Reconstruction of processes with long-range dependence

Rui Vilela Mendes
CMAF and IPFN
http://label2.ist.utl.pt/vilela/

Outline

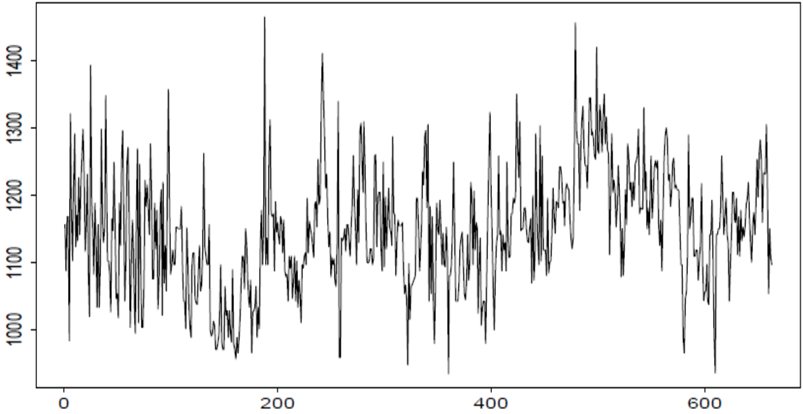
- What is long-range dependence?
- The Hurst exponent
- Tools and examples
    1. Gibbs measures and chains with complete connections
    2. Fractional processes
    3. $\epsilon-$machines

What is long-range dependence? Why is it important?

- **Fields:**
  Internet modelling
  Finance
  Climate studies
  Econometrics
  Hydrology
  Lingustics
  DNA sequencing
- **Issues:**
  Detection of long memory in the data
  Statistical estimation of parameters of long range dependence
  Limit theorems under long range dependence
  Simulation of long memory processes
  Scaling and fractal behavior
- Historically the first study that emphasized long range dependence was in hydrology (Hurst, 1951)

What is long-range dependence?

The anual minima of the water level in the Nile river



What is long-range dependence?

- Hurst R/S statistics. Let $S_i = \sum_{k=1}^{i} X_k$

$$\frac{R}{S}\left(X_1, \cdots, X_n\right) = \frac{\max_{0\leq i\leq n}\left(S_i - \frac{i}{n}S_n\right) - \min_{0\leq i\leq n}\left(S_i - \frac{i}{n}S_n\right)}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}S_n\right)^2\right)^{1/2}}$$

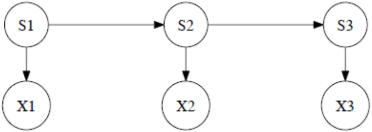
  Grows as $n^{0.74}$, not as $n^{0.5}$ as it would for independent events.
- **Long range dependence and the sum of autocovariances** Consider a stationary process with $\mathbb{E}\left(X\right) = 0$ and $\gamma\left(k\right) = \mathbb{E}\left(X_i X_{i+k}\right)$. If $\sum_{k=-\infty}^{\infty}\gamma\left(k\right)$diverges the process has **long-range dependence** (or long memory)
- Other notions

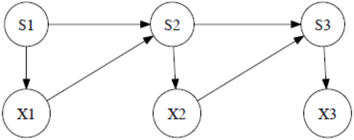
$$\begin{array}{rclcll}
\sum_{k=-n}^{n}\gamma\left(k\right) & \sim & n^{\alpha}L_1\left(n\right) & \text{as} & n\to\infty & 0<\alpha<1 \\
\gamma\left(k\right) & \sim & k^{-\beta}L_2\left(k\right) & \text{as} & k\to\infty & 0<\beta<1 \\
f\left(\nu\right) & \sim & \left|\nu\right|^{-\beta}L_3\left(\left|\nu\right|\right) & \text{as} & \nu\to 0 & 0<\beta<1
\end{array}$$

  L_1, L_2 and L_3 are slowly varying functions and $f\left(\nu\right) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{-i\nu k}\gamma\left(k\right)$ is the spectral density

What is long-range dependence?

- Long-range dependence corresponds to a slow decay of the autocorrelations or to large values of the spectral density near zero.
- Long-range dependence is sometimes associated to the self-similarity of the processes. In fact **some** self-similar processes have long-range dependence, but the two notions are independent.
- In the context of Hidden Markov Processes or Chains with Complete Connections there are also the notions of *Markov order*, *Cripticity order*, *Synchronization order*. When one of these is ∞ the process also has long memory





What is long-range dependence?

- Modeling tools

  - Gibbs measures

  - Chains with complete connections

  - Fractional stochastic processes

  - $\varepsilon-$machines

Statistical analysis of time series: Gibbs measures and Chains with complete connections

- Time series $\cdots X_{-2}X_{-1}X_0X_1X_2\cdots$
  $X \in Y$: the *state space*
  $Y^{\mathbb{Z}}$: the *path space*

Statistical analysis of time series: Gibbs measures and Chains with complete connections

- Time series $\cdots X_{-2} X_{-1} X_0 X_1 X_2 \cdots$
  $X \in Y$: the *state space*
  $Y^{\mathbb{Z}}$: the *path space*
- **Statistical properties: (3 levels)**
  (1) Expectation values of the observables
  (2) Probability measures on state space Y
  (3) Probability measures on path space Y^Z

Statistical analysis of time series: Gibbs measures and Chains with complete connections

- Time series $\cdots X_{-2}X_{-1}X_0X_1X_2\cdots$
  $X \in Y$: the *state space*
  $Y^{\mathbb{Z}}$: the *path space*
- **Statistical properties: (3 levels)**
  (1) Expectation values of the observables
  (2) Probability measures on state space Y
  (3) Probability measures on path space Y^{Z}
- *Level 1, 2 and 3- statistical indicators.*
  (*Mean partial sums, empirical measures* (pdf's) and *empirical process*)

Statistical analysis of time series: Gibbs measures and Chains with complete connections

- Time series $\cdots X_{-2}X_{-1}X_0X_1X_2\cdots$
  $X \in Y$: the *state space*
  $Y^{\mathbb{Z}}$: the *path space*
- **Statistical properties: (3 levels)**
  (1) Expectation values of the observables
  (2) Probability measures on state space Y
  (3) Probability measures on path space Y^Z
- *Level 1, 2 and 3- statistical indicators.*
  (*Mean partial sums, empirical measures* (pdf's) and *empirical process*)
- **Analysis and reconstruction of the process:**
  Purpose: To extract *Grammar* and *measure*

Statistical analysis of time series: Gibbs measures and Chains with complete connections

- Time series $\cdots X_{-2}X_{-1}X_0X_1X_2\cdots$
  $X \in Y$: the *state space*
  $Y^{\mathbb{Z}}$: the *path space*
- **Statistical properties: (3 levels)**
  (1) Expectation values of the observables
  (2) Probability measures on state space Y
  (3) Probability measures on path space Y^Z
- *Level 1, 2 and 3- statistical indicators.*
  (*Mean partial sums, empirical measures* (pdf's) and *empirical process*)
- **Analysis and reconstruction of the process:**
  Purpose: To extract *Grammar* and *measure*
- Examples:
  Hydrodynamic turbulence
  Market fluctuations:
  (there are analogies but the statistical indicators are different)

Statistical analysis of time series data: An example

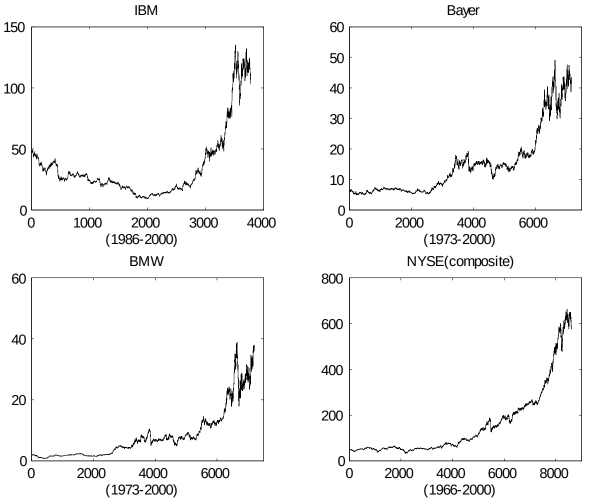
- Application of statistical tools requires:
  (i) Stationary or asymptotically stationary process
  (ii) Typical samples

Statistical analysis of time series data: An example

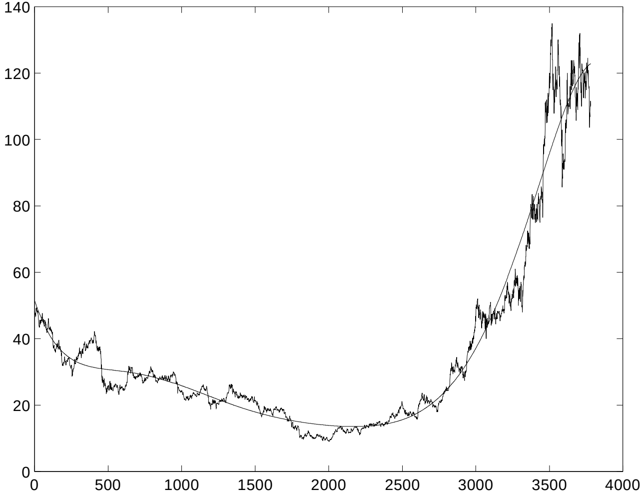
- Application of statistical tools requires:
  (i) Stationary or asymptotically stationary process
  (ii) Typical samples
- Market stocks as experimental probes revealing the mechanisms of the market process
  (i) $\Rightarrow$ preprocessing of the data

Statistical analysis of time series data

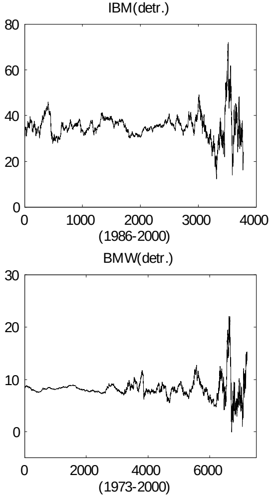
Daily data $p(t)$



Detrending by a polynomial

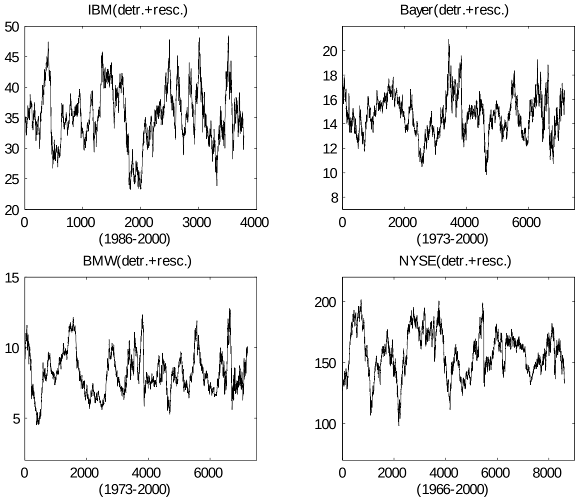


Detrended data $p(t) - q(t)$

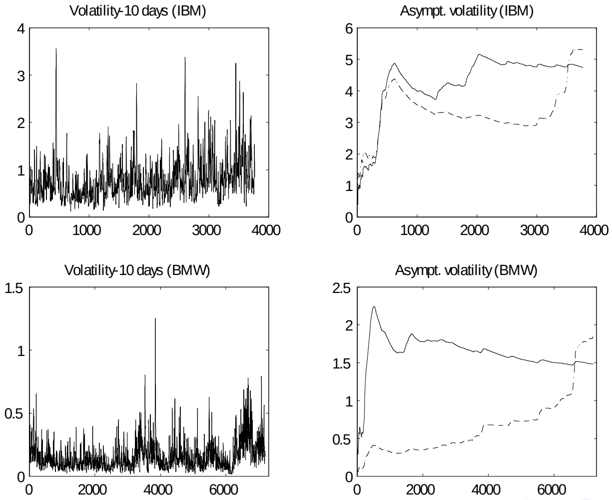


Statistical analysis of time series data

Detrended and rescaled data $x(t) = (p(t) - q(t))\,\frac{<p(t)>}{q(t)}$

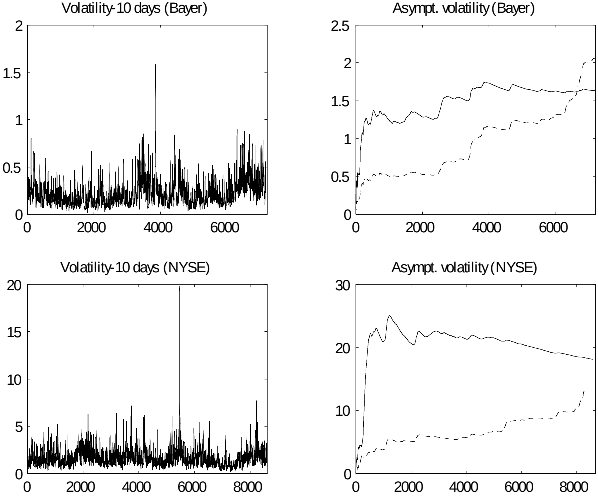


Ten-days window volatility and comparison of asymptotic volatility for the rescaled and non-rescaled data (IBM and BMW)



Statistical analysis of time series data

Ten-days window volatility and comparison of asymptotic volatility for the rescaled and non-rescaled data (Bayer and NYSE)



Statistical indicators

- $n-$days return

$$r(t,n) = \log p\,(t+n) - \log p\,(t)$$

Statistical indicators

- $n-$days return

$$r(t,n) = \log p\,(t+n) - \log p\,(t)$$

- (i) *Maximum* (over t) of $r(t,n)$

$$\delta\,(n) = \max_t \{r(t,n)\}$$

Statistical indicators

- $n-$days return

$$r(t,n) = \log p\,(t+n) - \log p\,(t)$$

- (i) *Maximum* (over t) of $r(t,n)$

$$\delta\,(n) = \max_t \{r(t,n)\}$$

- (ii) *Moments* of the distribution of $|r(t,n)|$

$$S_q(n) = \langle |r(t,n)|^q \rangle$$

Statistical indicators

- $n-$days return

$$r(t,n) = \log p\,(t+n) - \log p\,(t)$$

- (i) *Maximum* (over t) of $r(t,n)$

$$\delta\,(n) = \max_t \{r(t,n)\}$$

- (ii) *Moments* of the distribution of $|r(t,n)|$

$$S_q(n) = \langle |r(t,n)|^q \rangle$$

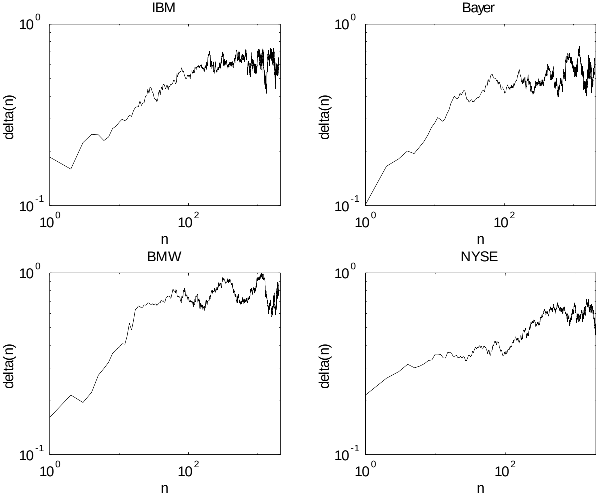
- (iii) If in some range ($n = 2$ to $n = 60$)

$$S_q(n) \sim n^{\chi(q)}$$

$\chi(q)$ is the *scaling exponent*

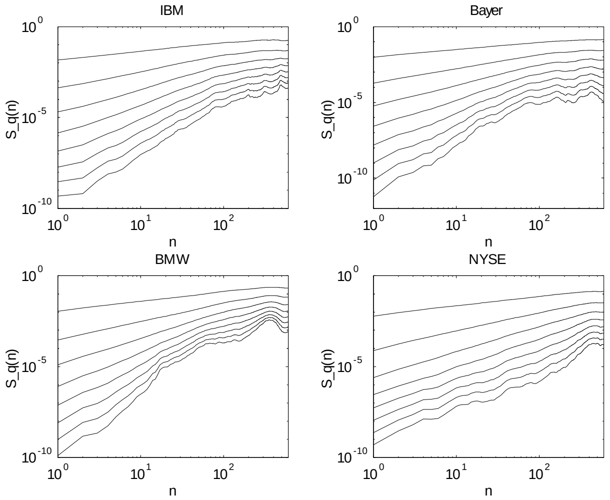
Statistical indicators

Maximum $\delta(n)$ of log-prices differences



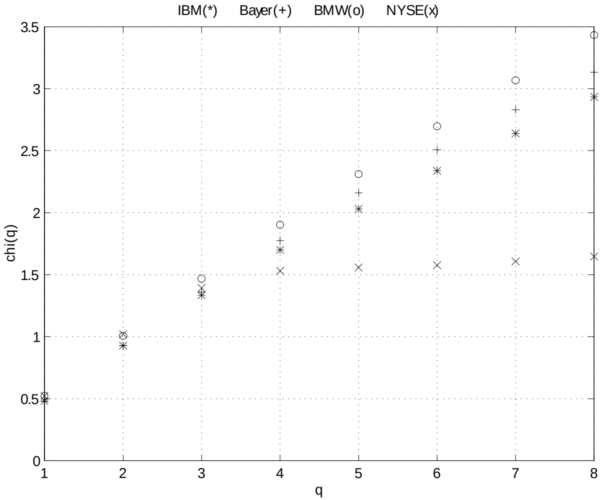
Statistical indicators

Moments of the $|r(t,n)|$ distribution



Statistical indicators

Scaling exponent $\chi(q)$



Statistical indicators

- Main conclusions:
  (a) $\delta(n)$ is log-concave and probably asymptotically constant for large r
  (b) $S_q(n)$ is a log-concave function of n with an inertial range
  (c) The scaling law $\chi(q)$ is an increasing concave function of q
  (d) $\chi(1)$ in the scaling region ($n = 2$ to $n = 60$) is close to 0.5
  (e) Scaling properties of NYSE somewhat different from the others

Statistical indicators

- Main conclusions:
  (a) $\delta(n)$ is log-concave and probably asymptotically constant for large r
  (b) $S_q(n)$ is a log-concave function of n with an inertial range
  (c) The scaling law $\chi(q)$ is an increasing concave function of q
  (d) $\chi(1)$ in the scaling region ($n = 2$ to $n = 60$) is close to 0.5
  (e) Scaling properties of NYSE somewhat different from the others

- Properties (a) to (c) are shared by fluid turbulence data, but with different values for the statistical indicators (in turbulence data $\chi(1) = \frac{1}{3}$, here $\chi(1) \approx 0.5 \Rightarrow$ essentially uncorrelated signal for $n \geq 2$)

Statistical indicators

- Main conclusions:
  (a) $\delta(n)$ is log-concave and probably asymptotically constant for large r
  (b) $S_q(n)$ is a log-concave function of n with an inertial range
  (c) The scaling law $\chi(q)$ is an increasing concave function of q
  (d) $\chi(1)$ in the scaling region ($n = 2$ to $n = 60$) is close to 0.5
  (e) Scaling properties of NYSE somewhat different from the others

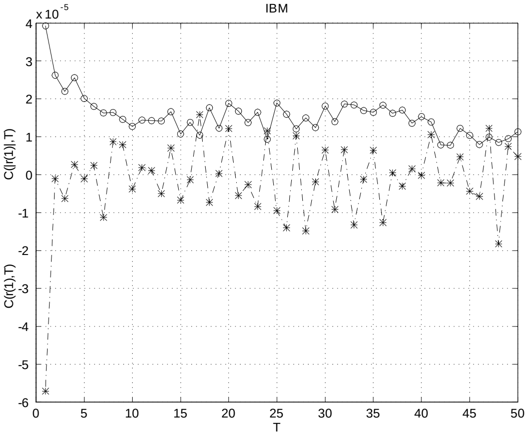
- Properties (a) to (c) are shared by fluid turbulence data, but with different values for the statistical indicators (in turbulence data $\chi(1) = \frac{1}{3}$, here $\chi(1) \approx 0.5 \Rightarrow$ essentially uncorrelated signal for $n \geq 2$)

- The behavior of the statistical indicators $\delta(n)$, $S_q(n)$ and $\chi(q) \Longrightarrow$ If the process is a topological Markov chain the transitions allowed by the transition matrix T must lie inside a strictly convex domain around the diagonal of T

Statistical indicators

Correlation function of one-day returns ($\star$) and its absolute value (o)

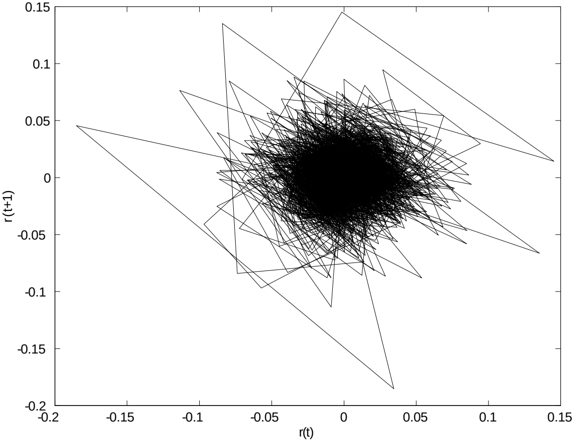
$$C\left(r(1),T\right)=\left\langle r\left(t+T,1\right)r\left(t,1\right)\right\rangle$$
$$C\left(\left|r(1)\right|,T\right)=\left\langle \left|r\left(t+T,1\right)\right|\left|r\left(t,1\right)\right|\right\rangle$$



Statistical indicators

Dynamics of one-day returns

$$r\left(t,1\right)\rightarrow r\left(t+1,1\right)$$



Why are Gibbs measures "natural" ?

- Events $\{X_i\}$
  $p_i =$ probability of event X_i

Looking for a Gibbs measure

Why are Gibbs measures "natural" ?

- Events $\{X_i\}$
  p_i = probability of event X_i
- Constraints:
  (i) Normalization, $\sum_i p_i = 1$
  (ii) Expectation value of known observables, $\sum_i p_i F_k(X_i) = C_k$

Why are Gibbs measures "natural" ?

- Events $\{X_i\}$
  $p_i =$ probability of event X_i
- Constraints:
  (i) Normalization, $\sum_i p_i = 1$
  (ii) Expectation value of known observables, $\sum_i p_i F_k(X_i) = C_k$
- Maximum entropy principle (to maximize the uncertainty about what is not known) = use the most unbiased estimate

$$S = -\sum_i p_i \log p_i + \lambda_0 \sum_i p_i + \sum_k \lambda_k \sum_i p_i F_k(X_i)$$

$$\frac{\partial S}{\partial p_i} = 0 \qquad \Longrightarrow \qquad -\log p_i - 1 + \lambda_0 + \sum_k \lambda_k F_k(X_i) = 0$$

$$p_i = \exp\left(-1 + \lambda_0 + \sum_k \lambda_k F_k(X_i)\right)$$

with $\lambda_0, \lambda_1, \cdots$ obtained from the constraints

Looking for a Gibbs measure

- Coding by a finite alphabet Σ
  Space Ω of orbits $\omega = i_1 i_2 \cdots i_k \cdots$, $i_k \in \Sigma$
  Dynamical law: a shift σ

$$\sigma\omega = i_2 \cdots i_k \cdots$$

Looking for a Gibbs measure

- Coding by a finite alphabet Σ
  Space Ω of orbits $\omega = i_1 i_2 \cdots i_k \cdots$, $i_k \in \Sigma$
  Dynamical law: a shift σ

$$\sigma\omega = i_2 \cdots i_k \cdots$$

- Grammar: set of allowed sequences in Ω

Looking for a Gibbs measure

- Coding by a finite alphabet Σ
  Space Ω of orbits $\omega = i_1 i_2 \cdots i_k \cdots$, $i_k \in \Sigma$
  Dynamical law: a shift σ

$$\sigma\omega = i_2 \cdots i_k \cdots$$

- Grammar: set of allowed sequences in Ω
- Sequences which coincide on the first n symbols: $n-$*cylinder* (or $n-$block) denoted by $[i_1 i_2 \cdots i_n]$
  Probability measures over the cylinders

Looking for a Gibbs measure

- Coding by a finite alphabet Σ
  Space Ω of orbits $\omega = i_1 i_2 \cdots i_k \cdots$, $i_k \in \Sigma$
  Dynamical law: a shift σ

$$\sigma\omega = i_2 \cdots i_k \cdots$$

- Grammar: set of allowed sequences in Ω
- Sequences which coincide on the first n symbols: $n-$*cylinder* (or $n-$block) denoted by $[i_1 i_2 \cdots i_n]$
  Probability measures over the cylinders
- **Gibbs measure**

$$c_1 \leq \frac{\mu\left(\left[i_1(\omega) i_2(\omega) \cdots i_n(\omega)\right]\right)}{\exp\left(-nP + \left(S_n\phi\right)(\omega)\right)} \leq c_2$$

$\left(S_n\phi\right)(\omega) = \sum_{k=0}^{n-1} \phi\left(\sigma^k \omega\right)$, ϕ being Hölder continuous function on Ω (the *potential)*
$P(\phi, G)$: a function depending on potential and grammar (the *pressure* of ϕ)

Looking for a Gibbs measure

- Relation to the entropy

$$h(\mu) = \lim_{n\to\infty} \frac{H_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i_1 \cdots i_n} \mu([i_1 i_2 \cdots i_n]) \log \mu([i_1 i_2 \cdots i_n])$$

Looking for a Gibbs measure

- Relation to the entropy

$$h(\mu) = \lim_{n\to\infty} \frac{H_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i_1\cdots i_n} \mu([i_1 i_2 \cdots i_n]) \log \mu([i_1 i_2 \cdots i_n])$$

- Variational principle: for each potential and grammar, $\sup_\eta \left\{ h(\eta) + \int \phi d\eta \right\}$ over all $\sigma-$invariant measures η is reached only for the Gibbs measure μ and equals the pressure

$$P(\phi, G) = h(\mu) + \int \phi d\mu$$

Looking for a Gibbs measure

- Relation to the entropy

$$h\left(\mu\right)=\lim_{n\to\infty}\frac{H_n}{n}=\lim_{n\to\infty}\frac{1}{n}\sum_{i_1\cdots i_n}\mu\left(\left[i_1 i_2\cdots i_n\right]\right)\log\mu\left(\left[i_1 i_2\cdots i_n\right]\right)$$

- Variational principle: for each potential and grammar, $\sup_\eta\left\{h\left(\eta\right)+\int\phi d\eta\right\}$ over all $\sigma-$invariant measures η is reached only for the Gibbs measure μ and equals the pressure

$$P\left(\phi,G\right)=h\left(\mu\right)+\int\phi d\mu$$

- Potential may be chosen such that $P=0$ (*normalized potential*). Then

$$\phi\left(\omega\right)=\lim_{n\to\infty}\log\frac{\mu\left(\left[i_1\left(\omega\right)\cdots i_n\left(\omega\right)\right]\right)}{\mu\left(\left[i_2\left(\omega\right)\cdots i_n\left(\omega\right)\right]\right)}$$

(Practical use hindered by poor statistics of large blocks)

Looking for a Gibbs measure

- **Gibbs measures for finite range potentials**
  (finite range potentials approximate uniformly any Hölder continuous potential)

Looking for a Gibbs measure

- **Gibbs measures for finite range potentials**
  (finite range potentials approximate uniformly any Hölder continuous potential)
- Property of range$-r$ potentials: for all values $i_1 i_2 \cdots i_n$ with $n \geq r$

$$\mu\left([i_1 \cdots i_n]\right) = \frac{\mu\left([i_1 \cdots i_r]\right) \times \cdots \times \mu\left([i_{n-r+1} \cdots i_n]\right)}{\mu\left([i_2 \cdots i_r]\right) \times \cdots \times \mu\left([i_{n-r+1} \cdots i_{n-1}]\right)} \qquad (1)$$

$\Longrightarrow$

$$h(\mu) = -\sum_{i_1 \cdots i_k} \mu\left([i_1 \cdots i_k]\right) \log \frac{\mu\left([i_1 \cdots i_k]\right)}{\mu\left([i_1 \cdots i_{k-1}]\right)} = H_k - H_{k-1}$$

for all $k \geq r$ if $r > 1$. If $r = 1$
$h(\mu) = H_1 H_k = -\sum_{i_1 \cdots i_k} \mu\left([i_1 \cdots i_k]\right) \log \mu\left([i_1 \cdots i_k]\right)$

Looking for a Gibbs measure

- **Gibbs measures for finite range potentials**
  (finite range potentials approximate uniformly any Hölder continuous potential)
- Property of range$-r$ potentials: for all values $i_1 i_2 \cdots i_n$ with $n \geq r$

$$\mu([i_1 \cdots i_n]) = \frac{\mu([i_1 \cdots i_r]) \times \cdots \times \mu([i_{n-r+1} \cdots i_n])}{\mu([i_2 \cdots i_r]) \times \cdots \times \mu([i_{n-r+1} \cdots i_{n-1}])} \quad (1)$$

$\Longrightarrow$

$$h(\mu) = -\sum_{i_1 \cdots i_k} \mu([i_1 \cdots i_k]) \log \frac{\mu([i_1 \cdots i_k])}{\mu([i_1 \cdots i_{k-1}])} = H_k - H_{k-1}$$

for all $k \geq r$ if $r > 1$. If $r = 1$
$h(\mu) = H_1 H_k = -\sum_{i_1 \cdots i_k} \mu([i_1 \cdots i_k]) \log \mu([i_1 \cdots i_k])$

- $\Longrightarrow$ **criterium to find the range of the potential**: range of the potential found when $H_k - H_{k-1}$ tends to a constant.
  Once the range is found, the potential may be constructed from the empirical weights $\widetilde{\mu}([i_1 \cdots i_k])$.

Looking for a Gibbs measure

- Another consequence of (1) is that for $k > r$

$$\mu\left(\left[i_1 \cdots i_{k+1}\right]\right) = \frac{\mu\left(\left[i_1 \cdots i_k\right]\right) \mu\left(\left[i_2 \cdots i_{k+1}\right]\right)}{\mu\left(\left[i_2 \cdots i_k\right]\right)}$$

Looking for a Gibbs measure

- Another consequence of (1) is that for $k > r$

$$\mu([i_1 \cdots i_{k+1}]) = \frac{\mu([i_1 \cdots i_k])\,\mu([i_2 \cdots i_{k+1}])}{\mu([i_2 \cdots i_k])}$$

- **Application to the market fluctuations**:
  Five-symbols code $\Sigma = \{-2, -1, 0, 1, 2\}$ for

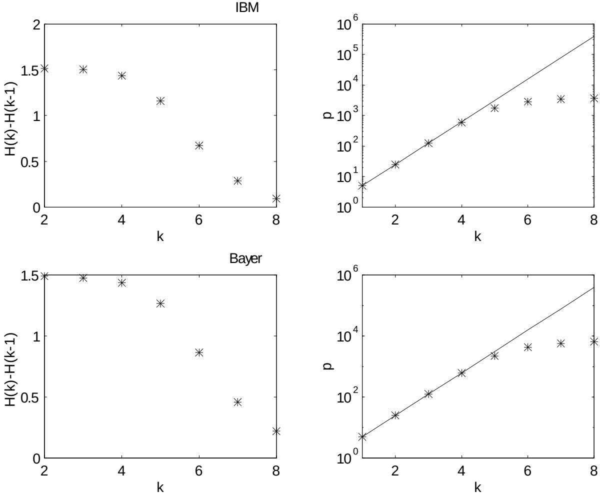
$$r(t) = \log p(t+1) - \log p(t)$$

Average $\overline{r(t)}$ and standard deviation $s = \sqrt{\overline{\left(r^2(t) - \overline{r(t)}^2\right)}}$

$$\begin{array}{rcl}
\left(r(t) - \overline{r(t)}\right) > s & \Longleftrightarrow & 2 \\
s \geq \left(r(t) - \overline{r(t)}\right) > \frac{s}{3} & \Longleftrightarrow & 1 \\
\frac{s}{3} \geq \left(r(t) - \overline{r(t)}\right) > -\frac{s}{3} & \Longleftrightarrow & 0 \\
-\frac{s}{3} \geq \sigma\left(r(t) - \overline{r(t)}\right) > -s & \Longleftrightarrow & -1 \\
-s \geq \left(r(t) - \overline{r(t)}\right) & \Longleftrightarrow & -2
\end{array}$$

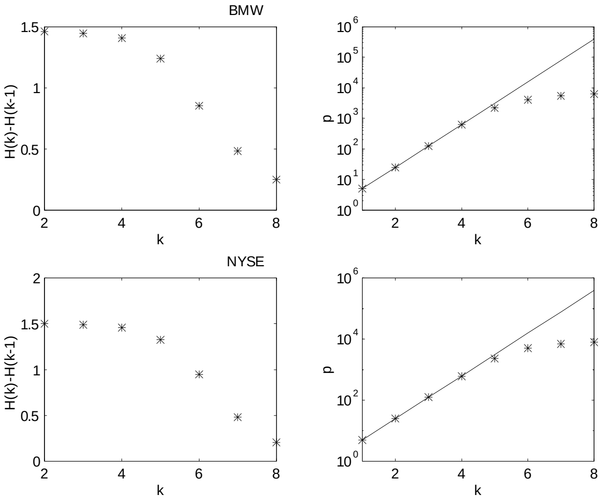
Looking for a Gibbs measure

$H_k - H_{k-1}$ and the number of occurring blocks of size k (IBM and Bayer)



Looking for a Gibbs measure

$H_k - H_{k-1}$ and the number of occurring blocks of size k (BMW and NYSE)



Looking for a Gibbs measure

- Behavior of $H_k - H_{k-1}$ quite different from hydrodynamic turbulence data

Looking for a Gibbs measure

- Behavior of $H_k - H_{k-1}$ quite different from hydrodynamic turbulence data
- To check whether a short-range potential is reliable :
  For successively higher k estimate

$$\mu_e\left(\left[i_1 \cdots i_{k+1}\right]\right) = \frac{\widetilde{\mu}\left(\left[i_1 \cdots i_k\right]\right) \widetilde{\mu}\left(\left[i_2 \cdots i_{k+1}\right]\right)}{\widetilde{\mu}\left(\left[i_2 \cdots i_k\right]\right)}$$

then compare with the empirically observed $\widetilde{\mu}\left(\left[i_1 \cdots i_{k+1}\right]\right)$

Looking for a Gibbs measure

- Behavior of $H_k - H_{k-1}$ quite different from hydrodynamic turbulence data
- To check whether a short-range potential is reliable :
  For successively higher k estimate

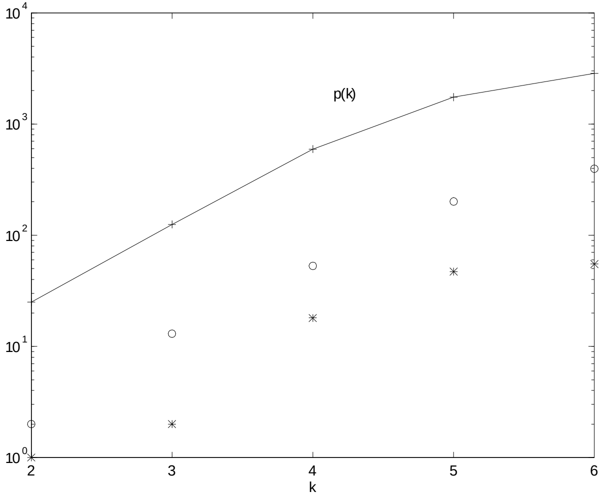
$$\mu_e\left(\left[i_1 \cdots i_{k+1}\right]\right) = \frac{\widetilde{\mu}\left(\left[i_1 \cdots i_k\right]\right) \widetilde{\mu}\left(\left[i_2 \cdots i_{k+1}\right]\right)}{\widetilde{\mu}\left(\left[i_2 \cdots i_k\right]\right)}$$

  then compare with the empirically observed $\widetilde{\mu}\left(\left[i_1 \cdots i_{k+1}\right]\right)$
- Standard deviation of the relative positive errors

$$\varepsilon_k = \max\left(0, \frac{\widetilde{\mu}\left(\left[i_1 \cdots i_{k+1}\right]\right) - \mu_e\left(\left[i_1 \cdots i_{k+1}\right]\right)}{\frac{1}{2}\left(\widetilde{\mu}\left(\left[i_1 \cdots i_{k+1}\right]\right) + \mu_e\left(\left[i_1 \cdots i_{k+1}\right]\right)\right)}\right)$$

Looking for a Gibbs measure

Underestimation errors one (o) and two (*) standard deviations away from the mean and the total number $p\,(k)$ of observed blocks



Looking for a Gibbs measure

- Large number of large deviation errors
  Correspond to blocks involving 2 and -2
  Large deviations misrepresented by empirically constructed measure

Looking for a Gibbs measure

- Large number of large deviation errors
  Correspond to blocks involving 2 and -2
  Large deviations misrepresented by empirically constructed measure
- ?? $\Rightarrow$ non-Gibbsian measure
  or
  ?? $\Rightarrow$ Gibbsian measure with long-range potential (sharp rise of $H_k - H_{k-1}$ at $k = 2$ followed by a very slow increase)

Looking for a Gibbs measure

- Large number of large deviation errors
  Correspond to blocks involving 2 and -2
  Large deviations misrepresented by empirically constructed measure
- ?? $\Rightarrow$ non-Gibbsian measure
  or
  ?? $\Rightarrow$ Gibbsian measure with long-range potential (sharp rise of $H_k - H_{k-1}$ at $k = 2$ followed by a very slow increase)
- Large deviation analysis applied to the calculation of H_k consistent with this hypothesis

Looking for a Gibbs measure

- Large number of large deviation errors
  Correspond to blocks involving 2 and -2
  Large deviations misrepresented by empirically constructed measure
- ?? $\Rightarrow$ non-Gibbsian measure
  or
  ?? $\Rightarrow$ Gibbsian measure with long-range potential (sharp rise of $H_k - H_{k-1}$ at $k = 2$ followed by a very slow increase)
- Large deviation analysis applied to the calculation of H_k consistent with this hypothesis
- In any case it needs an approach suited to deal with long-memory processes

Chains with complete connections

- **Chain with complete connections** (CCC)

Chains with complete connections

- **Chain with complete connections** (CCC)

1. $\forall a_i \in \Sigma$

$$P\left(X_1 = a_1, \cdots, X_n = a_n\right) > 0$$

Chains with complete connections

- **Chain with complete connections** (CCC)

1. $\forall a_i \in \Sigma$
$$P\left(X_1 = a_1, \cdots, X_n = a_n\right) > 0$$
2. The limit
$$\begin{aligned} & \lim_{m\to\infty} P\left(X_0 = a_0 | X_j = a_j, -m \leq j \leq -1\right) \\ = \quad & P\left(X_0 = a_0 | X_j = a_j, j \leq -1\right) \end{aligned}$$
exists $\forall a_i, j \leq -1$

Chains with complete connections

- **Chain with complete connections** (CCC)

1. $\forall a_i \in \Sigma$
$$P\left(X_1 = a_1, \cdots, X_n = a_n\right) > 0$$
2. The limit
$$\begin{aligned} & \lim_{m\to\infty} P\left(X_0 = a_0 | X_j = a_j, -m \leq j \leq -1\right) \\ = \quad & P\left(X_0 = a_0 | X_j = a_j, j \leq -1\right) \end{aligned}$$
exists $\forall a_i, j \leq -1$
3. There is a sequence $(\gamma_m)_{m\geq 1}$ with $\lim_{m\to\infty} \gamma_m = 0$, such that for all $\{a_j, b_j \in \Sigma, j \leq -1\}$ with $a_j = b_j$ for $-m \leq j \leq -1$
$$\left|\left(\frac{P\left(X_0 = a_0 | X_j = a_j, j \leq -1\right)}{P\left(X_0 = a_0 | X_j = b_j, j \leq -1\right)} - 1\right)\right| \leq \gamma_m$$

Looking for a Gibbs measure

- **Chain with complete connections and summable decay** (CCCSD) CCC with $\sum \gamma_m < \infty$

Looking for a Gibbs measure

- **Chain with complete connections and summable decay** (CCCSD) CCC with $\sum \gamma_m < \infty$
- Conditions 1. and 2. implicitly assumed for the pre-processed data Decays γ_m estimated from a typical sample of the process. From the empirical probabilities

$$P\left(a_0 | a_1 \cdots a_m A\right) = \frac{P\left(a_0 a_1 \cdots a_m A\right)}{P\left(a_1 \cdots a_m A\right)}$$

A a block of arbitrary length

$$g\left(a_0 a_1 \cdots a_m\right) = \left(\frac{\max_A P\left(a_0 | a_1 \cdots a_m A\right)}{\min_A P\left(a_0 | a_1 \cdots a_m A\right)} - 1\right)$$

$$\gamma_m = \max_{a_0 a_1 \cdots a_m} g\left(a_0 a_1 \cdots a_m\right)$$

If the statistics for long blocks is poor $\Rightarrow$ large fluctuations in γ_m

Looking for a Gibbs measure

- **Chain with complete connections and summable decay** (CCCSD) CCC with $\sum \gamma_m < \infty$
- Conditions 1. and 2. implicitly assumed for the pre-processed data Decays γ_m estimated from a typical sample of the process. From the empirical probabilities

$$P(a_0 | a_1 \cdots a_m A) = \frac{P(a_0 a_1 \cdots a_m A)}{P(a_1 \cdots a_m A)}$$

A a block of arbitrary length

$$g(a_0 a_1 \cdots a_m) = \left(\frac{\max_A P(a_0 | a_1 \cdots a_m A)}{\min_A P(a_0 | a_1 \cdots a_m A)} - 1 \right)$$

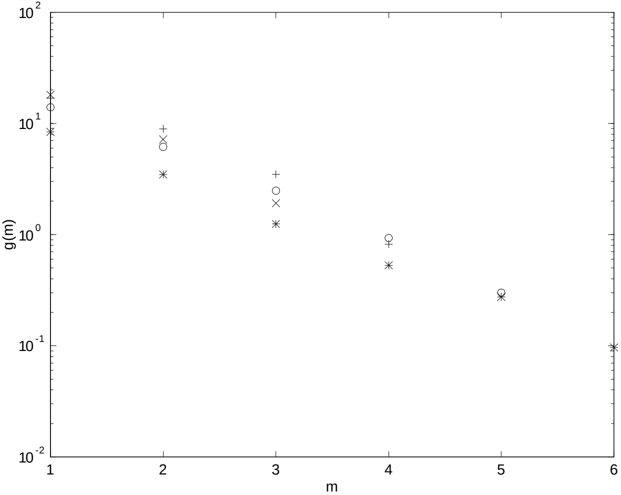
$$\gamma_m = \max_{a_0 a_1 \cdots a_m} g(a_0 a_1 \cdots a_m)$$

If the statistics for long blocks is poor $\Rightarrow$ large fluctuations in γ_m

- Better estimate of the decay behavior with $g(m) = \overline{g(a_0 a_1 \cdots a_m)}$, the average being taken over all sets $a_0 a_1 \cdots a_m$ of size m.

Chains with complete connections

$g(m)$ computed using A blocks of length 5 to 8 ($\times, +, o, *$)



Chains with complete connections

- Result compatible with exponential decay
  $\Rightarrow$ summability of the γ_m's

Chains with complete connections

- Result compatible with exponential decay
  $\Rightarrow$ summability of the γ_m's
- **A CCC-process with summable decays is the $\overline{d}-$limit of its Markov approximations of order k**

Chains with complete connections

- Result compatible with exponential decay
  $\Rightarrow$ summability of the γ_m's
- **A CCC-process with summable decays is the $\overline{d}-$limit of its Markov approximations of order k**
- $\overline{d}-$*distance*
  Coupling between two processes $X = \{X_n\}$ and $Y = \{Y_n\}$ is another process $\left\{\widetilde{X}_n, \widetilde{Y}_n\right\}$ over $\Sigma \times \Sigma$ such that the marginal probabilities of $\widetilde{X}$ and $\widetilde{Y}$ coincide with those of X and Y

$$\overline{d}(X, Y) = \inf \left\{ P\left(\widetilde{X}_0 \neq \widetilde{Y}_0\right) : \left\{\widetilde{X}_n, \widetilde{Y}_n\right\} \begin{array}{l}\text{is a stationary} \\ \text{coupling of X and Y}\end{array} \right\}$$

Chains with complete connections

- Result compatible with exponential decay
  $\Rightarrow$ summability of the γ_m's
- **A CCC-process with summable decays is the $\overline{d}-$limit of its Markov approximations of order k**
- $\overline{d}-$*distance*
  Coupling between two processes $X = \{X_n\}$ and $Y = \{Y_n\}$ is another process $\left\{\widetilde{X}_n, \widetilde{Y}_n\right\}$ over $\Sigma \times \Sigma$ such that the marginal probabilities of $\widetilde{X}$ and $\widetilde{Y}$ coincide with those of X and Y

$$\overline{d}\,(X, Y) = \inf\left\{P\left(\widetilde{X}_0 \neq \widetilde{Y}_0\right) : \left\{\widetilde{X}_n, \widetilde{Y}_n\right\} \begin{array}{l}\text{is a stationary}\\ \text{coupling of X and Y}\end{array}\right\}$$

- $\overline{d}-$distance tending to zero does mean that the processes will coincide after a certain time

Chains with complete connections

- Result compatible with exponential decay
  $\Rightarrow$ summability of the γ_m's
- **A CCC-process with summable decays is the $\overline{d}-$limit of its Markov approximations of order k**
- $\overline{d}-$*distance*
  Coupling between two processes $X = \{X_n\}$ and $Y = \{Y_n\}$ is another process $\left\{\widetilde{X}_n, \widetilde{Y}_n\right\}$ over $\Sigma \times \Sigma$ such that the marginal probabilities of $\widetilde{X}$ and $\widetilde{Y}$ coincide with those of X and Y

$$\overline{d}\left(X, Y\right) = \inf\left\{P\left(\widetilde{X}_0 \neq \widetilde{Y}_0\right) : \left\{\widetilde{X}_n, \widetilde{Y}_n\right\} \begin{array}{l}\text{is a stationary}\\ \text{coupling of X and Y}\end{array}\right\}$$

- $\overline{d}-$distance tending to zero does mean that the processes will coincide after a certain time
- **Perfect simulation** always understood in the $\overline{d}-$distance sense. It does not mean **perfect prediction** (Means that a process is constructed with the same conditional probabilities of the original one)

Chains with complete connections

- **Simulation scheme by the sequence of canonical Markov approximations of finite order k ($k-$CMA)**
  $k-$CMA of a process X is a Markov chain $Y^{(k)}$ of order k with conditional probabilities $P^{(k)}$
  $$P^{(k)}\left(a_0 | a_1 \cdots a_k\right) = P\left(a_0 | a_1 \cdots a_k\right) = \sum_A P\left(a_0 | a_1 \cdots a_k A\right)$$

Chains with complete connections

- **Simulation scheme by the sequence of canonical Markov approximations of finite order k ($k-$CMA)**
  $k-$CMA of a process X is a Markov chain $Y^{(k)}$ of order k with conditional probabilities $P^{(k)}$
  $$P^{(k)}\left(a_0|a_1\cdots a_k\right) = P\left(a_0|a_1\cdots a_k\right) = \sum_A P\left(a_0|a_1\cdots a_k A\right)$$
- For a CCC X with summable decays
  $$\overline{d}\left(X, Y^{(k)}\right) \leq C\gamma_k \tag{2}$$
  The property of the Markov approximation, essential for the approximation result (2), is
  $$\inf_A P\left(a_0|a_1\cdots a_k A\right) \leq P^{(k)}\left(a_0|a_1\cdots a_k\right) \leq \sup_A P\left(a_0|a_1\cdots a_k A\right) \tag{3}$$
  meaning that for Markov approximation schemes, other than the canonical one, Eq.(2) holds provided (3) is satisfied

Looking for a Gibbs measure

- **For the market fluctuation data**:
  $\leq k-$Markov approximation:
  i) Empirical transition probabilities $\widetilde{P}\left(a_0 | a_1 \cdots a_m\right)$ inferred from the probability of blocks of order $m+1$. up to m_{Max}
  ii) $\leq k-$Simulation: look at the current block $\left(a_1 \cdots a_k\right)$ and use the $k-$empirical probability to infer the next state a_0. If that block has not appeared in the trainig data, use the $k-1$ sized block $a_2 \cdots a_k$ and the $k-1$ order empirical probabilities

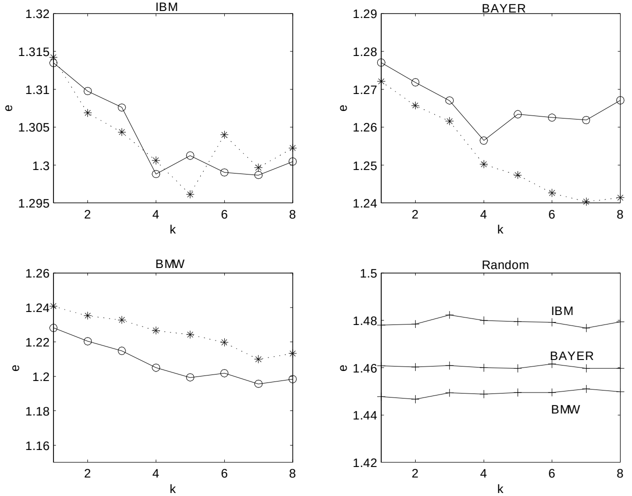
Looking for a Gibbs measure

- **For the market fluctuation data**:
  $\leq k-$Markov approximation:
  i) Empirical transition probabilities $\widetilde{P}\left(a_0|a_1\cdots a_m\right)$ inferred from the probability of blocks of order $m+1$. up to m_{Max}
  ii) $\leq k-$Simulation: look at the current block $\left(a_1\cdots a_k\right)$ and use the $k-$empirical probability to infer the next state a_0. If that block has not appeared in the trainig data, use the $k-1$ sized block $a_2\cdots a_k$ and the $k-1$ order empirical probabilities
- Averaged squared error

$$e^2 = \left\langle \left(\widetilde{a}_0 - a_0\right)^2 \right\rangle$$

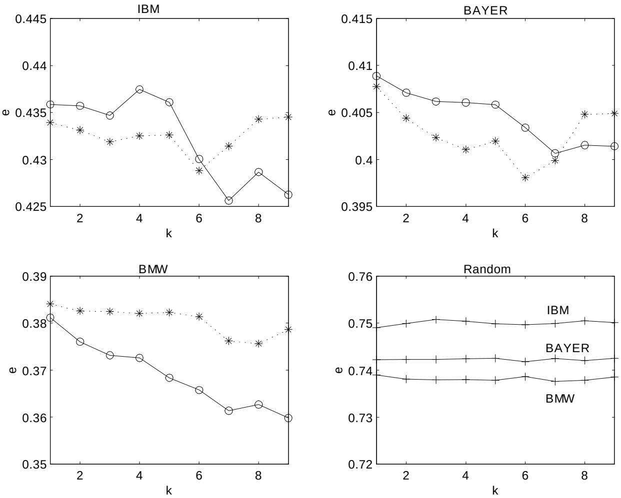
Chains with complete connections

The past predicting the future (o) and the future predicting the past (*), compared to random choice (5 symbols alphabet)



Chains with complete connections

The past predicting the future (o) and the future predicting the past (*), compared to random choice (3 symbols alphabet)



Chains with complete connections

Main conclusions:

- Average prediction better than random choice.
- Main improvement results from correct accounting of the two-symbol probabilities ($k = 1$)
- Small (but consistent) improvement using past information up $k = 4$ or 5. No significant improvement using higher order approximations
- Bulk of data represented by a short-memory process. Nevertheless there is evidence for a small long-memory component that is captured by the higher-order Markov approximations
- There is a maximum $k = k_m$ that should be used for the simulation process
- Inter-companies prediction: improvement coming from the one-symbol probabilities (as compared to random choice) is obtained. For the long-memory component the behavior is company-dependent.

General conclusions

1. Bulk of the market fluctuation process is a short-memory process. In addition it has a small long-memory component associated with the large fluctuations of the returns.
2. Existence of the long-memory component suggests the *chains with complete connections and summable decays* as a framework
3. Although the decays may be exponentially converging, the lack of accurate data for long blocks prevents an accurate description by a finite range Gibbs potential.
4. The sequence of empirical based $\leq k-$Markov approximations seems the most unbiased simulation of the process. Eventual convergence in the $\overline{d}-$distance sense, because of summable decays.

References

P. Doukhan, G. Oppenheim and M. S. Taqqu; *Theory and applications of long-range dependence*, Birkhäuser, Berlin 2003.
J.-R. Chazottes, E. Floriani and R. Lima; *Relative entropy and identification of Gibbs measures*, J. of Stat. Phys. 90 (1998) 697-725.
M. Iosifescu and S. Grigorescu; *Dependence with Complete Connections and its Applications*, Cambridge University Press, Cambridge, UK, 1990.
G. Maillard; *Introduction to chains with complete connections*, http://www.latp.univ-mrs.fr/~maillard/Greg/Publications_files/main.pdf
X. Bressaud, R. Fernandez and A. Galves; *Speed of $\overline{d}-$convergence for Markov approximations of chains with complete connections. A coupling approach*, Stoch. Proc. and Appl. 83 (1999) 127-138.
F. Comets, R. Fernandez and P. A. Ferrari; *Processes with long memory: Regenerative construction and perfect simulation*, arXiv:math.PR/0009204.
R. Vilela Mendes, R. Lima and T. Araújo; *A process reconstruction analysis of market fluctuations*, Int. J. of Theoretical and Applied Finance 5 (2002) 797-821.

Fractional processes. Self-similarity. Fractional Brownian motion (fBm)

- A process $\{X(t), t \geq 0\}$ is *selfsimilar* if for any a there is b such that
$$\{X(at)\} \stackrel{d}{=} \{bX(t)\}$$
$b = a^H$, process $H-$selfsimilar (or $H-$ss) - (Hurst exponent)

Fractional processes. Self-similarity. Fractional Brownian motion (fBm)

- A process $\{X(t), t \geq 0\}$ is *selfsimilar* if for any a there is b such that
$$\{X(at)\} \stackrel{d}{=} \{bX(t)\}$$
$b = a^H$, process $H-$selfsimilar (or $H-$ss) - (Hurst exponent)
- A process has *stationary increments (si)* if any distribution of
$$\{X(t+h) - X(t), t \geq 0\}$$
is independent of $t \geq 0$

Fractional processes. Self-similarity. Fractional Brownian motion (fBm)

- A process $\{X(t), t \geq 0\}$ is *selfsimilar* if for any a there is b such that

$$\{X(at)\} \stackrel{d}{=} \{bX(t)\}$$

  $b = a^H$, process $H-$selfsimilar (or $H-$ss) - (Hurst exponent)
- A process has *stationary increments (si)* if any distribution of

$$\{X(t+h) - X(t), t \geq 0\}$$

  is independent of $t \geq 0$
- *Theorem: If* $\{X(t), t \geq 0\}$ *is real-valued, H-ss with stationary increments and* $\mathbb{E}\left[X(1)^2\right] < \infty$*, then*

$$\mathbb{E}[X(t)X(s)] = \frac{1}{2}\left\{t^{2H} + s^{2H} - |t-s|^{2H}\right\}\mathbb{E}\left[X(1)^2\right]$$

  The simplest such process is a Gaussian process called fractional Brownian motion (fBm), $B_H(t)$, defined to have $\mathbb{E}[B_H(t)] = 0$. fBm is the unique Gaussian H-ss process with stationary increments

Fractional Gaussian noise and long-range dependence

- The process
$$Y_t = B_H\left(t+1\right) - B_H\left(t\right)$$
is called *fractional Gaussian noise*

Fractional Gaussian noise and long-range dependence

- The process

$$Y_t = B_H(t+1) - B_H(t)$$

  is called *fractional Gaussian noise*
- Renormalization group

$$T_N : Y_t \to (T_N Y)_t = \frac{1}{N^H} \sum_{i=t}^{t+N-1} Y_i$$

Fractional Gaussian noise and long-range dependence

- The process
$$Y_t = B_H(t+1) - B_H(t)$$
is called *fractional Gaussian noise*
- Renormalization group
$$T_N : Y_t \to (T_N Y)_t = \frac{1}{N^H} \sum_{i=t}^{t+N-1} Y_i$$
- *Theorem: Within the class of stationary sequences, fractional Gaussian noise is the only Gaussian fixed point of the renormalization group*

Fractional Gaussian noise and long-range dependence

- The process

$$Y_t = B_H(t+1) - B_H(t)$$

  is called *fractional Gaussian noise*
- Renormalization group

$$T_N : Y_t \to (T_N Y)_t = \frac{1}{N^H} \sum_{i=t}^{t+N-1} Y_i$$

- *Theorem: Within the class of stationary sequences, fractional Gaussian noise is the only Gaussian fixed point of the renormalization group*
- *Long-range dependence:* Let $\{X(t), t \geq 0\}$ be H-ss, si, 0<H<1 with $E\left[X(1)^2\right] < \infty$ and define

$$\xi(n) = X(n+1) - X(n)$$

$$r(n) = \mathbb{E}\left[\xi(0)\,\xi(n)\right] = \frac{1}{2}\left\{(n+1)^{2H} - 2n^{2H} + (n-1)^{2H}\right\}\mathbb{E}\left[X(1)^2\right]$$

Long-range dependence

- Then

$$
\begin{aligned}
r(n) \quad &\underset{n\to\infty}{\sim} \quad H(2H-1)\, n^{2H-2}\mathbb{E}\left[X(1)^2\right] \quad , \quad H \neq \tfrac{1}{2} \\
r(n) \quad &= \quad 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad H = \tfrac{1}{2}
\end{aligned}
$$

and

$$
\begin{array}{ccccc}
0 < H < \frac{1}{2} & , & \sum_{n=0}^{\infty} |r(n)| < \infty & & \\
H = \frac{1}{2} & , & \text{uncorrelated} & & \\
\frac{1}{2} < H < 1 & , & \sum_{n=0}^{\infty} |r(n)| = \infty & , & \text{long-range dependence}
\end{array}
$$

Long-range dependence

- Then

$$
\begin{aligned}
r(n) &\underset{n\to\infty}{\sim} H(2H-1)\, n^{2H-2}\mathbb{E}\left[X(1)^2\right] \quad , \quad H \neq \tfrac{1}{2} \\
r(n) &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad H = \tfrac{1}{2}
\end{aligned}
$$

  and

$$
\begin{array}{ccc}
0 < H < \frac{1}{2} & , & \sum_{n=0}^{\infty} |r(n)| < \infty \\
H = \frac{1}{2} & , & \text{uncorrelated} \\
\frac{1}{2} < H < 1 & , & \sum_{n=0}^{\infty} |r(n)| = \infty \quad , \quad \text{long-range dependence}
\end{array}
$$

- If $0 < H < \frac{1}{2}$, $r(n) < 0$ for $n \geq 1$ (negative correlation, anti-persistent process),

Long-range dependence

- Then

$$r(n) \underset{n\to\infty}{\sim} H(2H-1)\,n^{2H-2}\mathbb{E}\left[X(1)^2\right] \quad , \quad H \neq \tfrac{1}{2}$$

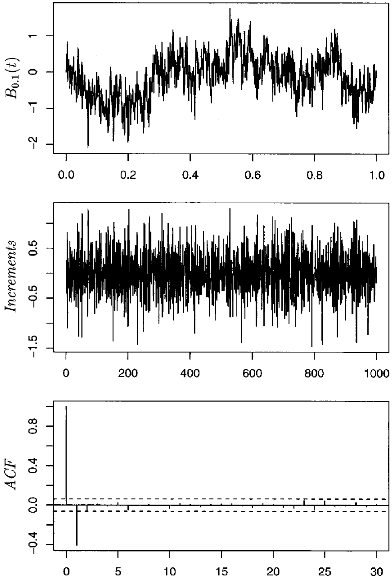
$$r(n) = 0 \qquad\qquad H = \tfrac{1}{2}$$

and

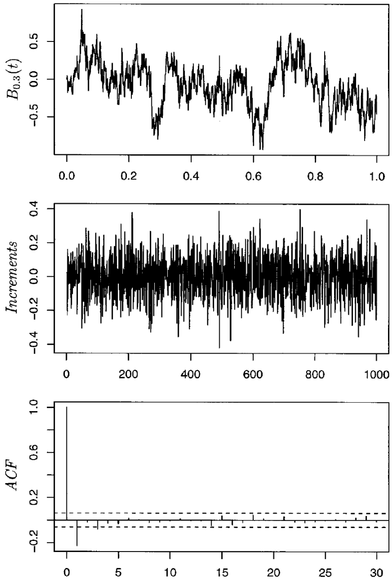
$$\begin{array}{ccccc} 0 < H < \frac{1}{2} & , & \sum_{n=0}^{\infty} |r(n)| < \infty & & \\ H = \frac{1}{2} & , & \text{uncorrelated} & & \\ \frac{1}{2} < H < 1 & , & \sum_{n=0}^{\infty} |r(n)| = \infty & , & \text{long-range dependence} \end{array}$$

- If $0 < H < \frac{1}{2}$, $r(n) < 0$ for $n \geq 1$ (negative correlation, anti-persistent process),
- If $\frac{1}{2} < H < 1$, $r(n) > 0$ for $n \geq 1$ (positive correlation, persistent process).

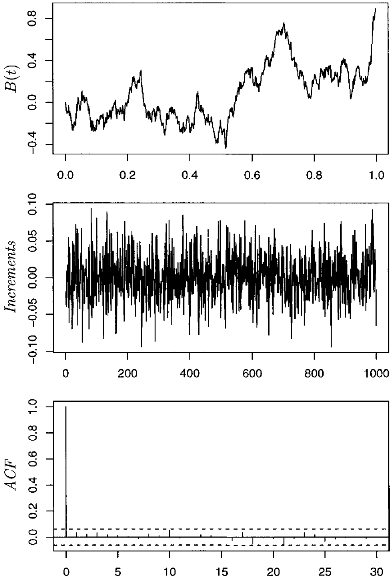
Plots of fractional Brownian motion (H=0.1)



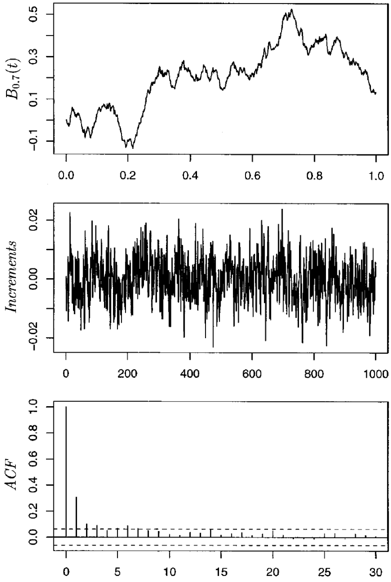
Plots of fractional Brownian motion (H=0.3)



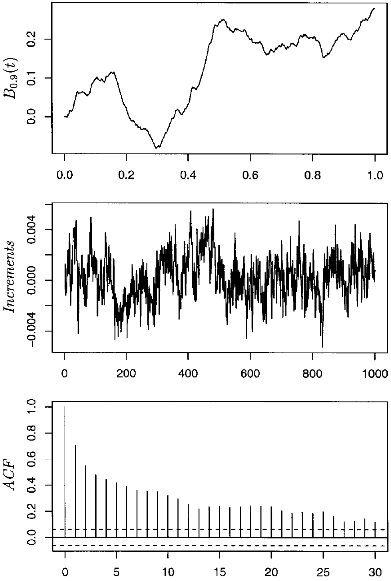
Plots of fractional Brownian motion (H=0.5)



Plots of fractional Brownian motion (H=0.7)



Plots of fractional Brownian motion (H=0.9)



Sample path properties

- Kolmogorov criterium for the existence of a continuous version of $X(t)$: $\exists\alpha \geq 1, \beta > 0, k > 0$ such that

$$\mathbb{E}\left[\left|X(t) - X(s)\right|^{\alpha}\right] \leq k\left|t - s\right|^{1+\beta}$$

Sample path properties

- Kolmogorov criterium for the existence of a continuous version of $X(t)$: $\exists \alpha \geq 1, \beta > 0, k > 0$ such that

$$\mathbb{E}\left[\left|X(t) - X(s)\right|^{\alpha}\right] \leq k\left|t - s\right|^{1+\beta}$$

- Hölder continuity of order γ

$$P\left\{\omega \in \Omega, \sup \frac{\left|X(t,\omega) - X(s,\omega)\right|}{\left|t - s\right|^{\gamma}} \leq \delta\right\} = 1$$

Sample path properties

- Kolmogorov criterium for the existence of a continuous version of $X(t)$: $\exists \alpha \geq 1, \beta > 0, k > 0$ such that

$$\mathbb{E}\left[|X(t) - X(s)|^{\alpha}\right] \leq k\,|t-s|^{1+\beta}$$

- Hölder continuity of order γ

$$P\left\{\omega \in \Omega, \sup \frac{|X(t,\omega) - X(s,\omega)|}{|t-s|^{\gamma}} \leq \delta\right\} = 1$$

- From

$$\mathbb{E}\left[|B_H(t) - B_H(s)|^{\alpha}\right] = \mathbb{E}\left[|B_H(1)|^{\alpha}\right] |t-s|^{\alpha H}$$

  it follows that fBm $\{B_H(t)\}$ has a continuous version $(P\{X(t) = B_H(t)\} = 1)$ the sample paths of which are Hölder continuous of order $\beta \in [0, H)$ and are almost surely nowhere locally Hölder continuous of order $\gamma > H$.

Sample path properties

- Kolmogorov criterium for the existence of a continuous version of $X(t)$: $\exists \alpha \geq 1, \beta > 0, k > 0$ such that
$$\mathbb{E}\left[|X(t) - X(s)|^{\alpha}\right] \leq k\,|t-s|^{1+\beta}$$
- Hölder continuity of order γ
$$P\left\{\omega \in \Omega, \sup \frac{|X(t,\omega) - X(s,\omega)|}{|t-s|^{\gamma}} \leq \delta\right\} = 1$$
- From
$$\mathbb{E}\left[|B_H(t) - B_H(s)|^{\alpha}\right] = \mathbb{E}\left[|B_H(1)|^{\alpha}\right]|t-s|^{\alpha H}$$
it follows that fBm $\{B_H(t)\}$ has a continuous version $(P\{X(t) = B_H(t)\} = 1)$ the sample paths of which are Hölder continuous of order $\beta \in [0, H)$ and are almost surely nowhere locally Hölder continuous of order $\gamma > H$.
- Sample paths of fBm have nowhere bounded variation and are not differentiable.

FBm for H not 1/2 is not a semimartingale

- *p - Variation of a process* $X(t)$

$$V_p(0, T) = \sup_{partitions} \sum_{k=1}^{n} \left| X(t_k) - X(t_{k-1}) \right|^p$$

FBm for H not 1/2 is not a semimartingale

- *p - Variation of a process* $X(t)$

$$V_p(0,T) = \sup_{partitions} \sum_{k=1}^{n} \left| X(t_k) - X(t_{k-1}) \right|^p$$

- Index of the p-variation

$$I(X, |0, T]) = \inf \{ p > 0;\ V_p(0,T) < \infty \}$$

FBm for H not 1/2 is not a semimartingale

- *p - Variation of a process* $X(t)$

$$V_p(0,T) = \sup_{partitions} \sum_{k=1}^{n} |X(t_k) - X(t_{k-1})|^p$$

- Index of the p-variation

$$I(X, |0,T]) = \inf\{p > 0; V_p(0,T) < \infty\}$$

- From $\mathbb{E}\left[|B_H(t) - B_H(s)|^{\alpha}\right] = \mathbb{E}\left[|B_H(1)|^{\alpha}\right] |t-s|^{\alpha H}$ it is easy to conclude that

$$I(B_H, |0,T]) = \frac{1}{H}$$

The index of a semimartingale must belong to $[0,1] \cup \{2\}$. Therefore $B_H(t)$ cannot be a semimartingale unless $H = \frac{1}{2}$

Representation of fractional Brownian motion by Wiener integrals

- "Time" representation

$$B_H(t) \overset{d}{=} \frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\left\{\int_{-\infty}^{0}\left((t-u)^{H-\frac{1}{2}}-(-u)^{H-\frac{1}{2}}\right)dB(u)+\int_{0}^{t}(t-u)^{H-\frac{1}{2}}\,dB(u)\right\}$$

Representation of fractional Brownian motion by Wiener integrals

- "Time" representation

$B_H(t) \overset{d}{=}$
$\frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\left\{\int_{-\infty}^{0}\left((t-u)^{H-\frac{1}{2}}-(-u)^{H-\frac{1}{2}}\right)dB(u)+\int_0^t (t-u)^{H-\frac{1}{2}}\,dB(u)\right\}$

- Finite interval representation

$$B_H(t) \overset{d}{=} C\int_0^t K(t,u)\,dB(u)$$

$K(t,u) =$
$\left\{\left(\frac{t}{u}\right)^{H-\frac{1}{2}}(t-u)^{H-\frac{1}{2}}-\left(H-\frac{1}{2}\right)u^{\frac{1}{2}-H}\int_u^t x^{H-\frac{3}{2}}(x-u)^{H-\frac{1}{2}}\,dx\right\}$

Representation of fractional Brownian motion by Wiener integrals

- "Time" representation

$$B_H(t) \overset{d}{=} \frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\left\{\int_{-\infty}^{0}\left((t-u)^{H-\frac{1}{2}}-(-u)^{H-\frac{1}{2}}\right)dB(u)+\int_0^t (t-u)^{H-\frac{1}{2}}\,dB(u)\right\}$$

- Finite interval representation

$$B_H(t) \overset{d}{=} C\int_0^t K(t,u)\,dB(u)$$

$$K(t,u)=\left\{\left(\frac{t}{u}\right)^{H-\frac{1}{2}}(t-u)^{H-\frac{1}{2}}-\left(H-\frac{1}{2}\right)u^{\frac{1}{2}-H}\int_u^t x^{H-\frac{3}{2}}(x-u)^{H-\frac{1}{2}}\,dx\right\}$$

- "Spectral" representation

$$B_H(t) \overset{d}{=} \frac{\Gamma\left(H+\frac{1}{2}\right)}{(2\pi)^{\frac{1}{2}}\,c(H)}\int_{\mathbb{R}}\frac{e^{ixt}-1}{ix}\,d\tilde{B}(x)$$

$$\tilde{B}(x)=B_1+iB_2,\ B_1(x)=B_1(-x),\ B_2(x)=-B_2(-x)$$

Representation of fractional Brownian motion

- (Paley-Wigner-type) series representation

$$B_H(t) = \sum_{n \in \mathbb{Z}} \frac{e^{2i\omega_n t/T} - 1}{2i\omega_n t/T} Z_n$$

convergent in $t \in [0, T]$.
The $\omega_n' s$ are the real zeros of the Bessel function J_{1-H} and the Z_n are independent complex-valued Gaussian random variables with mean zero and variance

$$\mathbb{E}|Z_n|^2 = \begin{cases} (2-2H)^{-1}\Gamma^{-2}(1-H)\left(\frac{\omega_n}{2}\right)^{-2H} J_{-H}^{-2}(\omega_n) V_T^{-1} & \omega_n \neq 0 \\ V_T^{-1} & \omega_n = 0 \end{cases}$$

$$V_T = \frac{\Gamma(3/2-H)}{2H\Gamma(H+1/2)\Gamma(3-2H)} T^{2-2H}$$

Stochastic integration for fractional Brownian motion

- The non semimartingale property for $H \neq \frac{1}{2}$ makes the standard construction not applicable.

Stochastic integration for fractional Brownian motion

- The non semimartingale property for $H \neq \frac{1}{2}$ makes the standard construction not applicable.
- If $1/2 < H < 1$ the zero quadratic variation allows for a pathwise construction of the integral

$$\int_a^b f(t,\omega)\,\delta B_H(t) := \lim_{\Delta t_k \to 0} \sum_k f(t_k,\omega)\left(B_H(t_{k+1}) - B_H(t_k)\right)$$

However with the pathwise definition one ends up, in general with

$$\mathbb{E}\left[\int f(t, B_H(t))\,\delta B_H(t)\right] \neq 0$$

leading, for example, to arbitrage in financial applications.

Stochastic integration for fractional Brownian motion

- The non semimartingale property for $H \neq \frac{1}{2}$ makes the standard construction not applicable.
- If $1/2 < H < 1$ the zero quadratic variation allows for a pathwise construction of the integral

$$\int_a^b f(t,\omega)\,\delta B_H(t) := \lim_{\Delta t_k \to 0} \sum_k f(t_k,\omega)\left(B_H(t_{k+1}) - B_H(t_k)\right)$$

However with the pathwise definition one ends up, in general with

$$\mathbb{E}\left[\int f(t, B_H(t))\,\delta B_H(t)\right] \neq 0$$

leading, for example, to arbitrage in financial applications.

- An alternative

$$\int_a^b f(t,\omega)\,dB_H(t) := \lim_{\Delta t_k \to 0} \sum_k f(t_k,\omega) \diamond \left(B_H(t_{k+1}) - B_H(t_k)\right)$$

$\diamond$ is the Wick product. This integral reduces to Ito's for $H = \frac{1}{2}$.

Fractional Ito formulas

- $H \in (0,1)$

$$f(t, B_H(t)) = f(0,0) + \int_0^t \frac{\partial f}{\partial s}(s, B_H)\, ds + \int_0^t \frac{\partial f}{\partial x}(s, B_H)\, dB_H(s) + H \int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_H(s))\, s^{2H-1} ds$$

Fractional Ito formulas

- $H \in (0,1)$

$$
\begin{aligned}
f(t, B_H(t)) &= f(0,0) + \int_0^t \frac{\partial f}{\partial s}(s, B_H)\, ds + \int_0^t \frac{\partial f}{\partial x}(s, B_H)\, dB_H(s) \\
&\quad + H \int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_H(s))\, s^{2H-1} ds
\end{aligned}
$$

- $H \in \left(\frac{1}{2}, 1\right)$ $\qquad dX_i(t) = \sum_{j=1}^m \sigma_{ij}(t, \omega)\, dB_H^j(t)$

$$
\begin{aligned}
f(t, X(t)) &= f(0, X(0)) + \int_0^t \frac{\partial f}{\partial s}(s, X(s))\, ds \\
&\quad + \int_0^t \sum_{i=1}^n \frac{\partial f}{\partial x^i}(s, X(s))\, dX_i(s) \\
&\quad + \int_0^t \left\{ \sum_{i,j=1}^m \frac{\partial^2 f}{\partial x^i \partial x^j}(s, X(s)) \sum_{k=1}^m \sigma_{ik}(s)\, D_{\phi,s}^k (X_j(s)) \right\}
\end{aligned}
$$

$D_{\phi,t} F = \int_{\mathbb{R}} \phi(s,t)\, D_s F ds$ and $\phi(s,t) = H(2H-1)\,|s-t|^{2H-2}$

References

P. Embrechts and M. Maejima; *Selfsimilar processes*, Princeton Univ. Press, 2002.
F. Biagini, B. Oksendal, A. Sulem, and N. Wallner; *An introduction to white noise theory and Malliavin Calculus for fractional Brownian motion*, Proc. Royal Soc. 460 (2004) 347-372.
F. Biagini, Y. Hu, B. Oksendal and T. Zhang; *Stochastic calculus for fractional Brownian motion and applications*, Springer 2007.
K. Dzhaparidze and H. van Zanten; *Krein's spectral theory and the Paley-Wiener expansion for fractional Brownian motion*, Annals of Prob. 33 (2005) 620-644.

Modeling application: The fractional volatility model

- Classical Mathematical Finance has, for a long time, been based on the assumption that the price process of market securities may be approximated by geometric Brownian motion (GBM)

$$dS_t = \mu S_t dt + \sigma S_t dB(t) \tag{4}$$

consistent with the fact that in liquid markets the autocorrelation of price changes decays to negligible values in a few minutes.
- Otherwise **GBM has serious shortcomings:**
  - Does not reproduce the empirical leptokurtosis
  - Does not explain why nonlinear functions of the returns exhibit significant positive autocorrelation (volatility clustering)
  - There is an essential memory component and a dynamical model for volatility is needed, σ in Eq.(4) being itself a process.
- This led to many deterministic and stochastic models for the volatility which fit the leptokurtosis but not always the long memory. In contrast with GBM, they mostly lack the kind of nice mathematical properties needed to develop the tools of mathematical finance.

The fractional volatility model

- **The fractional volatility model:** a model based on simple mathematical assumptions and reconstructed from the data.
- **Basic hypothesis:**
  (H1) The log-price process $\log S_t$ belongs to a product space $(\Omega_1 \times \Omega_2, P_1 \times P_2)$ of which (Ω_1, P_1) is the Wiener space and the second, (Ω_2, P_2), is a probability space to be reconstructed from the data. With $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ and $\mathcal{F}_{1,t}$ and $\mathcal{F}_{2,t}$ the σ-algebras in Ω_1 and Ω_2
  $$\log S_t\left(\omega_1, \omega_2\right)$$
  (H2) For each fixed ω_2, $\log S_t\left(\cdot, \omega_2\right)$ is a square integrable random variable in Ω_1.
- These principles and a careful analysis of the market data led, in an essentially unique way, to the following model:

$$\boxed{\begin{array}{lcl} dS_t & = & \mu S_t\, dt + \sigma_t S_t\, dB\left(t\right) \\ \log \sigma_t & = & \beta + \frac{k}{\delta}\left\{B_H\left(t\right) - B_H\left(t-\delta\right)\right\} \end{array}} \tag{5}$$

The fractional volatility model

- M1 – ***Consequence of H2*** (theor. 1.1.3 in Nualart)

$$\frac{dS_t(\bullet,\omega')}{S_t} = \mu(\bullet,\omega')dt + \sigma(\bullet,\omega')dB_t$$

with μ(,ω') and σ(,ω') processes in Ω

- σ(ω,ω') is called the *induced volatility*

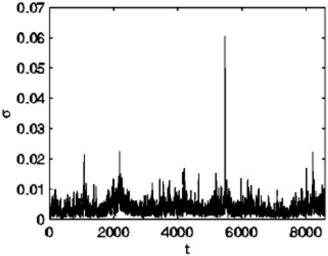
- E2 – ***To reconstruct σ(ω,ω') from the data***

From M1

$$\sigma_t^2(\bullet,\omega') = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left\{E\left(\log S_{t+\varepsilon} - \log S_t\right)^2\right\}$$

Approximated from the data by

$\sigma_t^2(\cdot,\omega') \approx \mathrm{var}(\log S_t)/\Delta t$



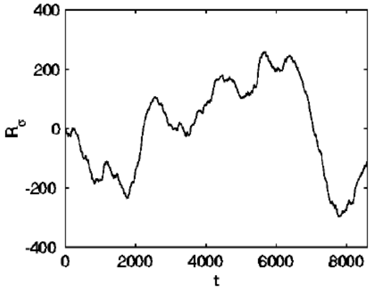
What does the data suggest for σ_t ?

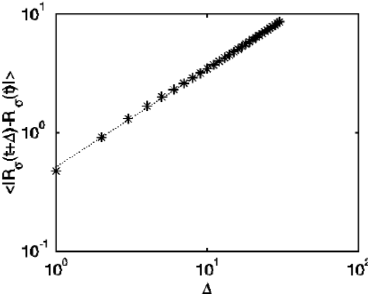
- σ_t **is not self similar**

$$E\left|\frac{\sigma(t+\Delta)-\sigma(t)}{\sigma(t)}\right| \neq \Delta^H$$

- **However $R_\sigma(t)$ is**
  $\Sigma \log \sigma(n\delta) = \beta t + R_\sigma(t)$
  $H \approx 0.8 - 0.9$

$$E\left|\frac{R_\sigma(t+\Delta)-R_\sigma(t)}{R_\sigma(t)}\right| = \Delta^H$$





What does the data suggest for σ_t ?

- Recall:
  If a process X_t has finite variance, stationary increments and is self-similar, then
  $Cov(X_s, X_t)=(|s|^{2H}+|t|^{2H}-|s-t|^{2H})E(X_1^2)$
  and the simplest such process is a zero-mean Gaussian process, Fractional Brownian motion B^H_t with long-range dependence for H>1/2
- H3 – ***Mathematical simplicity***
- E2+H3 – ***Consistency with the data and the choice of the simplest mathematical process implies*** :

$$\log \sigma_t = \beta + (k/\delta)\,(B^H_t - B^H_{t-\delta})$$

σ_t modeled by a stochastic exponential of fractional noise

The fractional volatility model

$$\boxed{\begin{array}{rcl} dS_t & = & \mu S_t \, dt + \sigma_t S_t \, dB(t) \\ \log \sigma_t & = & \beta + \frac{k}{\delta}\left\{B_H(t) - B_H(t-\delta)\right\} \end{array}}$$

- The data suggests values of H in the range $0.8 - 0.9$.
- The second equation leads to

$$\sigma(t) = \theta e^{\frac{k}{\delta}\{B_H(t) - B_H(t-\delta)\} - \frac{1}{2}\left(\frac{k}{\delta}\right)^2 \delta^{2H}}$$

  with $E[\sigma(t)] = \theta > 0$.
- Describes well **the statistics of price returns** for a large $\delta-$range in different markets and also implies **a new option pricing formula**, with "smile" deviations from Black-Scholes.

Some consequences

- **From**

$$\ln \sigma_t = \beta + \frac{k}{\delta}\left(B_H(t) - B_H(t-\delta)\right)$$

$$\sigma_t = \theta \exp\left[\frac{k}{\delta}\left(B(t) - B(t-\delta)\right) - \frac{1}{2}\left(\frac{k}{\delta}\right)^2 \delta^{2H}\right]$$

- **Log σ_t is a Gaussian process with mean β and covariance**

$$\psi(s,u) = \frac{k^2}{2\delta^2}\left\{|s-u+\delta|^{2H} + |u-s+\delta|^{2H} - 2|s-u|^{2H}\right\}$$

- **Then**

$$p_\delta(\sigma) = \frac{1}{\sqrt{2\pi}\sigma k \delta^{H-1}} \exp\left\{-\frac{(\log \sigma - \beta)^2}{2k^2\delta^{2H-2}}\right\}$$

Time scales and pdf's

- **and for the returns**

$$P_\delta(\log \frac{S_T}{S_t}) = \int_0^\infty d\sigma\, p_\delta(\sigma) p_\sigma(\log \frac{S_T}{S_t})$$

with

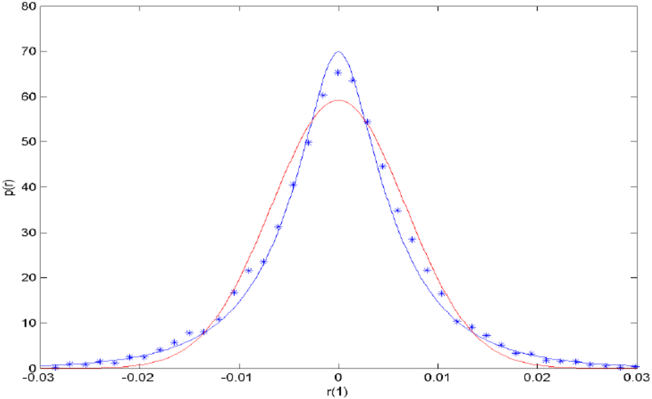
$$p_\sigma(\log \frac{S_T}{S_t}) = \frac{1}{\sqrt{2\pi\sigma^2 (T-t)}} \exp\left\{ \frac{\left(\log \frac{S_T}{S_t} - \left(\mu - \frac{\sigma^2}{2} \right)(T-t) \right)^2}{2\sigma^2 (T-t)} \right\}$$

$\Delta = T - t$, $r_0 = (\mu - \sigma^2/2)(T - t)$

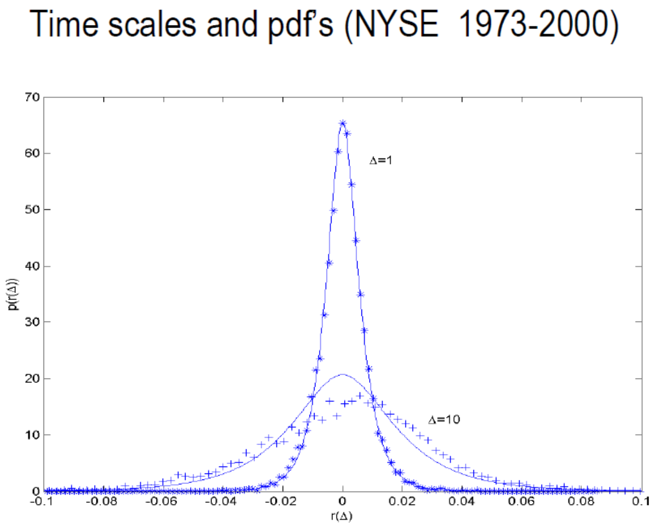
- **The probability distribution of the returns depends on the observation time scale δ**

Time scales and pdf's (NYSE 1973-2000)

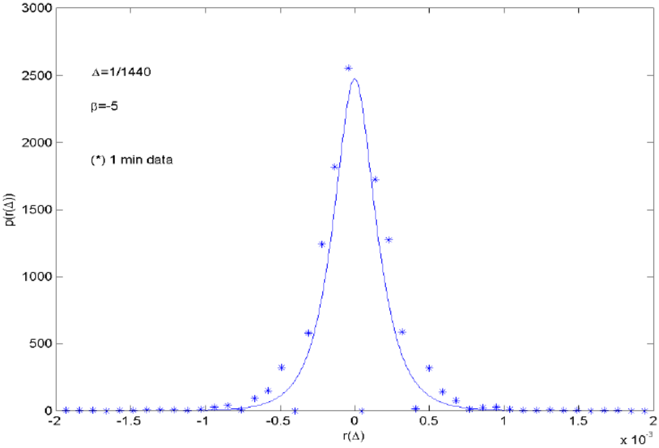
- **H=0.83 k=0.59 β= - 5 δ=1 Δ=1**



The fractional volatility model



Time scales and pdf's (USD-Euro 05-06 2001)



- The following nonlinear correlation of the returns

$$L(\tau) = \left\langle \left| r(t+\tau) \right|^2 r(t) \right\rangle - \left\langle \left| r(t+\tau) \right|^2 \right\rangle \langle r(t) \rangle$$

  is called *leverage* and the *leverage effect* is the fact that, for $\tau > 0$, $L(\tau)$ starts from a negative value whose modulus decays to zero whereas for $\tau < 0$ it has almost negligible values.

- In the form of Eqs.(5) the volatility process σ_t affects the log-price, but is not affected by it. Therefore, in its simplest form the fractional volatility model contains no leverage effect.

Leverage and the identification of the stochastic generators

Leverage may, however, be implemented in the model in a simple way by using an integral representation of the fractional Brownian motion

$$B_H(t) = \int_0^t K_H(t,s)\, dW_s$$

W_t being a Brownian motion and

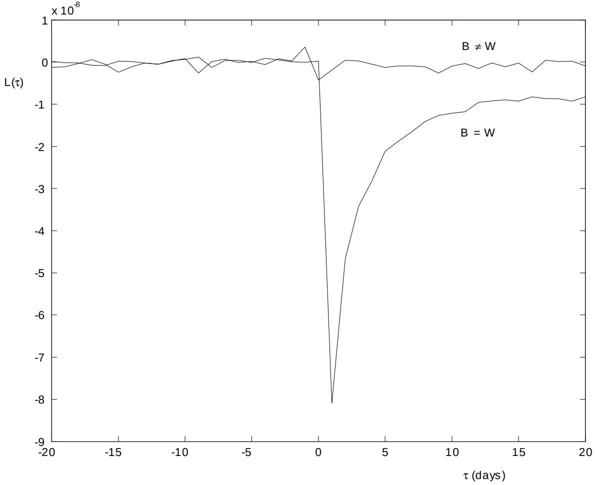
$$K_H(t,s) = C_H s^{\frac{1}{2}-H} \int_s^t (u-s)^{H-\frac{3}{2}} u^{H-\frac{1}{2}}\, du, \quad s < t$$

if one identifies the Brownian processes B_t in the equation for dS_t with W_t, that is, identifying the random generator of the log-price process with the stochastic integrator of the volatility, at least a part of the leverage effect is taken into account.

A new fractional volatility model is defined

$$\begin{aligned} dS_t &= \mu_t S_t dt + \sigma_t S_t dW_t \\ \log \sigma_t &= \beta + k' \int_{-\infty}^t (t-u)^{H-\frac{3}{2}}\, dW_u \end{aligned}$$

Leverage and the identification of the stochastic generators



Remarks and conclusions

- FVM describes well returns distribution and modifications to Black-Scholes.
- Universality through different markets. Related to limit-order book dynamics.
- Mathematical consistency. No-arbitrage.
- Leverage with identification of the generators of the stochastic processes. Completeness when the stochastic generators are identified.

References

D. Nualart; *The Malliavin calculus and related topics*, Springer 2006.
R. Vilela Mendes and M. J. Oliveira; *A data-reconstructed fractional volatility model*, Economics e-journal, discussion paper 2008-22.
R. Vilela Mendes; *A fractional calculus interpretation of the fractional volatility model*, Nonlinear Dyn. 55 (2009) 395–399.
R. Vilela Mendes; *The fractional volatility model: An agent-based interpretation,* Physica A: Stat. Mech. and its Applications, 387 (2008) 3987–3994.
R. Vilela Mendes, M. J. Oliveira and A. M. Meireles; *The fractional volatility model: No-arbitrage, leverage and completeness*, in publication

epsilon-machines

A process $\Pr(\overleftarrow{X}, \overrightarrow{X})$ is a *channel* with input distribution $\Pr(\overleftarrow{X})$ that transmits information from the *past* $\overleftarrow{X} = \ldots X_{-3}X_{-2}X_{-1}$ to the *future* $\overrightarrow{X} = X_0X_1X_2\ldots$

One wants to predict the future using information from the past. The prediction is probabilistic, specified by a distribution of possible futures $\overrightarrow{X}$ given a particular past $\overleftarrow{x}$: $\Pr(\overrightarrow{X}|\overleftarrow{x})$.

Implies that a good predictor needs to capture *all* of the information I shared between past and future: $\mathbf{E} = I[\overleftarrow{X}; \overrightarrow{X}]$, called the process's *excess entropy*.

Modeling:

Introduce an equivalence relation $\overleftarrow{x} \sim \overleftarrow{x}'$ that groups all histories giving rise to the same prediction:

$$\epsilon(\overleftarrow{x}) = \{\overleftarrow{x}' : \Pr(\overrightarrow{X}|\overleftarrow{x}) = \Pr(\overrightarrow{X}|\overleftarrow{x}')\}$$

Causal states:

$$\boldsymbol{\mathcal{S}} = \Pr(\overleftarrow{X}, \overrightarrow{X})/\sim$$

partitions the space of pasts into sets that are predictively equivalent
State-to-state transitions are denoted by matrices

$$T^{(x)}_{\mathcal{S}\mathcal{S}'} = \Pr(X = x, \mathcal{S}'|\mathcal{S})$$

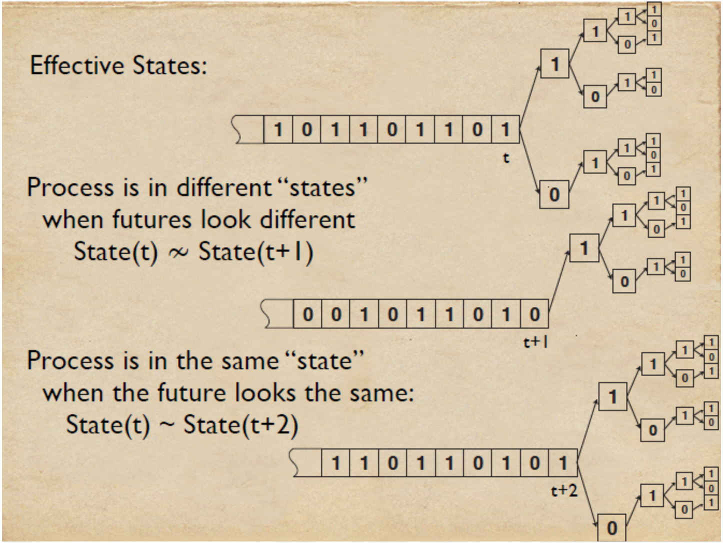
gives the probability of transitioning from one state $\mathcal{S}$ to the next $\mathcal{S}'$ on seeing measurement x.
The resulting model, consisting of the causal states and transitions, is called the process's $\epsilon-$machine
Causal states "shield" the future from the past:

$$\Pr(\overleftarrow{X}, \overrightarrow{X}|\mathcal{S}) = \Pr(\overleftarrow{X}|\mathcal{S})\Pr(\overrightarrow{X}|\mathcal{S}) \ .$$

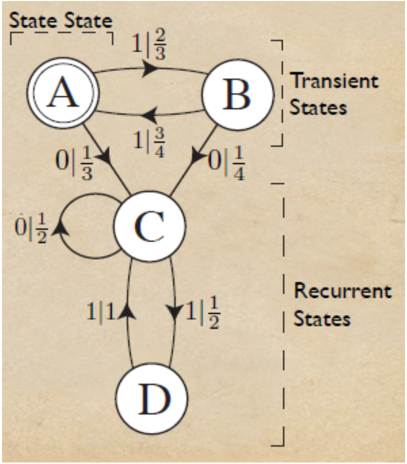
and are optimally predictive. Knowing which causal state a process is in, is just as good as having the entire past: $\Pr(\overrightarrow{X}|\mathcal{S}) = \Pr(\overrightarrow{X}|\overleftarrow{X})$. Causal states capture all of the information shared between past and future: $I[\mathcal{S};\overrightarrow{X}] = \mathbf{E}$.

epsilon-machines

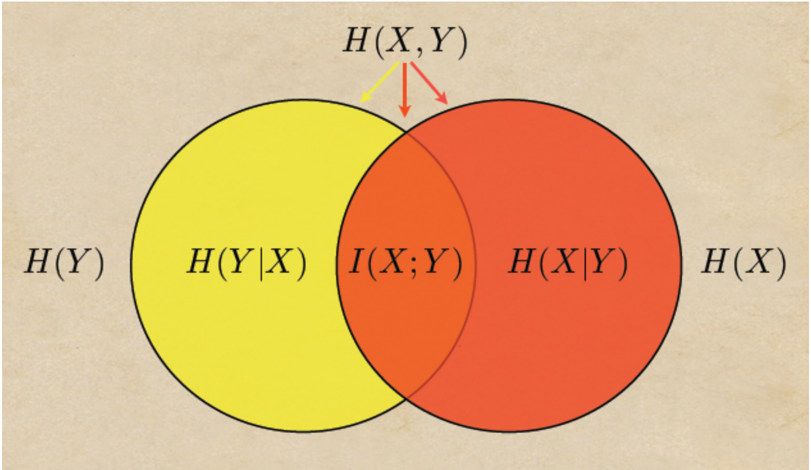
$\epsilon-$machines are *unifilar* : From the start state, each observed sequence $\ldots x_{-3}x_{-2}x_{-1}\ldots$ corresponds to one and only one sequence of causal states unlike general hidden Markov models.
For each state, each measurement symbol appears on at most one outgoing transition. Knowing the current state and measurement, the uncertainty in the next state vanishes: $H[\mathcal{S}_{t+1}|\mathcal{S}_t, X_t] = 0$. In summary, a process's $\epsilon-$machine is its unique minimal unifilar model.
There are algorithms to construct the $\epsilon-$machine from the data.


Effective States:
1 0 1 1 0 1 1 0 1
t
Process is in different "states"
when futures look different
State(t) ≁ State(t+1)
0 0 1 0 1 1 0 1 0
t+1
Process is in the same "state"
when the future looks the same:
State(t) ~ State(t+2)
1 1 0 1 1 0 1 0 1
t+2

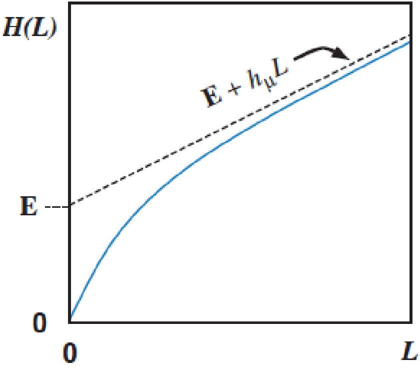
$\epsilon-$machine reconstructed for the even process

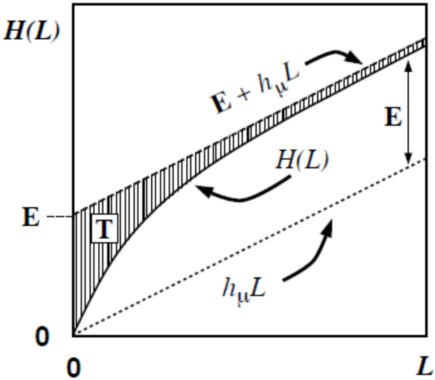


Entropies



Entropies





epsilon-machines: Markov order

A word w of length L is *synchronizing* if the Shannon entropy over the internal state, conditioned on w, is zero

$$Sync(w) \quad \Leftrightarrow \quad H[\mathcal{S}_\ell | X_{0:\ell} = w] = 0$$

Markov order (R)
Number of observations required to predict the future as accurately as if using the infinite past

$$R = \min\{\ell \mid \Pr(\overrightarrow{X} | \overleftarrow{X}) = \Pr(\overrightarrow{X} | X_{-\ell:0})\}$$

Equivalently, the Markov order is the length at which a process's *block-entropy curve* $H[X_{0:\ell}]$ reaches its asymptote.

$$R = \min\{\ell \mid H[X_{0:\ell}] = \mathbf{E} + \ell h_\mu\}$$

The length of data one must see before being able to predict upcoming symbols with the minimum error rate. R is considered ∞ when the condition does not hold for any finite ℓ.

epsilon-machines: Cripticity and synchronization orders

Cryptic Order (k_χ): The number of observations required to account for the asymptotic amount of state information not transmitted to the future, or the number of states that cannot be retrodicted given the infinite future

$$k_\chi = \min\{\ell \mid H[\mathcal{S}_\ell | \overrightarrow{X}] = 0\}$$

Equivalently is the length at which the *block-state entropy* $H[X_{0:\ell}, \mathcal{S}_\ell]$ reaches its asymptotic behavior.

$$k_\chi = \min\{\ell \mid H[X_{0:\ell}, \mathcal{S}_\ell] = \mathbf{E} + \ell h_\mu\}$$

k_χ is considered ∞ when the condition does not hold for any finite ℓ.

Synchronization Order (k_S): One is synchronized to a process after seeing word w if there is complete certainty in the state. The *synchronization order* k_S is the minimum length for which every word is a synchronizing word

$$k_S \equiv \min\{\ell \mid H[\mathcal{S}_\ell | X_{0:\ell}] = 0\}$$

k_S is considered ∞ when the condition does not hold for any finite ℓ.

Markov, cripticity and synchronization orders: Some properties

1- The synchronization order is

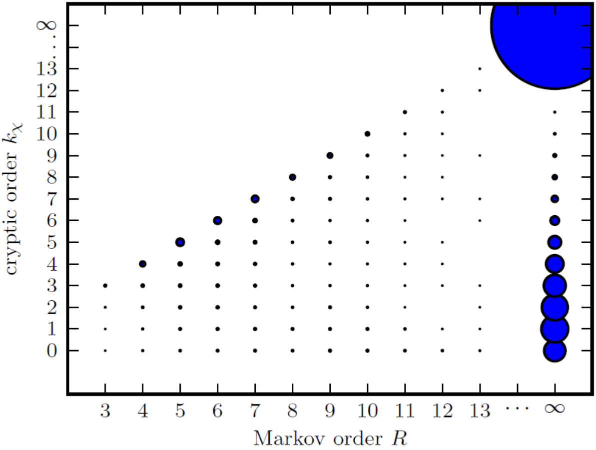
$$k_S = \max\{R, k_\chi\}$$

2- For $\epsilon-$machines

$$R = k_S$$

Genericity of infinite orders

Markov R and cryptic k_χ orders for all six-state, binary alphabet $\epsilon-$machines



References

C. R. Shalizi and J. P. Crutchfield; *Computational mechanics: Pattern and prediction, structure and simplicity*, J. Stat. Phys. 104 (2001) 817-879.
R. G. James, J. R. Mahoney, C. J. Ellison and J. P. Crutchfield; *Many roads to synchrony: Natural time scales and their algorithms*, arXiv:1010.5545.