CNR-IMATI Milano
December 14-16, 2004

WAVELETS AND SELF-SIMILARITY: THEORY AND APPLICATIONS

Lecture 4: Theory

P L A N

1. **SS + SSSI**
2. **STATIONARITY + LRD**
3. **SS + SSSI + Gaussian**
4. **ARFIMA**

General Results

■ We assume that all random processes discussed are real valued and defined on the same parameter space.

■ Two processes $X(t)$ and $Y(t)$, *equal in all finite dimensional distributions*, will be denoted as $X(t) \stackrel{d}{=} Y(t)$. This means that for any selection of "times" $0 \leq t_1 < t_2 < \ldots t_k < \infty$ random vectors $(X(\omega, t_1), \ldots, X(\omega, t_k))$ and $(Y(\omega, t_1), \ldots, Y(\omega, t_k))$ have the same distribution. Informally, processes equal-in-distribution are statistically indistinguishable.

■ Random process $X(t)$ is called stochastically continuous at t_0 if $\lim_{h \to 0} P(|X(t_0 + h) - X(t_0)| > \epsilon) = 0$, for any fixed $\epsilon > 0$.

■ Also, we consider processes not to be trivial. Process $X(t)$ is trivial if the distribution of random variable $X(\omega, t)$, t fixed is a point mass measure.

■ A random process $X(t), t > 0$ is called *self-similar* if for any $a > 0$, there exists $b > 0$ such that

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

■ (**Lamperti, 1962**) If random process $X(t), t \geq 0$ is nontrivial, stochastically continuous at 0, and self-similar, then there exists unique $H \geq 0$ such that $b = a^H$. If $X(0) = 0, a.s.$ then $H > 0$.

■ Standard definition of self-similar processes is as follows: Process $X(t),\ t \geq 0$ is self-similar, with self-similarity index H (H-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.

■ Uniqueness of H is not obvious from this definition, although, H is unique by the Lamperti's theorem. Also, from Definition it follows $X(0) = 0$.

■ Example: Standard Brownian Motion $B(t)$ is 1/2-ss. Indeed, the process $W(t) = 1/\sqrt{a}B(at)$ is standard Brownian motion, as well.

Stationarity and LRD

■ Let $\{X(t),\ t \in \mathbb{R}\}$ be a random process such that the autocovariance function

$$\gamma_X(r,s) = \mathcal{COV}(X(r), X(s)) = E(X(r) - EX(r))(X(s) - EX(s))$$

is finite for any pair $r, s \in \mathbb{R}$ The random process (time series) $\{X(t),\ t \in \mathbb{R}\}$ is said to be (weakly, second-order, or wide-sense) stationary if

(i) $E|X(t)|^2 < \infty$,

(ii) $EX(t) = m$, for all $t \in \mathbb{R}$, and

(iii) $\gamma_X(r,s) = \gamma_X(r+t, s+t)$ for all r, s, and $t \in \mathbb{R}$.

or

(iii)' $\mathcal{COV}(X_{t+h}, X_t) = \gamma_X(h)$.

■ If the index space for parameter t is not $\mathbb{R}$ but the set of integers, $\mathbb{Z}$ random process $X(t)$ is called random sequence or time series and often indexed as $X_t, t \in \mathbb{Z}$.

■ Examples:

(i) *White noise* is a stationary sequence Z_t such that $EZ_t = 0$ and $\gamma(h) = \sigma^2 \cdot \delta_h$; in notation, $Z_t \sim \mathcal{WN}(0, \sigma^2)$.

(ii) The *moving average* MA(q) process, defined as

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q},\ Z_t \sim \mathcal{WN}(0, \sigma^2),$$

has autocovariance function

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & |h| \le q \\ 0 & |h| > q. \end{cases}$$

(iii) The *autoregressive* AR(1) process, $X_t - \phi X_{t-1} = Z_t,\ Z_t \sim \mathcal{WN}(0, \sigma^2)$, has autocovariance function

$$\gamma(h) = \sigma^2 \frac{\phi^{|h|}}{1 - \phi^2}.$$

■ Fourier transformation of autocorrelation (or autocovariance) function leads to spectral density $f(\omega)$

$$f(\omega) = \int_R \gamma(h) e^{-ih\omega} dh,$$

which is non-negative by Wiener-Khinchine theorem. Properly normalized indeed represents a density in a probabilistic sense. Depending on the definition of Fourier transformation, in our case f should be divided by $2\pi\gamma(0)$.

■ The function $f(\omega)$ is also called power-spectrum since $E|X(t)|^2 = 1/(2\pi) \int_R f(\omega) d\omega$, and $E|X(t)|^2$ represents the "power" of zero-mean signal $X(t)$.

■ It is possible define a counterpart of a spectral density of a nonstationary process (pseudo-spectrum) if linear filtering produces a stationary process. An example is spectrum for the fractional Brownian motion (later).

■ A stationary process $Y(t)$ is called long-range dependent (LRD) process if its autocorrelation function or spectral density behave as

$$\gamma_Y(h) \sim C_\gamma |h|^{\alpha-1}, \; h \to \infty, \alpha \in (0,1),$$

or

$$f_Y(\omega) \sim C_f |\omega|^{-\alpha}, \; \omega \to 0, \alpha \in (0,1),$$

where C_γ and C_f are two related constants. These two relations are equivalent subject to mild assumptions.

SS + LRD

■ Let $X(t), t \in R$ be $H - ss$ process. If its increments are stationary, i.e, if the distribution of $X(t+h) - X(t)$ is independent of t, it will be called $H - sssi$ process.

The following theorem gives the form of autocorrelation function of any $H - sssi$ process with finite second moment.

Let $X(t), t \in \mathbb{R}$ be an $H - sssi$ process for which $E|X(1)|^2 < \infty$. Then,

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

Proof: From H-ss and stationarity of increments property,

$$
\begin{aligned}
EX(t)X(s) &= \frac{1}{2}\left[E(X(t)^2) + E(X(s)^2) - E[X(t) - X(s)]^2\right] \\
&= \frac{1}{2}\left[t^{2H}E(X(1)^2) + s^{2H}E(X(1)^2) - E[X(|t-s|) - X(0)]^2\right] \\
&= \frac{E|X(1)|^2}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right].
\end{aligned}
$$

Increments of SSSI processes

■ Let $X(t)$ be an H-sssi process with $0 < H < 1$ and $E|X(1)|^2 < \infty$. Define stationary sequence of random variables $Y(n)$ as

$$Y(n) = X(n+1) - X(n).$$

■ If $\gamma_Y(n)$ is the autocorrelation function for $Y(n)$, i.e., $\gamma_Y(n) = EY(n)Y(0)$, then if $H = 1/2$, $\gamma_Y(n) = 0$, for $n \geq 1$ and if $H \neq 1/2$, it is possible to find an explicit expression for $\gamma(n)$. Using the fact that $X(0) = 0$ we find,

$$
\begin{aligned}
\gamma_Y(n) = EY(1)Y(n) &= EX(1)(X(n+1) - X(n)) \\
&= E(X(1)X(n+1) - E(X(1)X(n)) \\
&= \frac{E|X(1)|^2}{2} \left[(n+1)^{2H} - n^{2H} + (n-1)^{2H} \right].
\end{aligned}
$$

■ If the expressions $(n \pm 1)^{2H}$ are replaced by their polynomial expansions $n^{2H} \pm 2Hn^{2H-1} + H(2H-1)n^{2H-2} + \ldots$, the following asymptotic result holds

$$
\lim_{n \to \infty} \frac{\gamma_Y(n)}{H(2H-1)E|X(1)|^2 \; n^{2H-2}} = 1. \tag{1}
$$

In other words, $\gamma_Y(n) = O(n^{2H-2})$.

■ Note that series $\sum_n |\gamma_Y(n)|$ converges if $2 - 2H > 1$ or, equivalently, if $0 < H < 1/2$. For such H, expression $2H - 1$ in (1) is negative, and correlations $\gamma_Y(n)$ are negative. If $1/2 < H < 1$ correlations $\gamma(n)$ are positive, but $\sum_n |\gamma_Y(n)| = \infty$, since $2 - 2H < 1$.

Aggregation

- Long range dependent process $Y(n)$ is asymptotically second-order self-similar, i.e., the second order moments of Y and aggregated time series $Y^{(m)}$ coincide.
- The aggregate series $Y^{(m)}(k)$ is defined as series of averages of non-overlapping blocks of size m from the sequence $Y(n)$,

$$Y^{(m)}(k) = \frac{Y(km - m + 1) + \cdots + Y(km)}{m},$$

-

$$VarY^{(m)} \sim 1/m \; \gamma_Y(0) + \sum_{k=1}^{m-1} k^{2H-2}(m-k) \sim m^{2H-2}.$$

Informally, $Y^{(m)}(k)$ and $Y(n)$ look similar at all scales. This asymptotic behavior of the variance of aggregated process, $VarY^{(m)} \sim m^{2H-2}$, can be used for inference about H.

Self-Similar Processes as Fixed Points of Renormalization Groups

- For a sequence of random variables $Y_0, Y_1, \dots, Y_n, \dots$, $H > 0$, and $N \geq 1$ define the transformation

$$T_{N,H} : Y \longmapsto T_{N,H} Y = \left\{ (T_{N,H} Y)_j,\ j = 0, 1, 2, \dots \right\},$$

where

$$(T_{N,H} Y)_j = \frac{1}{N} \sum_{k=jN}^{(j+1)N-1} Y_k,\ j = 0, 1, 2, \dots$$

- It is easy to verify that

$$T_{M,H} \circ T_{N,H} = T_{MN,H},$$

and the sequence of transformations $\{T_{N,H}, N = 1, 2, \dots\}$ forms a multiplicative semigroup, called *renormalization group of index H.*

■ A stationary sequence $Y = \{Y_0, Y_1, Y_2, \dots\}$ is H-ss, if Y is a fixed point of a renormalization group $\{T_{N,H}, N = 1, 2, \dots\}$, i.e., for any $N \geq 1$ and any finite set of indices J,

$$\left\{ (T_{N,H} Y)_j ,\ j \in J \right\} \stackrel{d}{=} \{Y_j ,\ j \in J\} .$$

■ Theorem: Let $\{X(t), t \geq 0\}$ be H-sssi process with $H > 0$. Then the increment process $Y_j = X(j+1) - X(j),\ j = 0, 1, 2, \dots$ is a fixed point of the renormalization group $\{T_{N,H}, N = 1, 2, \dots\}$.

Proof. Since $\{X(t), t \geq 0\}$ is H-ss for $H > 0$, by Lamperti's Theorem, $X(0) = 0$. Fix any $k > 0$, and real numbers $a_1, \ldots, a_k$. Fix $N \geq 1$. Then,

$$\begin{aligned}
\sum_{i=0}^{k} a_i (T_{N,H} Y)_i &= \sum_{i=0}^{k} a_i \, \frac{1}{N^H} \sum_{j=iN}^{(i+1)N-1} Y_j \\
&= \sum_{i=0}^{k} a_i \, \frac{1}{N^H} (X((i+1)N) - X(iN)) \\
&\stackrel{d}{=} \sum_{i=0}^{k} a_i (X(i+1) - X(i)) \\
&= \sum_{i=0}^{k} a_i Y_i.
\end{aligned}$$

Time domain exponent n^{2H-2} corresponds to frequency domain exponent ω^{1-2H}. Thus frequency scaling is,

$$\alpha = 2H - 1,$$

Fractional Brownian Motion (fBm) and Fractional Gaussian Noise (fGn)

■ Fractional Brownian motion (fBm)is a generalization of standard Brownian motion (Wienner Process). The Brownian motion $B(t)$ is standardly defined as a random process satisfying the following four requirements:

(i) $B(0) = 0$,

(ii) For any choice n and $0 \leq t_1 < t_2 < \cdots < t_n$, the increments $B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent and stationary;

(iii) For fixed t, $B(t)$ is the Gaussian random variable with zero mean and variance t, and

(iv) $B(t)$ is a continuous function of t, a.s.

■ It is straightforward to check that the Brownian motion is an 1/2-sssi process, for $W(t) = a^{-1/2}B(at)$ conforms to properties (i)-(iv).

■ The Brownian motion is a Gaussian process and Gaussian processes are fully determined by their second order properties. Thus, the Brownian motion is *unique* Gaussian process having covariance function
$\gamma(t,s) = EB(t)B(s) = 1/2(t+s-|t-s|) = \min\{t,s\}$.

■ If H-sssi process is Gaussian, it is unique and it is called **fractional Brownian motion.**

Def. A zero mean Gaussian process $B_H(t)$ is called fractional Brownian motion with Hurst exponent H, if

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

where $E|X(1)|^2 = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}$.

■ The process $B_H(t)$ is unique, in the sense that class of all fractional Brownian motions with exponent H coincides with the class of all Gaussian $H-ss$ processes. However, a Gaussian process is $H-ss$ with independent increments, if and only if it $H = 1/2$, i.e., if it is a Brownian motion.

■ Alternative Definition (1)

(i) $B_H(t)$ has stationary increments,

(ii) $B_H(0) = 0$ and $\mathbb{E}B_H(t) = 0 \forall t$

(iii) $\mathbb{E}B_H(t)^2 = |t|^{2H}, \forall t$

(iv) $B_H(t)$ is a Gaussian process

(v) $B_H(t)$ has continuous paths.

■ The difference process, $Y_n = B_H(n+1) - B_H(n)$ is called fractional Gaussian noise (fGn). The covariance function of fGn is

$$\gamma(h) = \frac{E|X(1)|^2}{2}\left[(h+1)^{2H} - 2h^{2H} + (h-1)^{2H}\right],$$

which is, as we discussed, in agreement with general H-sssi processes.

■ Since $B_H(t)$ is the unique Gaussian H-sssi process, it follows from Renormalization Theorem that in the class of Gaussian stationary sequences, the fractional Gaussian noise is *unique* fixed point of the renormalization group $\{T_{N,H}, N = 1, 2, \ldots\}$.

■ Alternative Definition (2): Mandelbrot and Van Ness (1968)

$$\begin{aligned} B_H(0) &= 0, \text{ and} \\ B_H(t) &= C \cdot \left[\int_{-\infty}^{t} (t-s)^{H-1/2} B(ds) - \int_{0}^{t} (-s)^{H-1/2} B(ds) \right] \\ &= \int K(t,s)\ B(ds), \quad \text{where } C = \Gamma(H+1/2)/(\Gamma(2H+1)\sin(\pi H))^{1/2}. \end{aligned}$$

■ $B_H(t)$ is selfsimilar with self-similarity index H

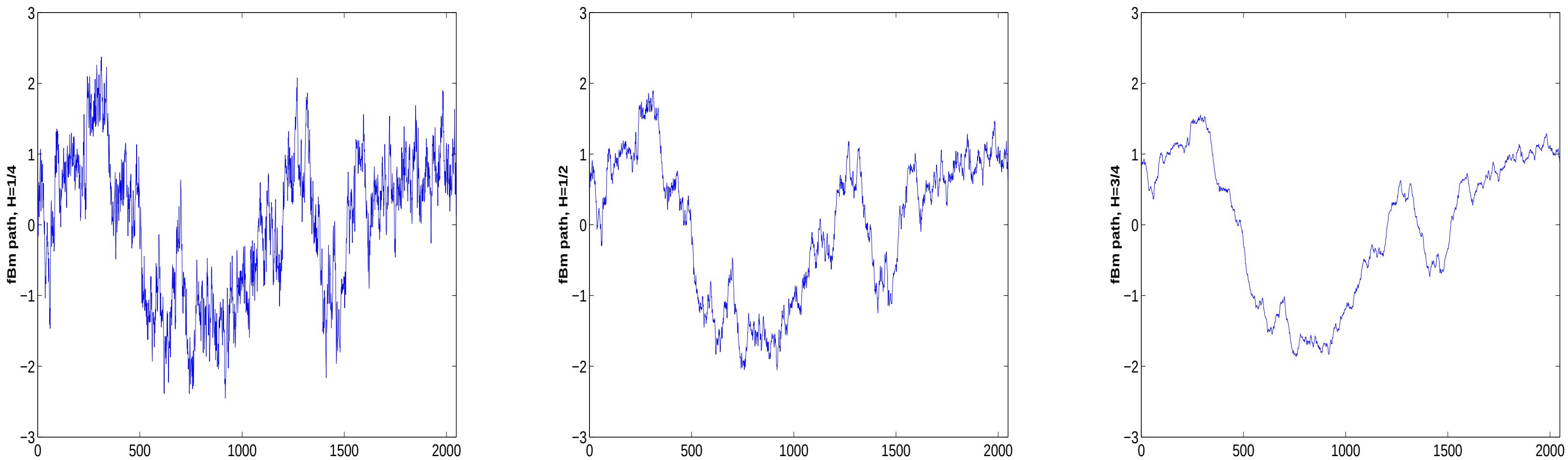
$$\begin{aligned} B_H(ct) &= \int K(ct,s)\ B(ds) = c^{H-1/2} \int K(t,s/c)\ B(ds) \\ &= c^{H-1/2} \int K(t,v)\ B(dv) = c^{H-1/2} c^{1/2} \int K(t,v)\ B(dv) = c^H B_H(t). \end{aligned}$$

■ Similarly to integral representation of Mandelbrot and Van Ness, fBm allows the so called harmonizable representation,

$$B_H(t) = \int_{\mathbb{R}} \frac{e^{it\omega} - 1}{|\omega|^{H+1/2}} B(d\omega),$$

where $B(d\omega)$ is the Wiener measure.

■ Sample paths of fractional Brownian motion are behaving similarly to those of standard Brownian motion. They are continuous almost surely for all $H \in (0, 1)$ and nowhere differentiable. The fractal (Hausdorff) dimension of sample paths is $D = 2 - H$. That means that for small H (say, $H < 0.5$) the sample paths are quite irregular and *space-filling.*



■ Simulated paths of fractional Brownian motion, (a) $H = 1/4$, (b) $H = 1/2$, and (c) $H = 3/4$.

■ It is interesting that sample paths of fractional Brownian motions are continuous in H, a result of Peltier and Lévy-Véhel (2000).

Convex Rearrangements of Davidov

■ Let $\Delta(i) = B_H(\frac{i+1}{n}) - B_H(\frac{i}{n})$, for $i = 1, \ldots, n-1$. Let $\Delta_{1:n} \leq \ldots \Delta_{n:n}$ be the corresponding order statistics. Define the polynomial
$VB_{H,n}(t) = \sum_{i=0}^{[nt]-1} \Delta_{i:n} + (nt - [nt])\Delta_{[nt]:n}$.

■ **[Th]** Phillpe and Thilly (2000) showed that

$$\frac{VB_{H,n}(t)}{n^{1-H}\sqrt{C}} \to L(t),$$

where $L(t) = -\frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}\Phi^{-1}(t)$, and Φ is the standard Gaussian cdf.

■ This result can be utilized to estimate H.

Autoregressive, Fractionally Integrated, Moving Average Processes (ARFIMA)

- ARFIMA(p, d, q) introduced by Granger and Joyeux (1980) and Hosking (1981).

- It can be used for statistical modeling of time series with long memory. ARMA models (when $d = 0$) or ARIMA (when d is a positive integer) are special cases.

- In the terns of standard time back-lag operator B defined as $B^k Y(n) = Y(n-k),\ k = 0, 1, \ldots,$ the ARFIMA(p, d, q) model can be represented as:

$$\Phi_p(B)(1-B)^d Y(n) = \Theta_q(B)\epsilon(n), \quad n = 1, 2, \ldots$$

where $\Phi_p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ is the autoregressive polynomial and $\Theta_q(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ is the moving average polynomial; p and q are integers, and d is a real number from (-1/2, 1/2). The innovations $\epsilon(n)$ are assumed i.i.d. normal with zero mean and variance σ^2.

■ The ARFIMA(p, d, q) model can be interpreted as an ARMA(p, q) process

$$\Phi(B)Y(n) = \Theta(B)Z(n), n = 1, 2, \ldots$$

where the noise process $Z(n)$ is a fractionally differenced Gaussian noise,

$$(1 - B)^d Z(n) = \epsilon(n),\ \ \epsilon(n) \sim N(0, \sigma^2).$$

We will focus on ARFIMA$(0, d, 0)$ since, for example, ARFIMA(p, d, q) model can be generated by appropriate filtering of ARFIMA$(0, d, 0)$. For example, in MATLAB, to obtain an ARFIMA(p, d, q) simulation from an ARFIMA$(0, d, 0)$ run, one applies `filter` function,

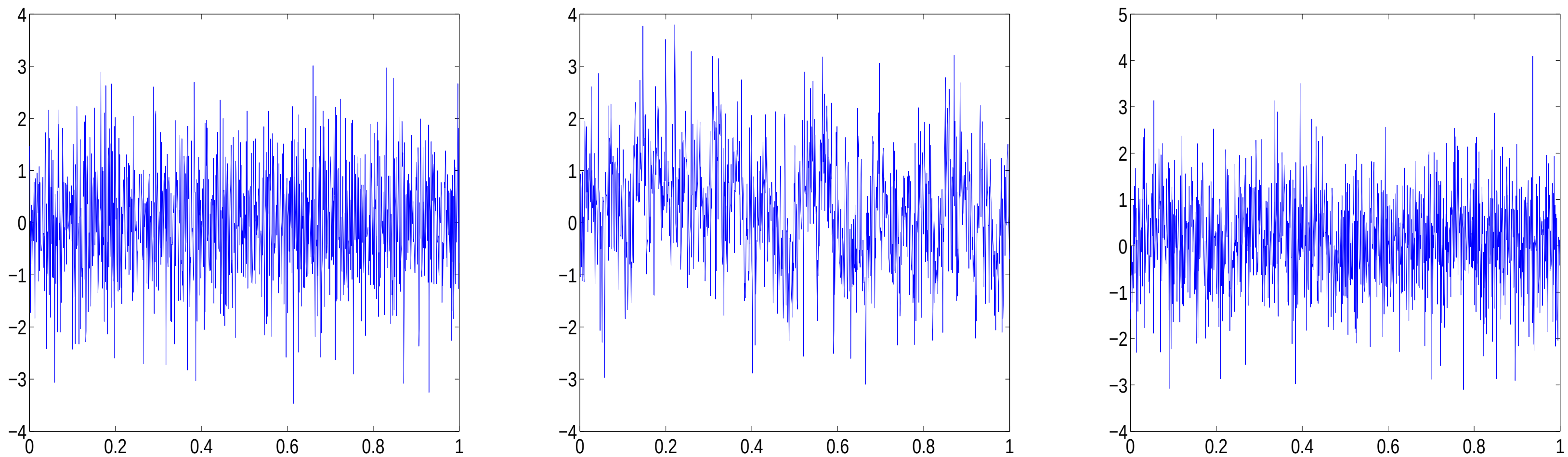
`arfimapdq = filter( b, a, arfima0d0 ),` where `a` $= [1 \quad -\phi_1 \quad -\phi_2 \quad \ldots \quad -\phi_p]$ and `b` $= [1 \quad \theta_1 \quad \theta_2 \quad \ldots \quad \theta_q]$.

■ The fractional difference operator can be expresses by the following binomial expansion:

$$
\begin{aligned}
(1-B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\
&= 1 - dB - \frac{1}{2} d(1-d)B^2 - \frac{1}{6} d(1-d)(2-d)B^3 - \cdots + \\
&\qquad (-1)^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} B^k + \ldots
\end{aligned}
$$

■ It holds $(-1)^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} = \prod_{0<j\leq k} \frac{k-1-d}{k} = c_k$, so ARFIMA$(0, d, 0)$ is an AR(∞) process. In particular, ARFIMA$(0, d, 0)$ is close to fractional Gaussian noise with parameter $H = d + 1/2$.

■ For $d > 1/2$ the ARFIMA process is not stationary, although it can be differenced to a stationary process. For $-0.5 < d < 0$ the process is called intermediate memory or *overdifferenced*, see Brockwell and Davis (1993). The ARFIMA model is stationary with long memory when when $0 < d < 1/2$, which is the most interesting case.



(a) ARFIMA(0,-0.3,0); (b) ARFIMA(0,0.3,0),; (c) ARFIMA(5,0.3,4) for $\phi = [3/10 \;\; 43/90 \;\; -19/90 \;\; -1/30 \;\; 1/45]; \theta = [-4/5 \;\; -11/45 \;\; 16/45 \;\; -4/45]$.

■ The process $Z(n)$ can also be represented as an moving average process of infinite order, since

$$Z(n) = (1-B)^{-d}\epsilon(n),\ \epsilon(n) \sim \mathcal{N}(0,\sigma^2) \tag{2}$$

can be written as

$$\begin{aligned} Z(n) &= \sum_{k=0}^{\infty} \binom{d+k-1}{k} \epsilon(n-k) \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}\ \epsilon(n-k). \end{aligned}$$

The covariance function of $Z(t)$ is given as

$$\begin{aligned} \gamma_Z(h) &= E(Z(n)Z(n+h)) = \\ &= \gamma(0)\frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(h+1-d)} \\ &= \gamma(0)\prod_{0<k\leq h} \frac{k-1+d}{k-d},\ h = 1,2,\ldots \end{aligned} \tag{3}$$

where $\gamma(0) = \sigma^2 \sum_{k=0}^{\infty} \binom{d+k-1}{k}^2 = \frac{\sigma^2}{\sqrt{\pi}\cdot 4^d}\frac{\Gamma(1/2-d)}{\Gamma(1-d)}$.

■ Autocorrelation function can be represented as

$$\rho(h) = \gamma(h)/\gamma(0) = \prod_{0<k\leq h} \frac{k-1+d}{k-d} = \frac{d(1+d)\ldots(h-1+d)}{(1-d)(2-d)\ldots(h-d)}, \quad h = 1, 2, \ldots.$$

■ Since $\frac{\Gamma(h+a)}{\Gamma(h+b)} \sim h^{a-b}$, when h is large, $\rho(h) \sim \frac{\Gamma(1-d)}{\gamma(d)} h^{2d-1}$. Obviously, when $0 < d < 0.5$, the exponent $-1 < 2d-1 < 0$ and the series $\sum_h \rho(h)$ diverges ("long memory").

■ By filtering considerations, the spectral density of an ARFIMA(p, d, q) process is

$$f(\omega) = \frac{\sigma^2}{2\pi} \, |1 - e^{-i\omega}|^{-2d} \frac{|\Theta(e^{-i\omega})|^2}{|\Phi(e^{-i\omega})|^2}.$$

Since $|1 - e^{-i\omega}|^2 = 4\sin^2(\omega/2) \sim 4(\frac{\omega}{2})^2$, as $\omega \to 0$, the spectral density of ARFIMA(p, d, q) process behaves as

$$f(\omega) \sim \frac{\sigma^2}{2\pi} \, |\omega|^{-2d} \frac{|\Theta(1)|^2}{|\Phi(1)|^2},$$

when ω is close to 0.

■ Expressions for autocovariance function for general ARFIMA(p, d, q) are complicated.