SIMDI 226 - I
Time series analysis : I

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

Telecom ParisTech

December 4, 2012

Plan

Examples of applications

Time series analysis based on stochastic modeling is applied in various fields :

- Health : physiological signal analysis (image analysis).
- Engineering : monitoring, anomaly detection, localizing/tracking.
- Audio data : analysis, synthesis, coding.
- Ecology : climatic data, hydrology.
- Econometrics : economic/financial data.
- Insurance : risk analysis.

Heartbeats

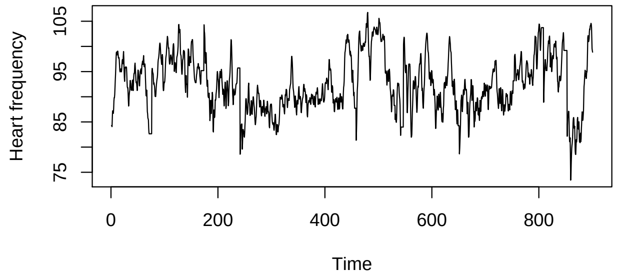

Figure: Heart rate of a resting person over a period of 900 seconds. This rate is defined as the number of heartbeats per unit of time. Here the unit is the minute and is evaluated every 0.5 seconds.

Internet traffic

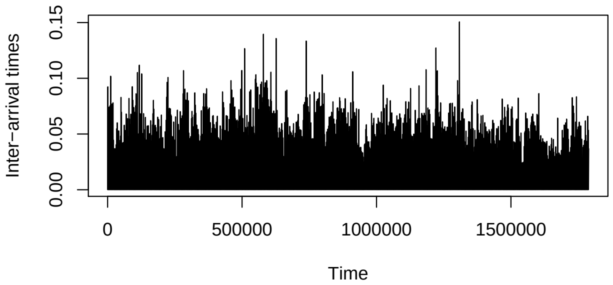

Figure: Inter-arrival times of TCP packets, expressed in seconds, obtained from a 2 hours record of the traffic going through an Internet link. `http://ita.ee.lbl.gov/`.

Speech audio data

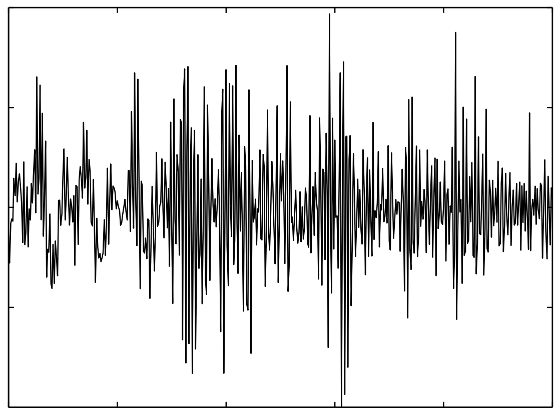

Figure: A speech audio signal with a sampling frequency equal to 8000 Hz. Record of the unvoiced fricative phoneme *sh* (as in *sh*arp).

Climatic data: wind speed

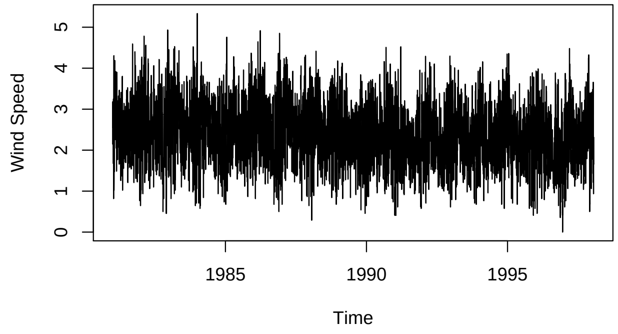

Figure: Daily record of the wind speed at Kilkenny (Ireland).

Climatic data: temperature changes

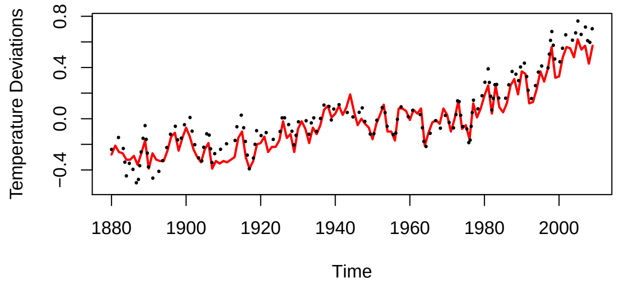

Figure: Global mean land-ocean temperature index (solid red line) and surface-air temperature index (dotted black line).
`http://data.giss.nasa.gov/gistemp/graphs/`.

Gross National Product of the USA

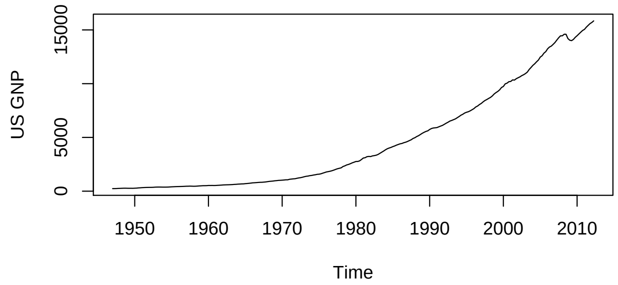

Figure: Quarterly growth national product (GNP) of the USA in Billions of $s.
`http://research.stlouisfed.org/fred2/series/GNP`.

GNP quarterly rate

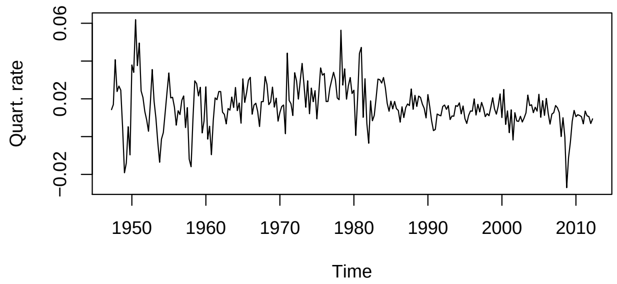

Figure: Quarterly rate of the US GNP.

Financial index

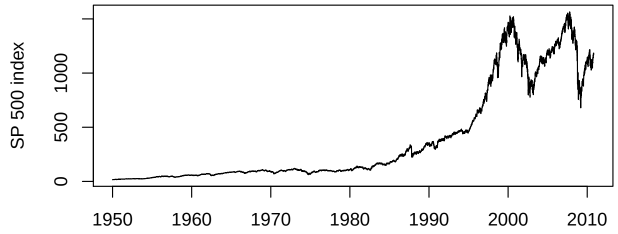

Figure: Daily open value of the Standard and Poor 500 index. This index is computed as a weighted average of the stock prices of 500 companies traded at the New York Stock Exchange (NYSE) or NASDAQ.

Financial index: log returns

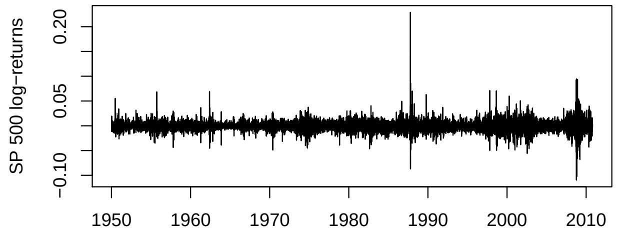

Figure: SP500 log-retruns.

Main goals of time series analysis

- Stochastic modelling : seasonal/linear trend + noise + "structural properties".
- Statistical inference : estimate the parameters of the model, classify or detect (signal presence, anomalies).
- Forecasting : based on a stochastic model, use historical data to predict future values.
- Filtering and tracking : estimate hidden variables and track them.

Stochastic modelling

Definition : random processes

A random or stochastic process valued in $(E, \mathcal{E})$ and indexed on T is a collection of random variables $(X_t)_{t \in T}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We will generally consider t as a time index, in which case $T = \mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}_+$... A spatial index can also be considered, say $T = \mathbb{R}^d$.

Note that a random vector of length n can be seen as a random process $(X_t)_{t \in T}$ with $T = \{1, \dots, n\}$.

Random path

Definition : path

Let $(X_t)_{t \in T}$ be a random process defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The path of the random experiment $\omega \in \Omega$ is defined as $(X_t(\omega))_{t \in T}$ viewed as an element of E^T.

We shall denote by X the random variable $\omega \mapsto (X_t(\omega))_{t \in T}$ valued in $(E^T, \mathcal{E}^{\otimes T})$, where $\mathcal{E}^{\otimes T}$ is the smallest σ-field containing the cylinder sets

$$A_1 \times \dots A_n \times E^{T \setminus \{1, \dots, n\}}, \qquad n \geq 1, \quad A_1, \dots, A_n \in \mathcal{E} .$$

It is also also the smallest σ-field on E^T which makes ξ_t measurable for all $t \in T$, where ξ_t is the canonical projection $\xi_t : (x_s)_{s \in T} \mapsto x_t$ from E^T to $(E, \mathcal{E})$.

Finite distributions

Define

$$\mathcal{I}(T) = \{I \subset T, I \text{ finite}\} .$$

Definition : law

Let $(X_t)_{t \in T}$ be a random process. The law of the process in the sense of finite distributions is defined as the collection of probability distributions

$$\mathbb{P} \circ X^{-1} \circ \Pi_I^{-1} \left(\prod_{t \in I} A_t \right) = \mathbb{P} \left(X_t \in A_t, t \in I \right), \qquad I \in \mathcal{I}(T),$$

where Π_I is the canonical projection $(x_s)_{s \in T} \mapsto (x_s)_{s \in I}$.

We will denote

$$X \stackrel{\text{fidi}}{=} Y ,$$

when X and Y have the same law in the sense of finite distributions.

From finite distributions to path distribution

Note that if $J \subset I$ are in $\mathcal{I}(T)$, then, for all $A \in \mathcal{E}^{\otimes J}$,

$$\mathbb{P} \circ X^{-1} \circ \Pi_J^{-1}(A) = \mathbb{P} \circ X^{-1} \circ \Pi_I^{-1} \left(A \times E^{I \setminus J} \right) . \tag{1}$$

Theorem : Kolmogorov

Let $(E, \mathcal{E})$ be a measurable space, T an arbitrary set of indices and $(\nu_I)_{I \in \mathcal{I}(T)}$ such that each ν_I is a probability on $(E^I, \mathcal{E}^{\otimes I})$. The two following assertions are equivalent.

(i) $(\nu_I)_{I \in \mathcal{I}(T)}$ satisfies the compatibility condition (1) for all $J \subseteq I$.

(ii) There is a unique probability ν_T on $(E^T, \mathcal{E}^{\otimes T})$ such that $\nu_I = \nu_T \circ \Pi_I^{-1}$ for all $I \in \mathcal{I}(T)$.

What is all this about?

In practice, we do not start with $X_t(\omega)$ defined for all $t \in T$ and $\omega \in \Omega$. Instead we have the following steps:

Step 1 Start with a collection of compatible finite distributions $(\nu_I)_{I\in\mathcal{I}(T)}$ on the measurable spaces $(E^I, \mathcal{E}^{\otimes I})$.

Step 2 Deduce the probability space $(E^T, \mathcal{E}^{\otimes T}, \nu_T)$ by the Kolmogorov theorem.

Step 3 Define the process X using the canonical process $X_t = \xi_t$, for all $t \in T$. Hence we get a process X on $(\Omega, \mathcal{F}, \nu_T)$ with the desired finite distributions.

Step 4 Define new processes by filtering X, for instance $W_t = g_t(X)$ where $g_t : E^T \to F$ is measurable for all t, or equivalently, $Y = g(X)$ where $g : E^T \to F^T$ is measurable.

Complementary facts

- In Step 1, the simplest possible finite distributions are often used, see the following item.
- Simulating a random process is often done by following the same steps with ν_I being the probability distribution of independent uniform random variables.
- Let $T = \mathbb{N}, \mathbb{Z}, \mathbb{R}_+$ or $\mathbb{R}$. The process X is adapted to a given filtration $(\mathcal{F}_t)_{t\in T}$ if, for all $t \in T$, X_t is $\mathcal{F}_t$-measurable. Example: natural filtration $\mathcal{F}_t = \sigma(X_s,\ s \le t)$.
- If the paths a.s. belong to a subset A of E^T which is endowed with a metric: e.g. $\mathcal{C}([0,1]) \subset \mathbb{R}^{[0,1]}$, it might be interesting to define the path as a random element of A endowed with the corresponding Borel σ-field.

Independent random variables

Let $(\nu_t)_{t\in T}$ be a collection of probability measures on $(E, \mathcal{E})$.
For all $I \in \mathcal{I}$, set the probability on E^I with independent component having marginal distributions $(\nu_t)_{t\in I}$:

$$\nu_I = \bigotimes_{t\in I} \nu_t \ , \qquad (2)$$

Then $(\nu_I)_{I\in\mathcal{I}}$ satisfies the compatibility condition (1).
We thus obtain a collection of independent random variables X_t such that $X_t \sim \nu_t$ for all $t \in T$.
If $\nu_t = \nu$ for all $t \in T$ we say that $(X_t)_{t\in T}$ is a is a collection of i.i.d. (independent and identically distributed) random variables with marginal distribution ν.

Gaussian processes

Let T be an arbitrary set of indices. Let $\mu = (\mu_t)_{t\in T}$ be real-valued and $(\gamma_{s,t})_{s,t\in T}$ be such that, for all $I \in \mathcal{I}(T)$

$$\Gamma_I = [\gamma_{s,t}]_{s,t\in I} \quad \text{is symmetric non-negative definite .}$$

Then there exists a process $(X_t)_{t\in T}$ on a probability space $(\Omega, \mathcal{F}, \xi)$ such that, for all $I \in \mathcal{I}(T)$

$$\mathbb{P} \circ X^{-1} \circ \Pi_I^{-1} = \mathcal{N}\left((\mu_t)_{t\in I}, \Gamma_I\right) \ .$$

We will denote $X \sim \mathcal{N}(\mu, \gamma)$ and say that X is a Gaussian process with mean μ and covariance γ.

Shift and backshift operators

Suppose that $T = \mathbb{Z}$ or $T = \mathbb{N}$.

Definition : Shift and backshift operators

Let the shift operator $\mathrm{S} \ : \ E^T \to E^T$ be defined by

$$\mathrm{S}(x) = (x_{t+1})_{t\in T} \quad \text{for all} \quad x = (x_t)_{t\in T} \in E^T \ .$$

For all $\tau \in T$, we define S^τ by

$$\mathrm{S}^\tau(x) = (x_{t+\tau})_{t\in T} \quad \text{for all} \quad x = (x_t)_{t\in T} \in E^T \ .$$

The operator S^{-1} is called the backshift operator.

Strict stationarity

Definition : Strict stationarity

Let $X = (X_t)_{t \in T}$ be a random process defined on $(\Omega, \mathcal{F}, \xi)$ with $T = \mathbb{Z}$ or $T = \mathbb{N}$. We say that X is stationary in the strict sense if

$$X \overset{\text{fidi}}{=} \mathrm{S} \circ X ,$$

which is equivalent to

$$\mathbb{P} \circ X^{-1} = \mathbb{P} \circ X^{-1} \circ \mathrm{S}^{-1} .$$

Examples

Examples based on finite distributions

- A constant process, $X_t = X_0$ for all $t \in T$ is stationary.
- A sequence of independent random variables is strictly stationary if and only if they are i.i.d.
- $X \sim \mathcal{N}(\mu, \Gamma)$ is stationary if and only if $\mu_t = \mu_0$ and $\gamma_{s,t} = \gamma_{s-t,0}$ for all $s, t \in T$.

Examples based on stationarity preserving filters

Examples of filters $g : E^T \to F^T$ preserving stationarity :

- $g = \sum_k h_k \mathrm{S}^{-k} : x \mapsto h \star x$ for a finitely supported sequence h.
- $g : x \mapsto \left(h \circ \Pi_I \circ \mathrm{S}^t(x)\right)_{t \in T}$ with $h : E^I \to E$ with $I \in \mathcal{I}(T)$.
- Time reversing operator: $g : (x_t)_{t \in \mathbb{Z}} \mapsto (x_{-t})_{t \in \mathbb{Z}}$.

L^2 space

We set $E = \mathbb{C}^d$. We denote

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X \ \mathbb{C}^d\text{-valued r.v. such that } \mathbb{E}\left[|X|^2\right] < \infty \right\} .$$

$(L^2, \langle , \rangle)$ is a Hilbert space with

$$\langle X, Y \rangle = \mathbb{E}\left[X^T \overline{Y}\right] .$$

Definition : L^2 Processes

The process $\mathbf{X} = (\mathbf{X}_t)_{t \in T}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{C}^d$ is an L^2 process if $\mathbf{X}_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for all $t \in T$.

Mean and covariance functions

Let $\mathbf{X} = (\mathbf{X}_t)_{t \in T}$ be an L^2 process.

- Its mean function is defined by $\mu(t) = \mathbb{E}[\mathbf{X}_t]$,
- Its covariance function is defined by

$$\Gamma(s,t) = \mathrm{cov}(\mathbf{X}_s, \mathbf{X}_t) = \mathbb{E}\left[\mathbf{X}_s \mathbf{X}_t^H\right] - \mathbb{E}[\mathbf{X}_s]\,\mathbb{E}[\mathbf{X}_t]^H .$$

Scalar case

Let $\mathbf{X} = (\mathbf{X}_t)_{t \in T}$ be an L^2 process with mean function μ and covariance function Γ. This is equivalent to say that for all $\mathbf{u} \in \mathbb{C}^d$, $\mathbf{u}^H \mathbf{X}$ is a scalar L^2 process with mean function $\mathbf{u}^H \mu$ and covariance function $\mathbf{u}^H \Gamma \mathbf{u}$.

Covariance function, examples

Hermitian symmetry, non-negative definiteness ($d = 1$)

For all $I \in \mathcal{I}(T)$, $\Gamma_I = \mathrm{Cov}([X(t)]_{t \in I}) = [\gamma(s,t)]_{s,t \in I}$ is a hermitian non-negative definite matrix.

Examples

- L^2 random variables with marginals $(\nu_t)_{t \in T}$ have mean $\mu(t) = \int x\, \nu_t(\mathrm{d}x) = \nu_t(\mathrm{Id})$ and covariance

$$\Gamma(s,t) = \begin{cases} \nu_t(\mathrm{Id}\mathrm{Id}^H) - \nu_t(\mathrm{Id})\nu_t(\mathrm{Id})^H & \text{if } s = t, \\ 0 & \text{otherwise.} \end{cases}$$

- A Gaussian process is an L^2 process whose law is entirely determined by its mean and covariance functions.

Weakly stationary processes

Let $T = \mathbb{Z}$. Let X be an L^2 strictly stationary process with mean function μ and covariance function Γ. Then $\mu(t) = \mu(0)$ and $\gamma(s,t) = \gamma(s-t, 0)$ for all $s, t \in T$.

Definition: Weak stationarity

We say that a random process X is weakly stationary with mean μ and autocovariance function $\gamma : \mathbb{Z} \to \mathbb{C}$ if it is L^2 with mean function $t \mapsto \mu$ and covariance function $(s,t) \mapsto \gamma(s-t)$.

The autocorrelation function is defined (when $\gamma(0) > 0$) as

$$\rho(t) = \frac{\gamma(t)}{\gamma(0)} .$$

Examples based on finite distributions

An L^2 strictly stationary process is weakly stationary.

- The constant L^2 process has constant autocovariance function.
- A sequence of L^2 i.i.d. random variables is called a strong white noise, denoted by $X \sim \mathrm{IID}(\mu, \sigma^2)$.
- An L^2 process X with constant mean μ and constant diagonal covariance function equal to σ^2 is called a weak white noise. It is denoted by $X \sim \mathrm{WN}(\mu, \sigma^2)$.

Examples based on stationarity preserving linear filters

Let X be weakly stationary with mean μ and autocovariance γ.

Define $Y = g(X)$. In the following examples, Y is weakly stationary with mean μ' and autocovariance γ' for

- $g =$time reversing operator: $\mu' = \mu$ and $\gamma' = \gamma$.
- $g = \sum_k \mathrm{h}_k \, \mathrm{S}^{-k} : x \mapsto \mathrm{h} \star x$ for a finitely supported sequence h :
  - $\mu' = \mu \sum_k \mathrm{h}_k$
  - $\gamma'(\tau) = \sum_{\ell,k} \mathrm{h}_k \overline{\mathrm{h}_\ell} \gamma(\tau + \ell - k)$.

Heartbeats : autoregression

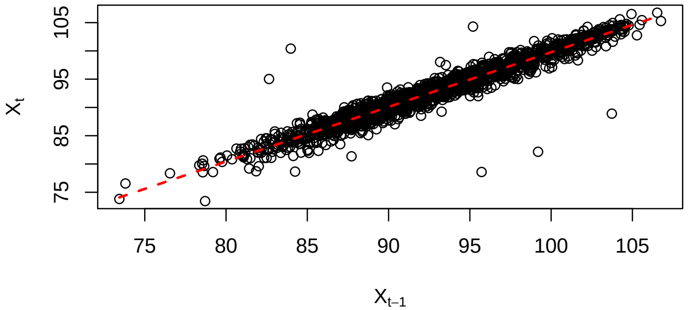

Figure: Illustration of $\gamma(1)$: X_t VS X_{t-1} for the heartbeats data. The red dashed line is the best linear fit.

Empirical estimates

Suppose you want to estimate the mean and the autocovariance from a sample $X_1, \ldots, X_n$. Define the empirical mean as

$$\widehat{\mu}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \; ,$$

and the empirical autocovariance and autocorrelation functions

$$\widehat{\gamma}_n(h) = \frac{1}{n} \sum_{k=1}^{n-|h|} (X_k - \widehat{\mu}_n)(X_{k+|h|} - \widehat{\mu}_n) \quad \text{and} \quad \widehat{\rho}_n(h) = \frac{\widehat{\gamma}_n(h)}{\widehat{\gamma}_n(0)} \; .$$

Heartbeats : autocorrelation (empirical)

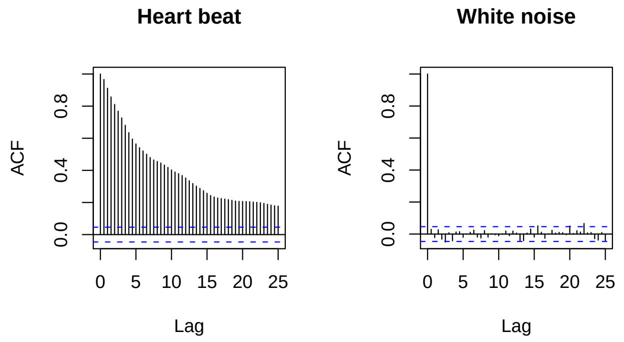

Figure: Left : empirical autocorrelation $\widehat{\rho}_n(h)$ of heartbeat data for $h = 0, \ldots, 100$. Right : the same from a simulated white noise sample with same length.

Spectral measure

Given a function $\gamma : \mathbb{Z} \to \mathbb{C}$, does there exist a weakly stationary process $(X_t)_{t\in\mathbb{Z}}$ with autocovariance γ?

Herglotz Theorem

Let $\gamma : \mathbb{Z} \to \mathbb{C}$. Then the two following assertions are equivalent:

(i) γ is hermitian symmetric and non-negative definite.

(ii) There exists a non-negative measure ν on the torus $\mathbb{T}$, called the spectral measure, such that, for all $t \in \mathbb{Z}$, $\gamma(t) = \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}\lambda t} \, \nu(\mathrm{d}\lambda)$.

Spectral density

If moreover $\gamma \in \ell^1(\mathbb{Z})$, these assertions are equivalent to

$$f(\lambda) = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}\lambda t} \gamma(t) \geq 0 \text{ for all } \lambda \in \mathbb{R} ,$$

and ν has density f ($\nu(\mathrm{d}\lambda) = f(\lambda)\mathrm{d}\lambda$).

f is called the spectral density.

Examples

- Let $X \sim \mathrm{WN}(\mu, \sigma^2)$. Then $f(\lambda) = \frac{\sigma^2}{2\pi}$.
- Let X be a weakly stationary process with spectral measure ν. Let $Y = \sum_k \mathrm{h}_k \, \mathrm{S}^{-k} \circ X$ for a finitely supported sequence h. Then Y is weakly stationary process with spectral measure ν' having density $\lambda \mapsto \left|\sum_k \mathrm{h}_k \mathrm{e}^{-\mathrm{i}\lambda k}\right|^2$ with respect to ν,

$$\nu'(\mathrm{d}\lambda) = \left|\sum_k \mathrm{h}_k \mathrm{e}^{-\mathrm{i}\lambda k}\right|^2 \nu(\mathrm{d}\lambda) .$$

A special one : the harmonic process

Let $(A_k)_{1\le k\le N}$ be N real valued L^2 random variables. Denote $\sigma_k^2 = \mathrm{Var}(A_k)$. Let $(\Phi_k)_{1\le k\le N}$ be N i.i.d. random variables with a uniform distribution on $[-\pi,\pi]$, and independent of $(A_k)_{1\le k\le N}$. Define

$$X_t = \sum_{k=1}^{N} A_k \cos(\lambda_k t + \Phi_k) , \qquad (3)$$

where $(\lambda_k)_{1\le k\le N} \in [-\pi,\pi]$ are N frequencies. The process (X_t) is called an *harmonic process*. It satisfies $\mathbb{E}[X_t] = 0$ and, for all $s,t \in \mathbb{Z}$,

$$\mathbb{E}[X_s X_t] = \frac{1}{2}\sum_{k=1}^{N} \sigma_k^2 \cos(\lambda_k (s-t)) .$$

Random fields with orthogonal increments

In the following we let $(\mathbb{X},\mathcal{X})$ be a measurable space.

Definition : Random fields with orthogonal increments

Let η be a non-negative measure on $(\mathbb{X},\mathcal{X})$. Let $W = (W(A))_{A\in\mathcal{X}}$ be an L^2 process indexed by $\mathcal{X}$. It is called a random field with orthogonal increments and intensity measure η if it satisfies the following conditions.

(i) For all $A \in \mathcal{X}$, $\mathbb{E}[W(A)] = 0$.

(ii) For all $A, B \in \mathcal{X}$, $\mathrm{Cov}(W(A), W(B)) = \eta(A\cap B)$.

Consequence

For all $A, B \in \mathcal{X}$ such that $A\cap B = \emptyset$, $W(A)$ and $W(B)$ are uncorrelated and $W(A\cup B) = W(A) + W(B)$.

Example

We denote by δ_λ the Dirac mass at point λ.

Let λ_k, $k = 1, \dots, n$ be fixed elements of $\mathbb{X}$. Let $Y_1, \dots, Y_n$ be centered L^2 uncorrelated random variables with variances $\sigma_1^2, \dots, \sigma_n^2$. Then

$$W = \sum_{k=1}^{n} Y_k \, \delta_{\lambda_k}$$

is a random field with orthogonal increments and intensity measure

$$\eta = \sum_{k=1}^{n} \sigma_k^2 \, \delta_{\lambda_k} \; .$$

Stochastic integral

Let W be a random field with orthogonal increments defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with intensity measure η on $(\mathbb{X}, \mathcal{X})$.
The stochastic integral with respect to W is defined by the following steps.

Step 1 We set $W(\mathbb{1}_A) = W(A)$, which defines a unitary operator from $\{\mathbb{1}_A, \; A \in \mathcal{X}\} \subset L^2(\mathbb{X}, \mathcal{X}, \eta)$ to $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Step 2 Extend this unitary operator linearly on $\mathrm{Span}\,(\mathbb{1}_A, \; A \in \mathcal{X})$.

Step 3 Extend this unitary operator continuously to $L^2(\mathbb{X}, \mathcal{X}, \eta) = \overline{\mathrm{Span}}\,(\mathbb{1}_A, \; A \in \mathcal{X})$.

Step 4 One obtains a unitary operator $L^2(\mathbb{X}, \mathcal{X}, \eta) \to L^2(\Omega, \mathcal{F}, \mathbb{P})$, denoted by

$$W(g) = \int g \, \mathrm{d}W \; .$$

Application

Let W be a random field with orthogonal increments with intensity measure η on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$.
Define, for all $t \in \mathbb{Z}$,

$$X_t = \int e^{it\lambda} \, dW(\lambda) ,$$

Then we have $\mathbb{E}[X_t] = 0$ and

$$\mathrm{Cov}(X_s, X_t) = \langle X_s, X_t \rangle = \langle e^{is\cdot}, e^{it\cdot} \rangle = \int_{\mathbb{T}} e^{i(s-t)\lambda} \, d\eta(\lambda) ,$$

We get a centered weakly stationary process with spectral measure η.

Construction of the spectral random field

Step 1 Let $(X_t)_{t\in\mathbb{Z}}$ be a centered weakly stationary with spectral measure η. Define

$$\mathcal{H}_\infty^X = \overline{\mathrm{Span}}\,(X_t,\, t \in \mathbb{Z}) \ .$$

(the closure is taken in L^2.)

Step 2 As previously, we can extend $X_t \mapsto e^{it\cdot}$ linearly and continuously as a unitary operator from $\mathcal{H}_\infty^X$ to $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \eta)$.

Step 3 Since $\overline{\mathrm{Span}}\,(e^{it\cdot},\, t \in \mathbb{Z}) = L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \eta)$, this operator is bijective.

Step 4 Let $\widehat{X}$ be its inverse operator. Then $\widehat{X}$ is a random field with orthogonal increments with intensity measure η on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$.

Spectral representation

By construction, every $Y \in \mathcal{H}_\infty^X$ can be represented as

$$Y = \int g(\lambda) \, d\widehat{X}(\lambda) \ .$$

for some $g \in L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \eta)$.

In particular, for all $t \in \mathbb{Z}$,

$$X_t = \int e^{it\lambda} \, d\widehat{X}(\lambda) \ .$$

This expression is called the spectral representation of X.