Stationary and weakly stationary time series

François Roueff
http://perso.telecom-paristech.fr/~roueff/

Telecom Paris

BGDIA703 2023/2024

Outline

Basic (important) definitions

Definition : Data set

A data set is a collection of values, say $X_{1:n} = X_1, \ldots, X_n$. Time series data sets are usually sampled from recorded measurements.

Basic (important) definitions

Definition : Data set

A data set is a collection of values, say $X_{1:n} = X_1, \ldots, X_n$. Time series data sets are usually sampled from recorded measurements.

Definition : Model

A model is a collection of probability distributions. The data set is assumed to be distributed according to one of them.

Basic (important) definitions

Definition : Data set

A data set is a collection of values, say $X_{1:n} = X_1, \ldots, X_n$. Time series data sets are usually sampled from recorded measurements.

Definition : Model

A model is a collection of probability distributions. The data set is assumed to be distributed according to one of them.

Definition : Statistic

A statistic is any value which can be computed from the data.

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,
- a corresponding list of dates (days, years, seconds...)
- or, equivalently, a starting date (in some unit), and a frequency.

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,
- a corresponding list of dates (days, years, seconds...)
- or, equivalently, a starting date (in some unit), and a frequency. For instance :
  - a date in years and frequency$= 12$ corresponds to monthly data,

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,
- a corresponding list of dates (days, years, seconds...)
- or, equivalently, a starting date (in some unit), and a frequency. For instance :
  - a date in years and frequency$= 12$ corresponds to monthly data,
  - a date in years and frequency$= 4$ corresponds to quarterly data,

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,
- a corresponding list of dates (days, years, seconds...)
- or, equivalently, a starting date (in some unit), and a frequency. For instance :
  - a date in years and frequency$= 12$ corresponds to monthly data,
  - a date in years and frequency$= 4$ corresponds to quarterly data,
  - a date in days and frequency$= 1$ corresponds to daily data,
  - $\vdots$

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,
- a corresponding list of dates (days, years, seconds...)
- or, equivalently, a starting date (in some unit), and a frequency. For instance :
  - a date in years and frequency$= 12$ corresponds to monthly data,
  - a date in years and frequency$= 4$ corresponds to quarterly data,
  - a date in days and frequency$= 1$ corresponds to daily data,
  - $\vdots$
- Remarks :
  - There may be missing values (usually expressed as 'NA')

Raw data set

A time series $X_1, \ldots, X_n$ is usually presented as

- a list of real values $X_1, \ldots, X_n$ in a data or spreadsheet file,
- a corresponding list of dates (days, years, seconds...)
- or, equivalently, a starting date (in some unit), and a frequency. For instance :
  - a date in years and frequency$= 12$ corresponds to monthly data,
  - a date in years and frequency$= 4$ corresponds to quarterly data,
  - a date in days and frequency$= 1$ corresponds to daily data,
  - $\vdots$
- Remarks :
  - There may be missing values (usually expressed as 'NA')
  - In the case of multivariate time series, each variable usually corresponds to a column (so each row corresponds to a date).

Example : US GNP data set

```
# Title:                 Gross National Product
# Source:                U.S. Department of Commerce
# Frequency:             Quarterly
DATE,VALUE
1947-01-01,238.1
1947-04-01,241.5
1947-07-01,245.6
1947-10-01,255.6
1948-01-01,261.7
1948-04-01,268.7
1948-07-01,275.3
1948-10-01,276.6
1949-01-01,271.3
1949-04-01,267.5
1949-07-01,268.9
```

⋮

First step : remove an additive trend

- Consider a time series $X_1, \ldots, X_n$.

First step : remove an additive trend

- Consider a time series $X_1, \ldots, X_n$.
- It often includes an additive trend, for instance:

First step : remove an additive trend

- Consider a time series $X_1, \ldots, X_n$.
- It often includes an additive trend, for instance:
  - a polynomial trend,

First step : remove an additive trend

- Consider a time series $X_1, \dots, X_n$.
- It often includes an additive trend, for instance:
  - a polynomial trend,
  - a seasonal trend.

First step : remove an additive trend

- Consider a time series $X_1, \ldots, X_n$.
- It often includes an additive trend, for instance:
  - a polynomial trend,
  - a seasonal trend.
- Then $X = D + Y$ where D belongs to a finite dimensional space V and Y is a centered random process.

First step : remove an additive trend

- Consider a time series $X_1, \dots, X_n$.
- It often includes an additive trend, for instance:
  - a polynomial trend,
  - a seasonal trend.
- Then $X = D + Y$ where D belongs to a finite dimensional space V and Y is a centered random process.
- Trends can be estimated or removed by
  - Fitting the trend using least squares,

$$\widehat{D} = \operatorname*{argmin}_{d \in V} \sum_t |X_t - d_t|^2 \ .$$

First step : remove an additive trend

- Consider a time series $X_1, \ldots, X_n$.
- It often includes an additive trend, for instance:
  - a polynomial trend,
  - a seasonal trend.
- Then $X = D + Y$ where D belongs to a finite dimensional space V and Y is a centered random process.
- Trends can be estimated or removed by
  - Fitting the trend using least squares,

$$\widehat{D} = \underset{d \in V}{\operatorname{argmin}} \sum_t |X_t - d_t|^2 \;.$$

  - Or applying a well chosen filter F_ψ, such that $\mathrm{F}_\psi(D) = 0$ and thus

$$\mathrm{F}_\psi(X) = \mathrm{F}_\psi(Y) \;.$$

R code example: Johnson and Johnson trend adjustment

trend-adjustment.html

Second step : choose a stochastic model on the random part

In time series analysis, one is interested in modeling the time dependence in the trend adjusted data $Y_1, \ldots, Y_n$.

Second step : choose a stochastic model on the random part

In time series analysis, one is interested in modeling the time dependence in the trend adjusted data $Y_1, \ldots, Y_n$.

This can be done by using

- a parametric model.

Second step : choose a stochastic model on the random part

In time series analysis, one is interested in modeling the time dependence in the trend adjusted data $Y_1, \ldots, Y_n$.

This can be done by using

- a parametric model.

Example

$Y_1, \ldots, Y_n$ is the sample of a Gaussian ARMA(p, q) model with (unknown) parameter $\vartheta = (\theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p, \sigma^2)$.

Second step : choose a stochastic model on the random part

In time series analysis, one is interested in modeling the time dependence in the trend adjusted data $Y_1, \ldots, Y_n$.

This can be done by using

- a parametric model.

Example

$Y_1, \ldots, Y_n$ is the sample of a Gaussian ARMA(p, q) model with (unknown) parameter $\vartheta = (\theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p, \sigma^2)$.

- a non-parametric model.

Second step : choose a stochastic model on the random part

In time series analysis, one is interested in modeling the time dependence in the trend adjusted data $Y_1, \ldots, Y_n$.

This can be done by using

- a parametric model.

Example

$Y_1, \ldots, Y_n$ is the sample of a Gaussian ARMA(p, q) model with (unknown) parameter $\vartheta = (\theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p, \sigma^2)$.

- a non-parametric model.

Example

$Y_1, \ldots, Y_n$ is the sample of a centered stationary Gaussian process with (unknown) autocovariance γ (or spectral density f).

Third step : estimate parameters, test hypotheses

Once a model is fixed for $Y_1, \ldots, Y_n$, it can be used to

- ▷ Estimate a parameter of the model such as ϑ, $\gamma(t)$, σ^2, f, ...

Third step : estimate parameters, test hypotheses

Once a model is fixed for $Y_1, \ldots, Y_n$, it can be used to

- ▷ Estimate a parameter of the model such as ϑ, $\gamma(t)$, σ^2, f, ...

  $\rightarrow$ Define an estimator, say $\widehat{\vartheta}_n$, which is a statistic based on the sample $Y_1, \ldots, Y_n$.

Third step : estimate parameters, test hypotheses

Once a model is fixed for $Y_1, \ldots, Y_n$, it can be used to

- Estimate a parameter of the model such as ϑ, $\gamma(t)$, σ^2, f, …

  $\rightarrow$ Define an estimator, say $\widehat{\vartheta}_n$, which is a statistic based on the sample $Y_1, \ldots, Y_n$.

- Test hypotheses, for instance

$$H_0 = \{Y \text{ is white noise}\} \quad \text{against} \quad H_1 = \{Y \text{ is ARMA}(p,q)\}$$

Third step : estimate parameters, test hypotheses

Once a model is fixed for $Y_1, \ldots, Y_n$, it can be used to

- Estimate a parameter of the model such as ϑ, $\gamma(t)$, σ^2, f, ...

  $\rightarrow$ Define an estimator, say $\widehat{\vartheta}_n$, which is a statistic based on the sample $Y_1, \ldots, Y_n$.

- Test hypotheses, for instance

$$H_0 = \{Y \text{ is white noise}\} \quad \text{against} \quad H_1 = \{Y \text{ is ARMA}(p,q)\}$$

  $\rightarrow$ Define a statistical test, say

$$\delta = \begin{cases} 1 & \text{if } T_n > t_n \,, \\ 0 & \text{otherwise} \,, \end{cases}$$

  where T_n is a statistic based on the sample $Y_1, \ldots, Y_n$ and t_n is a threshold.

Stationary and ergodic models

- We see $Y_1, \dots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.

Stationary and ergodic models

- We see $Y_1, \ldots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.

Stationary and ergodic models

- We see $Y_1, \ldots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.

Stationary and ergodic models

- We see $Y_1, \ldots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.
- Ergodic means that observing one path $(Y_t)_{t \in T}$ allows one to recover the distribution entirely.

Stationary and ergodic models

- We see $Y_1, \ldots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.
- Ergodic means that observing one path $(Y_t)_{t \in T}$ allows one to recover the distribution entirely.
- Examples:
  - An IID process is stationary and ergodic;

Stationary and ergodic models

- We see $Y_1, \dots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.
- Ergodic means that observing one path $(Y_t)_{t \in T}$ allows one to recover the distribution entirely.
- Examples:
  - An IID process is stationary and ergodic;
  - A sequence of independent variables that are not identically distributed is not stationary;

Stationary and ergodic models

- We see $Y_1, \ldots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.
- Ergodic means that observing one path $(Y_t)_{t \in T}$ allows one to recover the distribution entirely.
- Examples:
  - An IID process is stationary and ergodic;
  - A sequence of independent variables that are not identically distributed is not stationary;
  - a random walk is not stationary but its increments are;

Stationary and ergodic models

- We see $Y_1, \dots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.
- Ergodic means that observing one path $(Y_t)_{t \in T}$ allows one to recover the distribution entirely.
- Examples:
  - An IID process is stationary and ergodic;
  - A sequence of independent variables that are not identically distributed is not stationary;
  - a random walk is not stationary but its increments are;
  - A sequence of variables $(Y_t)_{t \in \mathbb{Z}}$ that is constant, *i.e.* $Y_t = Y_0$ for all t, is stationary but is not ergodic;

Stationary and ergodic models

- We see $Y_1, \ldots, Y_n$ as a finite sample of a stochastic process $(Y_t)_{t \in T}$, with $T = \mathbb{N}$ or $\mathbb{Z}$, distributed according to a statistical model.
- It is crucial to work with stationary and ergodic models.
- Stationary means that the model is shift invariant, more details hereafter.
- Ergodic means that observing one path $(Y_t)_{t \in T}$ allows one to recover the distribution entirely.
- Examples:
  - An IID process is stationary and ergodic;
  - A sequence of independent variables that are not identically distributed is not stationary;
  - a random walk is not stationary but its increments are;
  - A sequence of variables $(Y_t)_{t \in \mathbb{Z}}$ that is constant, *i.e.* $Y_t = Y_0$ for all t, is stationary but is not ergodic;
  - A Markov chain on a finite state space can be made stationary by choosing the initial state adequately. If it is irreducible, then it is ergodic.

R code example: dependent data

non-iid-data.html

Random processes

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

or

  - A random path X valued in the trajectory space $(E^T, \mathcal{E}^{\otimes T})$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

Random processes

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

or

  - A random path X valued in the trajectory space $(E^T, \mathcal{E}^{\otimes T})$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$.
- We can go back and forth from one definition to another
  - Define $X : \Omega \to E^T$ by $X(\omega) = (X_t(\omega))_{t \in T}$.
  - Define for all $t \in T$, $X_t : \Omega \to E$ by $X_t(\omega) = \xi_t(X(\omega))$, where $\xi_t : (x_s)_{s \in T} \mapsto x_t$ denotes the canonical projection from E^T to E.

Law of X

- Exactly as usual random variables, the random path X admits a distribution defined by

$$\mathbb{P}^X : A \mapsto \mathbb{P}^X(A) = \mathbb{P}(X \in A) ,$$

where A is a subset of possible trajectories ($A \in \mathcal{E}^{\otimes T}$).

Law of X

- Exactly as usual random variables, the random path X admits a distribution defined by

$$\mathbb{P}^X : A \mapsto \mathbb{P}^X(A) = \mathbb{P}(X \in A) ,$$

where A is a subset of possible trajectories ($A \in \mathcal{E}^{\otimes T}$).

- Alternatively, one can consider finite-dimensional distributions: for all finite subset $I = \{t_1, \ldots, t_n\} \subset T$ of indices,
  - denote by Π_I is the canonical projection $(x_t)_{t \in T} \mapsto (x_t)_{t \in I}$,
  - denote by X_I the random vector $(X_t)_{t \in I} = \Pi_I \circ X$,
  - denote by $\mathbb{P}^{X_I}$ the distribution of X_I, which is defined by

$$\mathbb{P}^{X_I}\left(A_{t_1} \times \cdots \times A_{t_n}\right) = \mathbb{P}\left(X_t \in A_t,\, t \in I\right),$$

  where $A_{t_1} \ldots A_{t_n}$ are subset of E (in $\mathcal{E}$).
  - The collection $\left(\mathbb{P}^{X_I}\right)_{I \in \mathcal{I}(T)}$ are called the fidi distributions.

- $\mathbb{P}^X$ can be constructed from $\left(\mathbb{P}^{X_I}\right)_{I \in \mathcal{I}(T)}$ (Kolmogorov theorem).

Independent and i.i.d. processes

- Let $(\nu_t)_{t \in T}$ be a collection of probability measures on $(E, \mathcal{E})$.

Independent and i.i.d. processes

- Let $(\nu_t)_{t\in T}$ be a collection of probability measures on $(E, \mathcal{E})$.
- An independent process $X = (X_t)_{t\in T}$ with marginals $(\nu_t)_{t\in T}$ is a process that satisfies

$$\mathbb{P}\left(X_t \in A_t \text{ for all } t \in I\right) = \prod_{t\in I} \mathbb{P}\left(X_t \in A_t\right) = \prod_{t\in I} \nu_t(A_t)\ .$$

Independent and i.i.d. processes

- Let $(\nu_t)_{t \in T}$ be a collection of probability measures on $(E, \mathcal{E})$.
- An independent process $X = (X_t)_{t \in T}$ with marginals $(\nu_t)_{t \in T}$ is a process that satisfies

$$\mathbb{P}\left(X_t \in A_t \text{ for all } t \in I\right) = \prod_{t \in I} \mathbb{P}\left(X_t \in A_t\right) = \prod_{t \in I} \nu_t(A_t) \; .$$

- If $\nu_t = \nu$ for all $t \in T$ we say that $(X_t)_{t \in T}$ is a independent and identically distributed (IID) with marginal ν.

Gaussian processes

- Let T be an arbitrary set of indices.

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 $I \in \mathcal{I}(T)$

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

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 $I \in \mathcal{I}(T)$

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

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

$$\mathbb{P}^{X_I} = \mathcal{N}\left((\mu_t)_{t \in I}, \Gamma_I\right) ,$$

which is equivalent to have for all $(x_s)_{s \in I} \in E^I$

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\sum_{s \in I} x_s X_s}\right] = \exp\left(\mathrm{i}\sum_{s \in I} x_s \mu_s - \frac{1}{2}\sum_{s \in I} x_s \gamma_{s,t} x_t\right) .$$

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 $I \in \mathcal{I}(T)$

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

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

$$\mathbb{P}^{X_I} = \mathcal{N}\left((\mu_t)_{t \in I}, \Gamma_I\right) ,$$

which is equivalent to have for all $(x_s)_{s \in I} \in E^I$

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\sum_{s \in I} x_s X_s}\right] = \exp\left(\mathrm{i}\sum_{s \in I} x_s \mu_s - \frac{1}{2}\sum_{s \in I} x_s \gamma_{s,t} x_t\right) .$$

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

Shift and backshift operators

- Suppose that $T = \mathbb{Z}$ or $T = \mathbb{N}$. Random processes are then called time series.

Shift and backshift operators

- Suppose that $T = \mathbb{Z}$ or $T = \mathbb{N}$. Random processes are then called time series.
- The shift operator $\mathrm{S} \; : \; E^T \to E^T$ is 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

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

Shift and backshift operators

- Suppose that $T = \mathbb{Z}$ or $T = \mathbb{N}$. Random processes are then called time series.
- The shift operator $\mathrm{S} \; : \; E^T \to E^T$ is 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, denoted by B.

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

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

Examples based on finite distributions

- A constant process,

$$X_t = X_0 \qquad \text{for all } t \in T$$

is stationary.

Examples based on finite distributions

- A constant process,

$$X_t = X_0 \qquad \text{for all } t \in T$$

  is stationary.
- A sequence of independent random variables is strictly stationary if and only if they are indentically distributed. (Thus it is an i.i.d. process).

Examples based on finite distributions

- A constant process,

$$X_t = X_0 \qquad \text{for all } t \in T$$

  is stationary.
- A sequence of independent random variables is strictly stationary if and only if they are indentically distributed. (Thus it is an i.i.d. process).
- Gaussian processes : $X \sim \mathcal{N}(\mu, \Gamma)$ 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

- Let $g : E^{\mathbb{Z}} \to F^{\mathbb{Z}}$.

Examples based on stationarity preserving filters

- Let $g : E^{\mathbb{Z}} \to F^{\mathbb{Z}}$.
- Suppose that X is stationary.

Examples based on stationarity preserving filters

- Let $g : E^{\mathbb{Z}} \to F^{\mathbb{Z}}$.
- Suppose that X is stationary.
- Is $g(X)$ stationary ?
  - If
    $$g \circ \mathrm{S} = \mathrm{S} \circ g \;,$$
    then the answer is yes:
    $$g(X) = g \circ X \stackrel{\text{fidi}}{=} g \circ \mathrm{S} \circ X = \mathrm{S} \circ g \circ X = \mathrm{S} \circ g(X) \;.$$

Examples based on stationarity preserving filters

- Let $g : E^{\mathbb{Z}} \to F^{\mathbb{Z}}$.
- Suppose that X is stationary.
- Is $g(X)$ stationary ?
  - If
$$g \circ \mathrm{S} = \mathrm{S} \circ g \;,$$
    then the answer is yes:
$$g(X) = g \circ X \overset{\text{fidi}}{=} g \circ \mathrm{S} \circ X = \mathrm{S} \circ g \circ X = \mathrm{S} \circ g(X) \;.$$
  - Time reversing operator: $g : (x_t)_{t\in\mathbb{Z}} \mapsto (x_{-t})_{t\in\mathbb{Z}}$. Here
$$g \circ \mathrm{S} = \mathrm{S}^{-1} \circ g \;.$$

Starting from an IID sequence

- Consider the case where $(\epsilon_t)_{t \in \mathbb{Z}}$ is IID (hence stationary).

Starting from an IID sequence

- Consider the case where $(\epsilon_t)_{t\in\mathbb{Z}}$ is IID (hence stationary).
- A linear process is defined by

$$X_t = \sum_{k\in\mathbb{Z}} \psi_k \epsilon_{t-k} \ ,$$

where $(\psi_t)_{t\in\mathbb{Z}}$ is an ℓ^2 sequence.

Starting from an IID sequence

- Consider the case where $(\epsilon_t)_{t\in\mathbb{Z}}$ is IID (hence stationary).
- A linear process is defined by

$$X_t = \sum_{k\in\mathbb{Z}} \psi_k \epsilon_{t-k} \ ,$$

where $(\psi_t)_{t\in\mathbb{Z}}$ is an ℓ^2 sequence.
- All stationary Markov processes $(X_n)_{n\in\mathbb{N}}$ can be defined iteratively as

$$X_{n+1} = f(X_n, \epsilon_n) \ ,$$

for a convenient mapping f.

Starting from an IID sequence

- Consider the case where $(\epsilon_t)_{t\in\mathbb{Z}}$ is IID (hence stationary).
- A linear process is defined by

$$X_t = \sum_{k\in\mathbb{Z}} \psi_k \epsilon_{t-k} \;,$$

  where $(\psi_t)_{t\in\mathbb{Z}}$ is an ℓ^2 sequence.
- All stationary Markov processes $(X_n)_{n\in\mathbb{N}}$ can be defined iteratively as

$$X_{n+1} = f(X_n, \epsilon_n) \;,$$

  for a convenient mapping f.
- Many stationary and ergodic models can be defined as a Causal "Bernoulli" shift:

$$X_t = \phi(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots)$$

  for a convenient mapping ϕ.

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.

Mean and covariance functions

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

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

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

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

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

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

Linear combinations $\longrightarrow$ 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 $\mathbf{u}^H\Gamma\mathbf{u}$.

Scalar case $E = \mathbb{C}$, examples

- If $E = \mathbb{C}$, then the auto-covariance function is $\mathbb{C}$-valued, and denoted by γ.

Scalar case $E = \mathbb{C}$, examples

- If $E = \mathbb{C}$, then the auto-covariance function is $\mathbb{C}$-valued, and denoted by γ.

Hermitian symmetry, non-negative definiteness

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.

Scalar case $E = \mathbb{C}$, examples

- If $E = \mathbb{C}$, then the auto-covariance function is $\mathbb{C}$-valued, and denoted by γ.

Hermitian symmetry, non-negative definiteness

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 independent random variables $(X_t)_{t \in \mathbb{Z}}$ have mean $\mu(t) = \mathbb{E}(X_t)$ and covariance

$$\gamma(s,t) = \begin{cases} \mathrm{var}(X_t) & \text{if } s = t, \\ 0 & \text{otherwise.} \end{cases}$$

- 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.

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 $s,t \in T$.

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 $\mu \in \mathbb{C}$ and autocovariance function $\gamma : \mathbb{Z} \to \mathbb{C}$ if it is L^2 with mean function $t \mapsto \mu$ and covariance function $(s,t) \mapsto \gamma(s-t)$.

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 $\mu \in \mathbb{C}$ and autocovariance function $\gamma : \mathbb{Z} \to \mathbb{C}$ if it is L^2 with mean function $t \mapsto \mu$ and covariance function $(s,t) \mapsto \gamma(s-t)$.

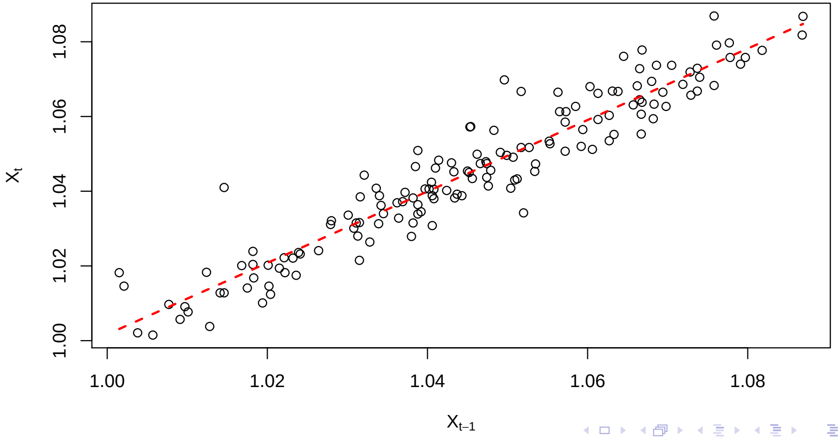
- The autocorrelation function is then defined (when $\gamma(0) > 0$) by

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \in [-1, 1] \, .$$

Autocorrelation=slope of regression line

We have, for all $t \in \mathbb{Z}$ and $h = 1, 2, \ldots,$

$$X_{t+h} = \text{Constant} + \rho(h) X_t + \epsilon_{t,h} \quad \text{with} \quad \epsilon_{t,h} \perp \operatorname{Span}(1, X_t) \ .$$



Partial Autocorrelation

- We can also write, for all $t \in \mathbb{Z}$ and $h = 1, 2, \ldots,$

$$X_t = \text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k} + \kappa(h) X_{t-h} + \epsilon_{t,h}$$

where

$$\epsilon_{t,h} \perp \mathrm{Span}\left(1, X_{t-1}, \ldots, X_{t-h}\right) \ .$$

Partial Autocorrelation

- We can also write, for all $t \in \mathbb{Z}$ and $h = 1, 2, \ldots,$

$$X_t = \text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k} + \kappa(h) X_{t-h} + \epsilon_{t,h}$$

  where

$$\epsilon_{t,h} \perp \mathrm{Span}\left(1, X_{t-1}, \ldots, X_{t-h}\right) \; .$$

- In the following, we plot successively
  - X_t as a function of X_{t-h}, compared to the regression line;

Partial Autocorrelation

- We can also write, for all $t \in \mathbb{Z}$ and $h = 1, 2, \ldots,$

$$X_t = \text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k} + \kappa(h) X_{t-h} + \epsilon_{t,h}$$

  where

$$\epsilon_{t,h} \perp \mathrm{Span}\left(1, X_{t-1}, \ldots, X_{t-h}\right) \ .$$

- In the following, we plot successively
  - X_t as a function of X_{t-h}, compared to the regression line;
  - X_t as a function of X_{t-h}, compared to the predictor
    $\text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k} + \kappa(h) X_{t-h}$.

Partial Autocorrelation

- We can also write, for all $t \in \mathbb{Z}$ and $h = 1, 2, \ldots,$

$$X_t = \text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k} + \kappa(h) X_{t-h} + \epsilon_{t,h}$$

where

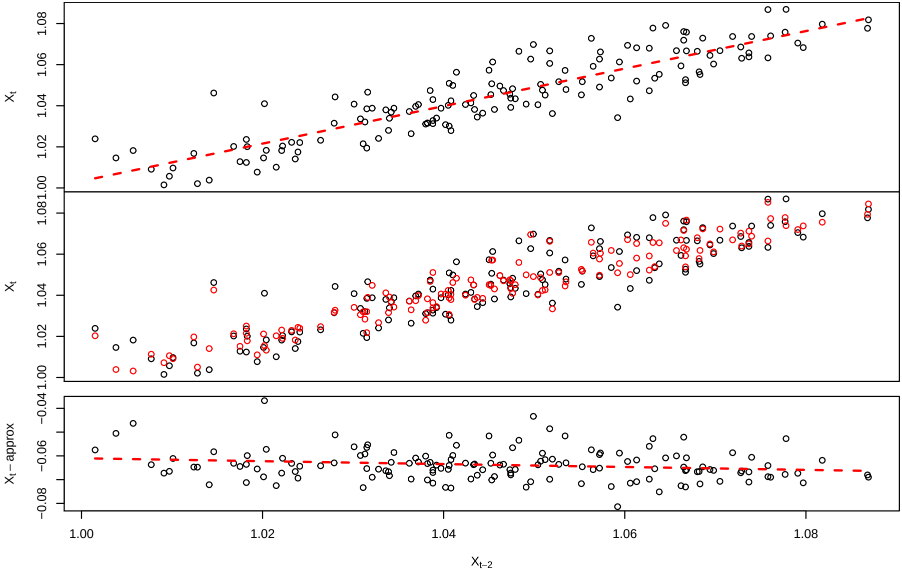
$$\epsilon_{t,h} \perp \mathrm{Span}\left(1, X_{t-1}, \ldots, X_{t-h}\right) \ .$$

- In the following, we plot successively
  - X_t as a function of X_{t-h}, compared to the regression line;
  - X_t as a function of X_{t-h}, compared to the predictor

    $$\text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k} + \kappa(h) X_{t-h}.$$

  - $X_t - \left(\text{Constant} + \sum_{k=1}^{h-1} \phi_k X_{t-k}\right)$ as a function of X_{t-h}, compared to the regression line $X_{t-h} \mapsto \kappa(h) X_{t-h}$.

Partial Autocorrelation=slope of partial regression



Examples

- An L^2 strictly stationary process is weakly stationary.

Examples

- An L^2 strictly stationary process is weakly stationary.
- The constant L^2 process has constant autocovariance function.

Examples

- An L^2 strictly stationary process is weakly stationary.
- The constant L^2 process has constant autocovariance function.

Strong and weak white noise

- A sequence of L^2 i.i.d. random variables is called a strong white noise, denoted by $X \sim \mathrm{IID}(\mu, \sigma^2)$.

Examples

- An L^2 strictly stationary process is weakly stationary.
- The constant L^2 process has constant autocovariance function.

Strong and weak white noise

- 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 $X \sim \mathrm{WN}(\mu, \sigma^2)$. (It does not have to be i.i.d.)

Examples based on stationarity preserving linear filters

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

Examples based on stationarity preserving linear filters

- Let X be weakly stationary with mean μ and autocovariance γ.
- In the following examples, $Y = g(X)$ is weakly stationary with mean μ' and autocovariance γ'.
  - Let g be the time reversing operator: $Y_t = X_{-t}$. Then

$$\mu' = \mu \qquad \text{and} \qquad \gamma' = \overline{\gamma} \; .$$

Examples based on stationarity preserving linear filters

- Let X be weakly stationary with mean μ and autocovariance γ.
- In the following examples, $Y = g(X)$ is weakly stationary with mean μ' and autocovariance γ'.
  - Let g be the time reversing operator: $Y_t = X_{-t}$. Then

$$\mu' = \mu \qquad \text{and} \qquad \gamma' = \overline{\gamma}\ .$$

  - Let $g = \sum_k \psi_k \, \mathrm{B}^k$: $Y_t = \sum_k \psi_k X_{t-k}$ with $\psi \in \ell^1$. Then

$$\mu' = \mu \sum_k \psi_k$$

$$\gamma'(\tau) = \sum_{\ell,k} \psi_k \overline{\psi_\ell} \gamma(\tau + \ell - k) \tag{1}$$

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?

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?
- A necessary and sufficient condition is that $\Gamma_I = [\gamma(s-t)]_{s,t\in I}$ is a hermitian non-negative definite matrix for all finite subset $I \subset \mathbb{Z}$.

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?
- A necessary and sufficient condition is that $\Gamma_I = [\gamma(s-t)]_{s,t\in I}$ is a hermitian non-negative definite matrix for all finite subset $I \subset \mathbb{Z}$.
- We say that γ is hermitian symmetric and non-negative definite.

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?
- A necessary and sufficient condition is that $\Gamma_I = [\gamma(s-t)]_{s,t\in I}$ is a hermitian non-negative definite matrix for all finite subset $I \subset \mathbb{Z}$.
- We say that γ is hermitian symmetric and non-negative definite.

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 finite non-negative measure ν on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ such that,

$$\text{for all } t \in \mathbb{Z}, \quad \gamma(t) = \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}\lambda t}\, \nu(\mathrm{d}\lambda)\ . \tag{2}$$

When these two assertions hold, ν is uniquely defined by (2).

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 (that is, $\nu(\mathrm{d}\lambda) = f(\lambda)\mathrm{d}\lambda$).

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 (that is, $\nu(\mathrm{d}\lambda) = f(\lambda)\mathrm{d}\lambda$).

Definition : spectral measure and spectral density

If γ is the autocovariance of a weakly stationary process X, the corresponding measure ν is called the spectral measure of X. Whenever the spectral measure ν admits a density f, it is called the spectral density function.

Examples

- Let $X \sim \mathrm{WN}(\mu, \sigma^2)$. Then $f(\lambda) = \frac{\sigma^2}{2\pi}$.

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 covariance function γ/spectral measure ν. Define

$$Y_t = \sum_k \psi_k X_{t-k}$$

  for a sequence $\psi \in \ell^1$.

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 covariance function γ/spectral measure ν. Define

$$Y_t = \sum_k \psi_k X_{t-k}$$

  for a sequence $\psi \in \ell^1$.
- Recall that Y is a weakly stationary process with covariance function

$$\gamma'(\tau) = \sum_{\ell,k} \psi_k \overline{\psi_\ell} \gamma(\tau + \ell - k) \ .$$

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 covariance function γ/spectral measure ν. Define

$$Y_t = \sum_k \psi_k X_{t-k}$$

for a sequence $\psi \in \ell^1$.
- Recall that Y is a weakly stationary process with covariance function

$$\gamma'(\tau) = \sum_{\ell,k} \psi_k \overline{\psi_\ell} \gamma(\tau + \ell - k) \; .$$

- Then Y is a weakly stationary process with spectral measure ν' having density $\lambda \mapsto \left|\sum_k \psi_k \mathrm{e}^{-\mathrm{i}\lambda k}\right|^2$ with respect to ν,

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

A special one : the harmonic process

Let $(A_k)_{1\le k\le N}$ be N real valued L^2 random variables. Denote $\sigma_k^2 = \mathbb{E}\left[A_k^2\right]$. Let $(\Phi_k)_{1\le k\le N}$ be N i.i.d. random variables with a uniform distribution on $[0, 2\pi]$, and independent of $(A_k)_{1\le k\le N}$. Define

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

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

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

Hence X is weakly stationary with autocovariance

$$\gamma(t) = \frac{1}{2}\sum_{k=1}^{N} \sigma_k^2 \cos(\lambda_k t) = \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}\lambda t} \left(\frac{1}{4}\sum_{k=1}^{N} \sigma_k^2 (\delta_{-\lambda_k}(\mathrm{d}\lambda) + \delta_{\lambda_k}(\mathrm{d}\lambda)) \right) \ .$$

Empirical estimates

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

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

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

Empirical estimates

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

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

- Define the empirical autocovariance and autocorrelation functions as

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

$$\widehat{\rho}_n(h) = \frac{\widehat{\gamma}_n(h)}{\widehat{\gamma}_n(0)} \;.$$

Periodogram

- The previous formula only work for $h = -n+1, \ldots, n-1$.

Periodogram

- The previous formula only work for $h = -n+1, \dots, n-1$.
- Define $\widehat{\gamma}_n(h) = \widehat{\rho}_n(h) = 0$ for all $|h| \geq n$.

Periodogram

- The previous formula only work for $h = -n+1, \ldots, n-1$.
- Define $\widehat{\gamma}_n(h) = \widehat{\rho}_n(h) = 0$ for all $|h| \geq n$.
- Now $\widehat{\gamma}_n$ is defined on $\mathbb{Z}$ and satisfies

$$\widehat{\gamma}_n(h) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda h} \, I_n(\lambda) \, \mathrm{d}\lambda \,,$$

where I_n is called the (raw) periodogram and is defined by

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{k=1}^{n} (X_k - \widehat{\mu}_n) \, \mathrm{e}^{-\mathrm{i}\lambda k} \right|^2 \,.$$

Periodogram

- The previous formula only work for $h = -n+1, \dots, n-1$.
- Define $\widehat{\gamma}_n(h) = \widehat{\rho}_n(h) = 0$ for all $|h| \geq n$.
- Now $\widehat{\gamma}_n$ is defined on $\mathbb{Z}$ and satisfies

$$\widehat{\gamma}_n(h) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda h} \, I_n(\lambda) \, \mathrm{d}\lambda \,,$$

where I_n is called the (raw) periodogram and is defined by

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{k=1}^{n} (X_k - \widehat{\mu}_n) \, \mathrm{e}^{-\mathrm{i}\lambda k} \right|^2 \,.$$

- $I_n(\lambda)$ can be seen as a (bad) estimator of the spectral density $f(\lambda)$.