac{\alpha}{\alpha-1}, 0, \pm\frac{t}{\alpha-1}$, where t the in number of days preceding the jump.

What is the Hurst index of $|r_n|^d$?



- X_1 i.i.d. with heavy tails.
- X_2 i.i.d. with light tails (or a SRD Markov process)
- Server has unit capacity

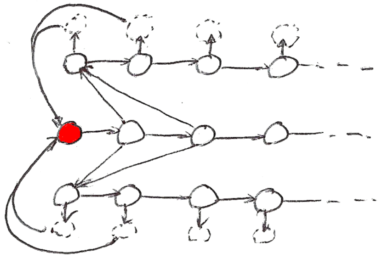
LRD behavior under LQF



- Let M_1, M_2 be MCs representing the two sources.
- $(M_1(n), M_2(n), Q_1(n), Q_2(n))$ is a MC (under any queue length based scheduling).
- Busy-idle function $1(Q_1(n) = 0)$ is LRD.

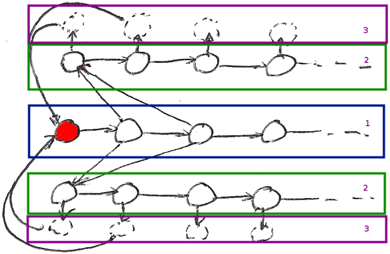
- Is $1(Q_2(n) = 0)$ LRD under LQF?

Solution: financial time series



- State 1 (red).

Solution: financial time series

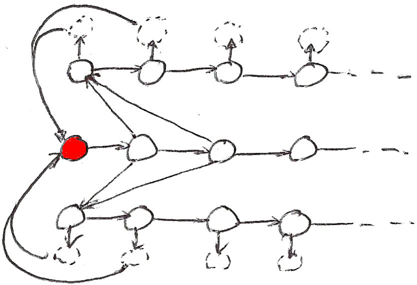


- State 1 (red).
- $c_1 = 0, c_2 = \frac{\alpha}{\alpha - 1}, c_3 = 0$
- In groups 1,2, $\rho - c_k = 0$. In group 3, ${}_1p_{ij} = 0$. (all returns are through state 1).
- Therefore the condition

$$\lim_{n\to\infty} \frac{1}{Q_{11}^{(n)}/\pi_1} \sum_{r=1}^{n} \sum_{i,j\in\mathcal{A}_k} \pi_i |(\rho(i) - c_k)(\rho(j) - c_k)|{}_1p_{ij}^{(r)} = 0 \quad \forall k$$

is satisfied.

Solution: financial time series



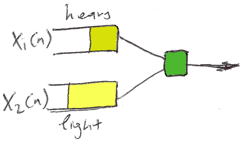
- We get

$$\lim_{n\to\infty} \frac{\text{var}(\sum_{r=1}^{n} \rho_i)}{R_{11}^{(n)}/\pi_1} = \frac{2}{3}(\mu - \frac{\alpha}{\alpha - 1})^2 + \frac{1}{3}\mu^2 > 0$$

  for any d for which $|r_n|^d$ has finite variance. ($d < \alpha/2$)

- $|r_n|^d$ has Hurst index $H = \frac{1}{2}(3 - \alpha)$.

Solution: LQF scheduling



- We took $\rho(n) = 1(Q_2(n) = 0)$
- It is enough to verify

$$\lim_{n\to\infty} \frac{1}{Q_{11}^{(n)}/\pi_1} \sum_{i,j:Q_2=0} \pi_i \sum_{r=1}^{n} {}_1p_{ij}^{(r)} = 0$$

- Note that $\sum_{i,j:Q_2=0} \pi_i \sum_{r=1}^{n} {}_1p_{ij}^{(r)}$ is the stationary time spent in the states $\{Q_2 = 0\}$ before the chain visits $(0, 0, 0, 0)$.
- Idle slot for Q_2 is exponentially distributed.
- Idle slot beginning at time n implies $Q_1(n-1) = 0$, $M_2(n) \leq 1$.
- With each idle period, there is a positive chance (namely $P(M_1(n) = 0, M_2(n) = 0)$) independent of what happened previously, that the chain visits $(0, 0, 0, 0)$.
- Thus, there are at most exponentially many idle periods of Q_2 before going to $(0, 0, 0, 0)$.
- Conclude $1(Q_2(n) = 0)$ has the same Hurst index as the chain $(M_1(n), M_2(n), Q_1(n), Q_2(n))$, which is determined by the tail index of X_1.