TIME SERIES I

VÍTĚZSLAV VESELÝ

Department of Statistics and Operations Research
University of Malta [1]
Email: vvese01@um.edu.mt

CONTENTS

Date: May 20, 2002.

[1] Visiting lecturer, 2-nd semester from 4-th February to 12-th July 2002

$s := v$ or $v =: s$ ...denoting expression v by symbol s.
iff stands for *if and only if*.

Sets and mappings:

- $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$ ...natural numbers, integers, real and complex numbers, respectively.
- $\mathbb{Z}_N := \{0, 1, \ldots, N-1\}$ ...residuals modulo $N \in \mathbb{N}$.
- $\mathbb{R}^+$ ...the set of all non-negative real numbers.
- $\exp X$ ...class of all subsets of the set X.
- $\operatorname{card} M$ ...cardinality of a set M.
- $(\cdot)^+ : \mathbb{R} \to \mathbb{R}^+$ ...mapping defined as $(x)^+ = \max(0, x)$.
- (a, b), $[a, b]$, $(a, b]$, $[a, b)$ ...intervals on real line.
- $\mathfrak{I}(a, b) = \{x \mid \min(a, b) < x < \max(a, b)\}$
  $\mathfrak{I}[a, b] = \{x \mid \min(a, b) \leq x \leq \max(a, b)\}$.
- $f(A) := \{y \in Y \mid y = f(x), x \in A \subseteq X\}$ ...range (image) of set A under mapping $f : X \to Y$.
- $f^{-1}(B) := \{x \in X \mid f(x) \in B\} \subseteq X$ ...inverse image of set $B \subseteq Y$ under mapping $f : X \to Y$.
- I_A ...indicator function of set $A \subseteq X$:
  $I_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{otherwise} \end{cases}.$
- $A_n \uparrow$ ...increasing or non-decreasing sequence of numbers or sets.
- $A_n \downarrow$ ...decreasing or non-increasing sequence of numbers or sets.
- $\sum_{i=1}^n A_i := \bigcup_{i=1}^n A_i$ ...union of a family of sets which are pairwise disjoint.
- $A^c := X - A$ ...complement of set $A \subseteq X$ in X where X is a priori known from the context.
- $\underline{A} := \liminf_{n\to\infty} A_n := \bigcup_{n=1}^\infty \bigcap_{j=n}^\infty A_j$ ...inferior limit of a sequence of sets.

- $\overline{A} := \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j$ ...superior limit of a sequence of sets.
- $A = \lim_{n\to\infty} A_n$ iff $\underline{A} = \overline{A}$, clearly
  $A_n \uparrow A$ implies $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$ and
  $A_n \downarrow A$ implies $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Vectors and matrices:

- $\boldsymbol{x} := [x_1, \ldots, x_n]^T$ ...vector of numbers (by default column vector if not stated otherwise).
- $\boldsymbol{x} + h := [x_1 + h, \ldots, x_n + h]^T,\ h \in \mathbb{C}$.
- $\boldsymbol{x_t} := [x_{t_1}, \ldots, x_{t_k}]^T \in \mathbb{C}^k$ where $\boldsymbol{t} = [t_1, \ldots, t_k]^T \in \mathbb{N}^k$, $t_i \in \{1, \ldots, n\}$ for $i = 1, \ldots, k$.
- $\boldsymbol{x}(i) := [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]^T$ for any $1 \le i \le n$.
- $f(\boldsymbol{x}) := f(x_1, \ldots, x_n),\ d\boldsymbol{x} := dx_1 \ldots dx_n$.
- $\mathbf{0}, \mathbf{0}_{n\times 1}$ ...vector of n zero entries.
- $\boldsymbol{A}$, $\boldsymbol{A}_{m\times n} := [a_{ij}] = [A(i,j)]$ ...matrix of size $m \times n$.
- $\mathcal{R}(\boldsymbol{A}) := \{\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}$ ...range space of matrix operator $\boldsymbol{A}$.
- $\mathcal{N}(A) := \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = 0\}$ ...null space (kernel) of matrix operator $\boldsymbol{A}$.
- $\boldsymbol{A}^T := [a_{ji}]$ ...matrix transpose.
- $\boldsymbol{A}^* := [\bar{a}_{ji}]$ ...matrix adjoint.
- $\boldsymbol{I}$, $\boldsymbol{I}_n := \boldsymbol{I}_{n\times n} = [\delta_{ij}]$ ...identity matrix of order n.
- $\det \boldsymbol{A}$ ...determinant of a square matrix $\boldsymbol{A}$.
- $\mathbf{0}$, $\mathbf{0}_{m\times n}$ ...zero matrix of size $m \times n$.
- $\operatorname{diag}(\boldsymbol{x}) := \begin{bmatrix} x_1 & 0 & \ldots & 0 \\ 0 & x_2 & \ldots & 0 \\ & & \vdots & \\ 0 & 0 & \ldots & x_n \end{bmatrix}$ ...diagonal matrix.
- $\boldsymbol{A}(i,:) := [a_{i1}, \ldots, a_{in}]$ ...i-th row of matrix $\boldsymbol{A}$ using MATLAB style.
- $\boldsymbol{A}(:,j) := [a_{1j}, \ldots, a_{mj}]^T$ ...j-th column of matrix $\boldsymbol{A}$ using MATLAB style.

- $\boldsymbol{A} := [r_1; \ldots; r_m] = [s_1, \ldots, s_n]$ ...forming matrix $\boldsymbol{A}$ row-by-row or columnwise using MATLAB style.
- $\boldsymbol{A} > 0$ (or $\boldsymbol{A} \geq 0$) ...positively (semi)definite (nonnegatively definite) matrix.
- $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{n} x_i \bar{y}_i = \boldsymbol{y}^* \boldsymbol{x}$ ...scalar (inner) product of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.
- $\|\boldsymbol{x}\| := \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ ...Euclidean norm of vector $\boldsymbol{x}$.

Random variables and random vectors:

- X ...random variable.
- $\mathbb{X} := [X_1, \ldots, X_n]^T$ ...(real) random vector, indexing conventions listed above for number vectors are adopted accordingly.
- $\mu := \mu_X := \mathrm{E}X$ ...expectation of random variable X.
- $\boldsymbol{\mu} := \boldsymbol{\mu}_{\mathbb{X}} := \mathrm{E}\mathbb{X} := [\mathrm{E}X_1, \ldots, \mathrm{E}X_n]^T$ ...expectation of random vector $\mathbb{X}$.
- $\sigma^2 := \sigma_X^2 := \mathrm{var}X := \mathrm{E}|X - \mathrm{E}X|^2 = \mathrm{E}|X|^2 - |\mathrm{E}X|^2 \geq 0$ ...variance of random variable X.
- $\sigma_{XY} := \mathrm{cov}(X, Y) := \mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y) = \mathrm{E}XY - (\mathrm{E}X)(\mathrm{E}Y)$ ...covariance of random variables X and Y.
- $\Sigma_{\mathbb{X}} := \mathrm{var}\mathbb{X} := [\mathrm{cov}(X_i, X_j)] = \mathrm{E}(\mathbb{X} - \mathrm{E}\mathbb{X})(\mathbb{X} - \mathrm{E}\mathbb{X})^T = \mathrm{E}\mathbb{X}\mathbb{X}^T - (\mathrm{E}\mathbb{X})(\mathrm{E}\mathbb{X})^T$ ...variance matrix of random vector $\mathbb{X}$.
- $\Sigma_{\mathbb{X}\mathbb{Y}} := \mathrm{cov}(\mathbb{X}, \mathbb{Y}) := [\mathrm{cov}(X_i, Y_j)] = \mathrm{E}(\mathbb{X} - \mathrm{E}\mathbb{X})(\mathbb{Y} - \mathrm{E}\mathbb{Y})^T = \mathrm{E}\mathbb{X}\mathbb{Y}^T - (\mathrm{E}\mathbb{X})(\mathrm{E}\mathbb{Y})^T$ ...covariance matrix of $\mathbb{X}$ and $\mathbb{Y}$.

It holds:

- $\mathrm{var}X = \mathrm{cov}(X, X)$.
- $\mathrm{cov}(Y, X) = \mathrm{cov}(X, Y)$.
- $\mathrm{cov}(\sum_r X_r, \sum_s Y_s) = \sum_r \sum_s \mathrm{cov}(X_r, Y_s)$ and hence in particular:

- $\mathrm{var}(X+Y) = \mathrm{var}X + \mathrm{cov}(X,Y) + \mathrm{cov}(Y,X) + \mathrm{var}Y = \mathrm{var}X + 2\mathrm{cov}(X,Y) + \mathrm{var}Y$.
- $\mathrm{cov}(\mathbb{X},\mathbb{X}) = \mathrm{var}\mathbb{X}$.
- $\mathrm{cov}(\mathbb{Y},\mathbb{X}) = \mathrm{cov}(\mathbb{X},\mathbb{Y})^T$ implies:
- $\mathrm{var}\mathbb{X} = (\mathrm{var}\mathbb{X})^T$ ... **variance matrix $\mathbb{X}$ is symmetrical**.
- Given number vectors $\boldsymbol{a}$ and $\boldsymbol{c}$, and matrices B and D of compatible sizes then
  $\mathrm{cov}(\boldsymbol{a}+B\mathbb{X},\boldsymbol{c}+D\mathbb{Y}) = \mathrm{cov}(B\mathbb{X},D\mathbb{Y}) = B\,\mathrm{cov}(\mathbb{X},\mathbb{Y})\,D^T$

  $$\Downarrow \mathbb{X} = \mathbb{Y}$$
- $\mathrm{var}(\boldsymbol{a}+B\mathbb{X}) = \mathrm{cov}(\boldsymbol{a}+B\mathbb{X},\boldsymbol{a}+B\mathbb{X}) = \mathrm{cov}(B\mathbb{X},B\mathbb{X}) = B\,\mathrm{var}(\mathbb{X})\,B^T$

  $$\Downarrow \boldsymbol{a} = \boldsymbol{0},\ B = \boldsymbol{b}^T$$
- $0 \le \mathrm{var}(\boldsymbol{b}^T\mathbb{X}) = \boldsymbol{b}^T \mathrm{var}\mathbb{X}\,\boldsymbol{b}$ implies:
- $\mathrm{var}\mathbb{X} \ge 0$ ... **variance matrix is non-negatively positive** and consequently it has non-negative eigen values λ_i and its square root matrix $\Sigma_{\mathbb{X}}^{\frac{1}{2}}$ having eigen values $\lambda_i^{\frac{1}{2}}$ may be constructed such that:
- $\Sigma_{\mathbb{X}} = \Sigma_{\mathbb{X}}^{\frac{1}{2}}\,\Sigma_{\mathbb{X}}^{\frac{1}{2}}$.
- $\mathrm{cov}(\sum_r \mathbb{X}_r, \sum_s \mathbb{Y}_s) = \sum_r \sum_s \mathrm{cov}(\mathbb{X}_r, \mathbb{Y}_s)$ and hence in particular:
- $\mathrm{var}(\mathbb{X}+\mathbb{Y}) = \mathrm{var}\mathbb{X} + \mathrm{cov}(\mathbb{X},\mathbb{Y}) + \mathrm{cov}(\mathbb{Y},\mathbb{X}) + \mathrm{var}\mathbb{Y} = \mathrm{var}\mathbb{X} + 2\mathrm{cov}(\mathbb{X},\mathbb{Y}) + \mathrm{var}\mathbb{Y}$.

2. Introduction

Definition 2.1. Random (stochastic) process X is a nonempty family ($T \neq \emptyset$) of (real) random variables defined on the same probability space $(\Omega, \mathcal{A}, P)$. We write $X := \{X_t \mid t \in T\}$ or simply $\{X_t\}$.
Special cases:
$T \subseteq \mathbb{R}$ ... **continuous-time process** or **random function**.
$T \subseteq \mathbb{Z}$ ... **discrete-time process**, **random sequence** or **time series**.

Remark 2.2.

- Indexing set T is usually ordered and interpreted as continuous or discrete time interval. It may be also disordered, for example coordinates of points on the plane (meteorology) or in 3-D space (geophysics).
- As $X_t : \Omega \to \mathbb{R}$ is a (measurable) mapping for each $t \in T$, the stochastic process may be viewed as a mapping $X : \Omega \times T \to \mathbb{R}$ as well.

Definition 2.3. For fixed $\omega \in \Omega$ we get function $x : T \to \mathbb{R}$ as an outcome of a random experiment: $x(t) := X_t(\omega)$. This function is called **sample-path (trajectory, realization, observation) of** X.

Remark 2.4. Figures 1.1–1.6 illustrate trajectories of various random processes (time series). Later on formulation *stochastic process* is related to the general case with any T in contrast with the formulation *time series* which assumes $T = \mathbb{Z}$ or sometimes $T = \mathbb{N}$.

Definition 2.5. If $X = \{X_t\}$ is a time series where X_t, $t \in T$, are all **mutually independent and identically distributed** with mean μ and variance σ^2, we shall write

$$\boxed{X \sim IID(\mu, \sigma^2)}$$

Example 2.6 (Examples of time series)**.**

(1) Sinusoid with random amplitude and phase (Fig. 1.1).
(2) Binary process of tossing a coin (cf. Fig. 1.4 as well).
(3) Random walk.
(4) Branching process.

Definition 2.7 (Consistent system of distribution functions of X)**.** Let us denote $\mathcal{T} := \{\boldsymbol{t} \,|\, \boldsymbol{t} = [t_1, t_2, \ldots, t_n] \in T^n,\ t_i \neq t_j,$ for $i \neq j,$ $n \in \mathbb{N}\}$. For each $\boldsymbol{t} \in \mathcal{T}$ of any size $n \in \mathbb{N}$ let $F_{\boldsymbol{t}}(\boldsymbol{x})$ be the joint distribution function of the marginal random vector $\mathbb{X}_{\boldsymbol{t}}$ being selected from the stochastic process $X = \{X_t\}_{t \in T}$ at time instants $t_1, t_2, \ldots, t_n$. The system $\{F_{\boldsymbol{t}}\}_{\boldsymbol{t} \in \mathcal{T}}$ describes completely the stochastic behaviour of X and is called **consistent system of distribution functions of** X (cf. the next theorem).

Theorem 2.8. *The system $\{F_{\boldsymbol{t}}\}_{\boldsymbol{t} \in \mathcal{T}}$ of definition 2.7 is called consistent because the following two consistency conditions hold for each $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$ and $n \in \mathbb{N}$:*

(i) *$F_{\boldsymbol{t_p}}(\boldsymbol{x_p}) = F_{\boldsymbol{t}}(\boldsymbol{x})$ for any permutation $\boldsymbol{p}$ of indices $\{1, 2, \ldots, n\}$.*
(ii) *$\lim_{x_i \to \infty} F_{\boldsymbol{t}}(\boldsymbol{x}) = F_{\boldsymbol{t}}(x_1, \ldots, x_{i-1,\infty}, x_{i+1}, \ldots, x_n) =:$
$=: F_{\boldsymbol{t}(i)}(\boldsymbol{x}(i))$ for any $i \in \{1, 2, \ldots, n\}$.*

Theorem 2.9 (Kolmogorov's theorem)**.** *Given T and $\mathcal{T}$ as of definition 2.7, let $\mathcal{F} := \{F_{\boldsymbol{t}}\}_{\boldsymbol{t} \in \mathcal{T}}$ be a consistent system of distribution functions. Then there exists a stochastic process $\{X_t\}_{t \in T}$ defined on a suitable probability space $(\Omega, \mathcal{A}, P)$ such that $\mathcal{F}$ is its system of distribution functions.*

Remark 2.10. Conditions (i) and (ii) of theorem 2.8 can be replaced by equivalent conditions formulated in terms of characteristic functions $\Phi_{\boldsymbol{t}}(\boldsymbol{u}) = \mathrm{E}(\exp(i\,\boldsymbol{u}^T \mathbb{X}_{\boldsymbol{t}})) = \mathrm{E}(\exp(i \sum_{j=1}^{n} u_j X_{t_j}))$, $\boldsymbol{u} \in \mathbb{R}^n$, which are associated with the distribution functions $F_{\boldsymbol{t}}$:

(i') $\Phi_{\boldsymbol{t_p}}(\boldsymbol{u_p}) = \Phi_{\boldsymbol{t}}(\boldsymbol{u})$ for any permutation $\boldsymbol{p}$ of indices $\{1, 2, \ldots, n\}$.
(ii') $\lim_{u_i \to 0} \Phi_{\boldsymbol{t}}(\boldsymbol{u}) = \Phi_{\boldsymbol{t}}(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_n) =:$
$=: \Phi_{\boldsymbol{t}(i)}(\boldsymbol{u}(i))$ for any $i \in \{1, 2, \ldots, n\}$.

Definition 2.11. We call a stochastic process **normal** or **gaussian** if every distribution function F_t of its consistent system $(\boldsymbol{t} \in \mathfrak{T})$ is a joint distribution function of normally distributed marginal random vector $\mathbb{X}_{\boldsymbol{t}}$.

Now we are about to introduce **moment functions** as analogies to the expectation and variance matrix of a random vector, which may be considered as a special case of a stochastic process with finite index set $T = \{1, 2, \ldots, n\}$.

Definition 2.12. Given stochastic processes $X = \{X_t\}_{t \in T}$ and $Y = \{Y_t\}_{t \in T}$, both on the same probability space, we define 1-st and 2-nd moment functions as follows.

(1) **mean of** X: $\mu_X : T \to \mathbb{R}$ by $\mu_X(t) := \mathrm{E}X_t$ provided that the expectations exist for all $t \in T$.
(2) **autocovariance function of** X: $\gamma_X : T \times T \to \mathbb{R}$ by $\gamma_X(r, s) := \mathrm{cov}(X_r, X_s)$ provided that the covariances exist for all $r, s \in T$.
(3) **variance of** X: $\sigma_X^2 : T \to \mathbb{R}^+$ by $\sigma_X^2(t) := \mathrm{cov}(X_t, X_t) = \gamma_X(t, t)$ provided the variances exist for all $t \in T$.
(4) **autocorrelation function of** X: $\rho_X : T \times T \to [-1, 1]$ by
$$\rho_X(r, s) := \begin{cases} \frac{\gamma_X(r,s)}{\sqrt{\gamma_X(r,r)}\sqrt{\gamma_X(s,s)}} & \text{for } \sqrt{\gamma_X(r,r)}\sqrt{\gamma_X(s,s)} \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$
provided that the correlations exist for all $r, s \in T$.
(5) **cross-covariance function of** X **and** Y:
$\gamma_{XY} : T \times T \to \mathbb{R}$ by $\gamma_{XY}(r, s) := \mathrm{cov}(X_r, Y_s)$ provided that the covariances exist for all $r, s \in T$.
(6) **cross-correlation function of** X **and** Y:
$\rho_{XY} : T \times T \to [-1, 1]$ by
$$\rho_{XY}(r, s) := \begin{cases} \frac{\gamma_{XY}(r,s)}{\sqrt{\gamma_X(r,r)}\sqrt{\gamma_Y(s,s)}} & \text{for } \sqrt{\gamma_X(r,r)}\sqrt{\gamma_Y(s,s)} \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$
provided that the correlations exist for all $r, s \in T$.

Theorem 2.13. *Given stochastic process $X := \{X_t \,|\, t \in T\}$ such that $\mathrm{E}|X_t|^2 < \infty$ for all $t \in T$, then $\mu_X(\cdot)$, $\gamma_x(\cdot,\cdot)$ and $\rho_X(\cdot,\cdot)$ exist as well. In such a case we say that X* ***has finite second moments****.*

Definition 2.14. Time series $X := \{X_t \,|\, t \in \mathbb{Z}\}$ is called **strictly stationary** if each distribution function from its consistent system $\{F_t\}_{t \in \mathfrak{T}}$, is shift-invariant (time-invariant): $F_t(\cdot) \equiv F_{t+h}(\cdot)$ for each $t \in \mathfrak{T}$ and $h \in \mathbb{Z}$.

Definition 2.15. Time series $X := \{X_t \,|\, t \in \mathbb{Z}\}$ is called **(weakly) stationary** if the following three conditions are fulfilled:

(1) X has finite second moments.
(2) $\gamma_X(r,s) = \gamma_X(r+h, s+h)$ for each $r, s, h \in \mathbb{Z}$.
(3) $\mu_X(\cdot) \equiv \mu_X$ is a constant function.

If only the first two are valid then X is called **covariance stationary**.

<u>*Remark*</u> 2.16.

(1) Clearly (2) implies with $r = s$ that the variance function of a stationary time series is a constant function as well: $\sigma_x^2(\cdot) \equiv \sigma_X^2$.
(2) If (3) holds, then $\gamma_X(r,s) = \mathrm{E}X_rX_s - (\mathrm{E}X_r)(\mathrm{E}X_s) = \mathrm{E}X_rX_s - \mu_x^2$ implies that (2) is equivalent (and might be thus substituted) with the condition: $\mathrm{E}X_rX_s = \mathrm{E}X_{r+h}X_{s+h}$ for each $r, s, h \in \mathbb{Z}$. We see altogether that all first and second moments are shift-invariant with weak stationarity. That is why weak stationarity is sometimes denoted as **2-nd order stationarity**.

<u>*Remark*</u> 2.17. Clearly condition (2) of definition 2.15 may be substituted be a modified condition
(2') $\gamma_x(r,s)$ depends only on the difference of arguments $r - s$.

That is why we can introduce autocovariance and autocorrelation function of a stationary time series as a function of one argument only:

$$\begin{aligned}\gamma_X(h) &:= \gamma_X(t+h,t)\\ \rho_X(h) &:= \rho_X(t+h,t) = \frac{\gamma_X(t+h,t)}{\sigma_X\sigma_X} = \frac{\gamma_X(h)}{\gamma_X(0)}\\ \sigma_X^2 &= \gamma_X(t,t) = \gamma_X(0)\end{aligned} \tag{2.1}$$

where $t,h \in \mathbb{Z}$ are arbitrary.

Theorem 2.18. *Every strictly stationary time series with finite second moments is stationary.*

Example 2.19.
In general stationarity does not imply strict stationarity (counter-example)

Theorem 2.20. *Every stationary gaussian time series is strictly stationary.*

Definition 2.21. Time series $X = \{X_t\}$ is called **white noise** with mean μ and variance σ^2, if $\mu_X(t) \equiv \mu$ and $\gamma_X(r,s) = \begin{cases}\sigma^2 & \text{for } r=s\\ 0 & \text{otherwise}\end{cases}$.
We write

$$\boxed{X \sim WN(\mu,\sigma^2).}$$

Stationary time series which is not white noise, is sometimes called **coloured noise**.

Remark 2.22. It is straightforward to verify the following implications:

$$X \sim IID(\mu,\sigma^2) \quad\Rightarrow\quad X \sim WN(\mu,\sigma^2) \quad\Rightarrow\quad X \text{ is stationary}.$$

Observe that neither of inverse implications holds in general (cf. example 2.19).

Example 2.23.

(1) Let $X_t(\omega) := A(\omega)\cos(\theta t) + B(\omega)\sin(\theta t)$, $t \in \mathbb{Z}$, $\theta \in [-\pi, \pi]$, $\operatorname{cov}(A, B) = 0$, $\mathrm{E}A = \mathrm{E}B = 0$, $\sigma_A^2 = \sigma_B^2 = 1$. Then $\{X_t\}$ is a stationary time series.

(2) Let $X_t := Z_t + \theta Z_{t-1}$, $\{Z_t\} \sim WN(0, \sigma^2)$, $t \in \mathbb{Z}$, $\theta \in \mathbb{R}$. Then $\{X_t\}$ is a stationary time series.

(3) Let $X_t := \begin{cases} Y_t & \text{for even } t \\ Y_t + 1 & \text{for odd } t \end{cases}$, $t \in \mathbb{Z}$, where $\{Y_t\}$ is a stationary time series. Then $\{X_t\}$ is a time series which is covariance stationary but not stationary.

(4) The random walk $\{S_t\}_{t\in\mathbb{Z}}$ from example 2.6(3) is neither stationary nor covariance stationary.

Remark 2.24 (Multivariate Time Series).
One can introduce the concept of m-**dimensional time series** ($m \in \mathbb{N}$) following the analogy with the univariate case ($m = 1$):
$\mathbb{X} := \{\mathbb{X}_t \,|\, t \in T\}$ where $\mathbb{X}_t = [X_{t,1}, \ldots, X_{t,m}]^T$ are m-dimensional random vectors on the same probability space $(\Omega, \mathcal{A}, P)$. We obtain univariate partial time series, vector mean function and matrix autocovariance/autocorrelation functions:
$X_i := \{X_{t,i} \,|\, t \in T\}$ ... i-**th partial time series**.
$\mu_{\mathbb{X}}(t) := [\mu_1(t), \ldots, \mu_m(t)]^T$ where $\mu_i(t) := \mathrm{E}X_{t,i} = \mu_{X_i}(t)$.
$\gamma_{\mathbb{X}}(r, s) := \Sigma_{\mathbb{X}}(r, s) := [\operatorname{cov}(X_{r,i}, X_{s,j})]_{i,j} = [\gamma_{ij}(r, s)]_{i,j}$ where
$\gamma_{ij}(r, s) := \gamma_{X_i X_j}(r, s)$ is clearly just the cross-covariance function of partial time series X_i and X_j.
$\rho_{\mathbb{X}}(r, s) := [\rho(X_{r,i}, X_{s,j})]_{i,j} = [\rho_{ij}(r, s)]_{i,j}$ where
$\rho_{ij}(r, s) := \rho_{X_i X_j}(r, s)$ is cross-correlation function of partial time series X_i and X_j.

The m-**dimensional stationarity** is to be established quite in analogy to definition 2.15 (see also remarks 2.16 and 2.17) simply assuming finite second moments for all partial time series in (1) and substituting $\gamma_{\mathbb{X}}$ for γ_X in (2) and $\mu_{\mathbb{X}}$ for μ_X in (3).

It is an easy exercise to prove the following statement:
$\mathbb{X}$ is stationary iff the following two conditions are fulfilled:

(a) Each partial time series X_i $(i = 1, \ldots, m)$ is stationary.
(b) $\gamma_{\mathbb{X}}(r, s) = \gamma_{\mathbb{X}}(r + h, s + h)$ for each $r, s, h \in \mathbb{Z}$.

Clearly the following relationships hold:

$$\rho_{\mathbb{X}}(r, s) = \left[\frac{\gamma_{ij}(r, s)}{\sqrt{\gamma_i(r, r)}\sqrt{\gamma_j(s, s)}} \right] \quad \text{in the general case}$$

and

$$\rho_{\mathbb{X}}(h) = \left[\frac{\gamma_{ij}(h)}{\sqrt{\gamma_i(0)}\sqrt{\gamma_j(0)}} \right] \quad \text{in the stationary case.}$$

where $\gamma_i := \gamma_{ii}$ is autocovariance function of i-th partial time series.

Definition 2.25. A bivariate function $f : T \times T \to \mathbb{R}$, $T \neq \emptyset$, is said to be **symmetric** or **non-negatively definite** if each square matrix $[f(t_i, t_j)]_{ij}$ of any size $n \in \mathbb{N}$ has the respective property for any choice of $\boldsymbol{t} := [t_1, \ldots, t_n] \in T^n$, i.e. all such matrices are symmetric or non-negatively definite.
A univariate function $g : \mathbb{Z} \to \mathbb{R}$ is said to be **symmetric** or **non-negatively definite** if the bivariate function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ defined by $f(r, s) := g(r - s)$ is symmetric or non-negatively definite.

Lemma 2.26. *Let f and g be functions as of definition 2.25. Then f (or g) is symmetric iff $f(s, r) = f(r, s)$ holds for any $r, s \in T$ (or $g(-t) = g(t)$ holds for any $t \in \mathbb{Z}$).*

Lemma 2.27. *The sum of two symmetric (or non-negatively definite) functions, which are both bivariate or univariate and defined on the same domain, is a symmetric (or non-negatively definite) function as well.*

Theorem 2.28 (Autocovarinace and autocorrelation function properties).
Let $X := \{X_t \mid t \in T\}$ *be a stochastic process with the autocovariance function* $\gamma_X(\cdot,\cdot)$ *[autocorrelation function* $\rho_X(\cdot,\cdot)$ *]. Then the following holds:*

(1) $\gamma_X(t,t) \geq 0$
[$\rho_X(t,t) = 1$ *if* $\gamma_X(t,t) = \sigma_X^2(t) \neq 0$, *or* $= 0$ *otherwise]* *for all* $t \in T$.

(2) $|\gamma_X(r,s)| \leq \sqrt{\gamma_X(r,r)}\sqrt{\gamma_X(s,s)}$ *[* $|\rho_X(r,s)| \leq 1$ *] for all* $r, s \in T$.

(3) γ_X *[* ρ_X *] is a symmetric and non-negatively definite function.*

Corollary 2.29 (for stationary time series).
Let $X := \{X_t \mid t \in \mathbb{Z}\}$ *be a stationary time series with the autocovariance function* $\gamma_X(\cdot)$ *[autocorrelation function* $\rho_X(\cdot)$ *]. Then the following holds:*

(1') $\gamma_X(0) \geq 0$
[$\rho_X(0) = 1$ *if* $\gamma_X(0) = \sigma_X^2 \neq 0$, *or* $= 0$ *otherwise].*

(2') $|\gamma_X(h)| \leq \gamma_X(0)$ *[* $|\rho_X(h)| \leq 1$ *] for all* $h \in \mathbb{Z}$.

(3') γ_X *[* ρ_X *] is a symmetric and non-negatively definite function.*

Theorem 2.30. *Given a function* $\gamma(\cdot,\cdot) : T \times T \to \mathbb{R}$ *(or* $\gamma(\cdot) : \mathbb{Z} \to \mathbb{R}$*) which is symmetric and non-negatively definite, then there exists a gaussian stochastic process (or stationary gaussian time series)* X *having autocovariance function* $\gamma_X = \gamma$.

Corollary 2.31. *Properties (1) and (2) (or (1') and (2')) are direct consequence of the property (3) (or (3')).*

Corollary 2.32. *Given two stochastic processes (stationary time series)* X *and* Y *with autocovariance functions* γ_X *and* γ_Y, *then there exists a stochastic process (stationary time series)* Z, *even gaussian, such that* $\gamma_Z = \gamma_X + \gamma_Y$.

Theorem 2.33.

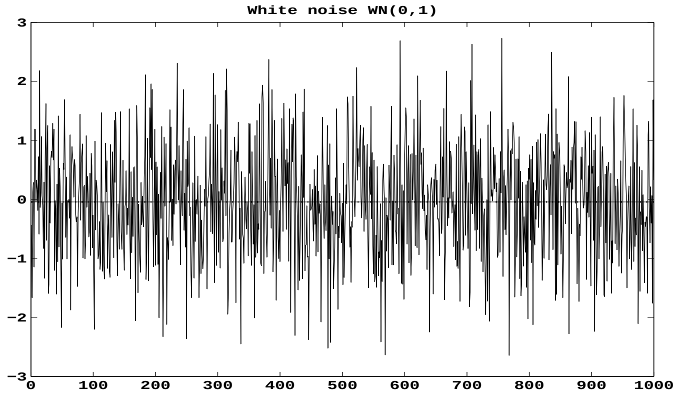
A function $\rho(h) := \begin{cases} 1 & \text{for } h = 0 \\ r & \text{for } h = \pm 1 \\ 0 & \text{for } |h| > 1 \end{cases}$ *can be an autocorrelation function of a suitable stationary time series* X *iff* $|r| \leq \frac{1}{2}$*. In such a case one possible choice is the time series* X *of example (2) in 2.23:* $X_t := Z_t + \theta Z_{t-1}$*,* $\{Z_t\} \sim WN(0, \sigma^2)$*,* $t \in \mathbb{Z}$*, with* $\theta = \frac{1 \pm \sqrt{1-4r^2}}{2r}$*.*

Definition 2.34 (Estimates of moment functions).
Let $\boldsymbol{x} = [x_1, \ldots, x_n]$ be n samples ($x_t = X_t(\omega)$ for $t = 1, \ldots, n$) of a stationary time series with mean μ, variance σ^2, autocovariance function $\gamma(\cdot)$ and autocorrelation function $\rho(\cdot)$. Their estimates are computed as follows:
$\widehat{\mu} := \frac{1}{n}\sum_{j=1}^{n} x_j$ ... estimate of μ;
$\widehat{\gamma}(h) := \frac{1}{n}\sum_{j=1}^{n-h}(x_{j+h} - \widehat{\mu})(x_j - \widehat{\mu})$, $0 \leq h \leq n-1$,
$\widehat{\gamma}(h) := \widehat{\gamma}(-h)$, $-(n-1) \leq h < 0$ (by symmetry 2.29(3') and 2.26);
... estimate of $\gamma(h)$;
$\widehat{\sigma}^2 := \widehat{\gamma}(0)$ ... estimate of the variance;
$\widehat{\rho}(h) := \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}$, $-(n-1) \leq h \leq n-1$ (see eq. (2.1))
... estimate of the autocorrelation function in the case of $\widehat{\gamma}(0) \neq 0$, otherwise $\widehat{\rho}(h) := 0$.

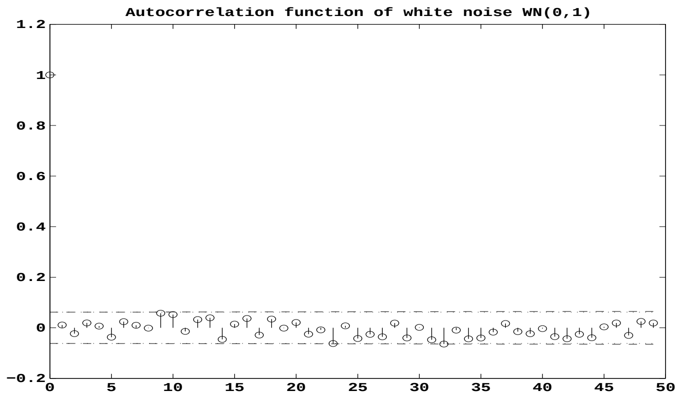
Theorem 2.35. *Let* $\mathbb{X} := [X_1, \ldots, X_n]$ *be the random subvector in* X *associated with sample vector* $\boldsymbol{x}$*. Then both the matrix*

$$\widehat{\Gamma}_n := [\widehat{\gamma}(i-j)]_{i,j} = \begin{bmatrix} \widehat{\gamma}(0) & \widehat{\gamma}(1) & \ldots & \widehat{\gamma}(n-1) \\ \widehat{\gamma}(1) & \widehat{\gamma}(0) & \ldots & \widehat{\gamma}(n-2) \\ \ldots & \ldots & \ldots & \ldots \\ \widehat{\gamma}(n-1) & \widehat{\gamma}(n-2) & \ldots & \widehat{\gamma}(0) \end{bmatrix}$$

which is an estimate of the variance matrix $\mathrm{var}\mathbb{X}$*, and the matrix* $\widehat{R}_n := \frac{\widehat{\Gamma}_n}{\widehat{\gamma}(0)}$ *which is an estimate of the correlation matrix* $\rho(\mathbb{X})$*, are symmetric and non-negatively definite.*



Gaussian white noise $WN(0,1)$



Autocorrelation function of gaussian white noise $WN(0,1)$

<u>*Remark*</u> 2.36.

(1) The estimate $\widehat{\gamma}(h)$ is not unbiased ($\mathrm{E}\widehat{\gamma}(h) \neq \gamma(h)$) because we divide the sample sum by n and not by the number of

degrees of freedom $n-1-h$. Let us observe that theorem 2.35 does not hold for the unbiased estimate (matrix $\widehat{\Gamma}_n$ looses the natural property of non-negative positiveness). Anyway, the estimate $\widehat{\gamma}(h)$ is **asymptotically unbiased** in the sense that $\mathrm{E}\widehat{\gamma}(h) \to \gamma(h)$ with $n \to \infty$. Morover it is **consistent in the quadratic mean** in the sense that $\mathrm{E}|\widehat{\gamma}(h) - \gamma(h)|^2 \to 0$ with $n \to \infty$, where the convergence is even faster than with the unbiased estimate.

(2) The estimate is reliable only for $n > 50$ and $h < \frac{n}{4}$.

(3) From the algebraic point of view $\widehat{\gamma}(h)$ may be written in the form of a dot product $\widehat{\gamma}(h) = \frac{1}{n}\langle \boldsymbol{x}_0, \boldsymbol{x}_h \rangle$ where $\boldsymbol{x}_h := [\underbrace{0,\dots,0}_{h}, x_1 - \widehat{\mu}, \dots, x_n - \widehat{\mu}, \underbrace{0,\dots,0}_{n-1-h}]$. Thus $\boldsymbol{x}_0$ represents the original sample vector (padded with $n-1$ zeros) and $\boldsymbol{x}_h$ its copy shifted by h.
Clearly $\|\boldsymbol{x}_0\|^2 = \|\boldsymbol{x}_h\|^2 = \sum_{j=1}^{n} |x_j - \widehat{\mu}|^2$. From the Schwarz inequality we have $|\langle \boldsymbol{x}_0, \boldsymbol{x}_h \rangle| \le \|\boldsymbol{x}_0\|^2$ resulting in $|\widehat{\gamma}(h)| \le \frac{1}{n}\|\boldsymbol{x}_0\|^2 = \frac{1}{n}\langle \boldsymbol{x}_0, \boldsymbol{x}_0 \rangle = \widehat{\gamma}(0)$. Hence we see that the estimate of the autocorrelation function preserves its natural property $|\widehat{\rho}(h)| \le 1$.

(4) In view of (3) the estimate $\widehat{\rho}(h)$ may be interpreted geometrically as a cosine of the angle between the original and shifted copy of $\boldsymbol{x}_0$ which is a measure of their linear dependence (similarity): zero means ortogonality (full linear independence=no correlations between them), ± 1 means linear dependence (full correlation: one of them is obtained as scalar multiple of the latter).

(5) Trend is indicated by correlations at great lags implying small decay of $\gamma(h)$ with $h \to \infty$. Periodic component is reflected by oscillatory behaviour of $\widehat{\gamma}(h)$ with the basic period of that component, or mixture of them if there is more than one such periodic component.