An Imprecise Introduction to Stochastic Processes

Stelios Arvanitis

2018

Acknowledgement. I am grateful to Anastasia Chatzilena that helped with the typesetting of the present notes.

Part I

1 General Definitions

Remember that a random variable is a real valued function between measurable spaces, which *preserves* the relevant measurability structures. Consider a probability measure, i.e. a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a non-empty carrier set, $\mathcal{F}$ is a set of subsets of Ω to which probabilities can be consistently attributed to (σ-algebra) and $\mathbb{P}$ is a probability measure (or distribution), $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ that satisfies positivity, normalization and countable additivity. Let $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be the set of reals with its (without loss of generality) Borel algebra. If $x : \Omega \to \mathbb{R}$ is a random variable, then x and $\mathbb{P}$ define a unique probability measure on $\mathbb{R}$ by transfer (push out measure). This is also termed as the marginal distribution of x, or the distribution that x follows.

Analogous is the definition of a random $n-$vector as a measurable function $\mathbf{x}_n : \Omega \to \mathbb{R}^n$. Again $\mathbf{x}_n$ and $\mathbb{P}$ define a unique probability measure on $\mathbb{R}^n$ by transfer (push out measure). Simultaneously $\mathbf{x}_n = (x_1, \ldots, x_n)$, i.e. a vector of random variables. In order for this couple of definitions to be consistent some consistency conditions must hold between the transfer measure on $\mathbb{R}^n$ with the marginal distributions of the random variables that comprise the vector. We can generalize the construction to obtain the notion of a stochastic process.

Definition 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, Θ a non empty set and $(\mathbb{R}, B_{\mathbb{R}})$ the measurable space consisting of $\mathbb{R}$ with some σ-algebra (usually the Borel). An $\mathbb{R}$-valued stochastic process over Θ (say x) is a Θ-collection of random variables on Ω, i.e.

$$x = \{x_\theta : \Omega \to \mathbb{R}, x_\theta \text{ measurable}, \theta \in \Theta\},$$

such that a probability measure is consistently defined via $\mathbb{P}$ and x on the set

$$\mathbb{R}^\Theta = \{f : \Theta \to \mathbb{R}\},$$

with respect to some appropriately constructed σ-algebra on $\mathbb{R}^{\Theta}$ via the consideration of $B_{\mathbb{R}}$ and Θ.

Remark 1. The definition can be extended so that the stochastic process to be a collection of random elements assuming their values in a more general, with appropriate properties space S (in the particular definition $S = \mathbb{R}$). In this respect, such a process is a random element defined on Ω with values in $S^{\Theta} \triangleq \{f : \Theta \to S\}$, equipped with an appropriate σ-algebra. In some cases it is desirable to define and study processes with values in some subspace of S^{Θ}(e.g. when S and Θ are also equipped with the necessary mathematical structure, the subspace of continuous functions from Θ to S, etc). Finally notice that the extension to a general S is almost straightforward when $S = \mathbb{R}^{\kappa}$, $\kappa \geqslant 1$, yet it can be more involved for more general S.

Remark 2. The definition implies that $\forall \omega \in \Omega$, $x(\omega) = f_{\omega} : \Theta \to \mathbb{R}$. Any such f_{ω} is termed as *a sample path* of the process.

Remark 3. The definition is not at all precise, since several notions from function spaces and measure theory would be needed. However it reveals the fact that x is "simply" an appropriately measurable function $\Omega \to \mathbb{R}^{\Theta}$, i.e. a construct that on each $\omega \in \Omega$ it assigns $x(\omega) : \Theta \to \mathbb{R}$ in a "measurable" manner. Hence, along with $\mathbb{P}$ it defines a probability measure on $\mathbb{R}^{\Theta}$, i.e. a distribution on a function space which is a quite complex mathematical entity. Part of the theory of stochastic processes revolves around the issue of representation of such a distribution by more familiar concepts. In any case a "classical" theorem, the Kolmogorov's Consistency Theorem (also known as *(Daniell)-Kolmogorov Extension Theorem*) implies that under some weak consistency conditions (involving permutations and integration) the distribution on $\mathbb{R}^{\Theta}$ is completely characterized by the set of finite dimensional distributions (fidi(s)).

Definition 2. The set of fidis of x is

$$\left\{\mathbb{P}_{x_{\theta_1}, x_{\theta_2}, \ldots, x_{\theta_\kappa}}, \theta_1, \theta_2, \ldots, \theta_\kappa \in \Theta^{\star}, \Theta^{\star} \neq \varnothing, \Theta^{\star} \text{ finite and ordered}\right\},$$

where

$$\mathbb{P}_{x_{\theta_1}, x_{\theta_2}, \ldots, x_{\theta_\kappa}} \triangleq \mathbb{P}\left(x_{\theta_1} \in \cdot, x_{\theta_2} \in \cdot, \ldots, x_{\theta_\kappa} \in \cdot\right),$$

i.e. it is the joint distribution of the random vector $\left(x_{\theta_1}, x_{\theta_2}, \ldots, x_{\theta_\kappa}\right)$.

Hence, iff the aforementioned conditions of the Consistency Theorem are satisfied, then the set of the joint distributions of the random vectors that can be constructed from the elements of x, completely specifies the distribution on $\mathbb{R}^{\Theta}$. [1] Hence the theorem represents a probability measure on a function space by a set of multivariate distributions, hence it reduces the former to the more familiar latter.

Example 1. The criteria we have encountered in the theory of M-estimation (likelihood functions, GMM criteria, sum of squares, etc) are typically stochastic processes defined over the relevant parameter spaces under the appropriate conditions.

[1] iff abbreviates if and only if.

Example 2. When $\Theta = \{1\}$ we recover the notion of the random variable, and when $\Theta = \{1, 2, \dots, n\}$ (as an ordered set) the notion of the $n-$random vector. (Describe the set of fidi's in the latter case!).

The following provides with a category of processes with familiar properties.

Definition 3. A *Gaussian* $\mathbb{R}$-valued process over Θ is a stochastic process for which every fidi is a normal distribution. It is possible to prove that a Gaussian process is completely characterized by the the following two functions: (i) the mean function $\mu : \Theta \to \mathbb{R}$ where $\mu(\theta) \triangleq \mathbb{E}(x_\theta)$ (ii) the covariance function (covariance kernel) $\Gamma : \Theta \times \Theta \to \mathbb{R}$ where $\Gamma(\theta, \theta^\star) \triangleq \mathbb{E}(x_\theta x_{\theta^\star}) - \mu(\theta)\mu(\theta^\star)$.

Example 3. $\Theta = \mathbb{Z}$, $z \sim N(0,1)$, $x_\theta = \theta z$ can be proven to be a Gaussian process with $\mu(\theta) = \mathbb{E}(x_\theta) = \theta\mathbb{E}(z) = 0\ \forall\theta \in \mathbb{R}$ and $\Gamma(\theta, \theta^\star) = \mathbb{E}(x_\theta x_{\theta^\star}) = \theta\theta^\star E(z^2) = \theta\theta^\star$.[2]

Definition 4. An $\mathbb{R}$-valued *time series* is a stochastic process where Θ is a totally ordered set, hence represents time.

Remark 4. Θ is usually a subset of $\mathbb{R}$ with the usual order. When Θ is an interval, then the time series is said to evolve in *continuous time* (the aforementioned Gaussian process is an example of a continuous time, time series). When Θ is discrete, the evolution is in *discrete time*. When $\Theta = \mathbb{N}$ or $\mathbb{Z}$ (or just countable), then x is also called a (possibly double)[3] random sequence. In what follows, mostly $\Theta = \mathbb{Z}$ (or $\mathbb{N}$ with slight modifications) and a typical element of Θ will be denoted as t with the obvious connotation.

In this context $x = (x_t)_{t\in\mathbb{Z}} \triangleq (\dots, x_{-1}, x_0, x_1, \dots)$, where x_t is necessarily a random variable for each $t \in \mathbb{Z}$.

Example 4. x is i.i.d.

Example 5. $x_t \sim N(t, t^2)$ and x independent.

2 Properties: (Strict) Stationarity

Given a process $x = (x_t)_{t\in\mathbb{Z}}$, consider the finite subset of $\mathbb{Z}$, $\{t_1, t_2, \dots, t_\kappa\}$ where $\kappa \in \mathbb{N}$. This defines the fidi $\mathbb{P}_{x_{t_1}, x_{t_2}, \dots, x_{t_\kappa}}$. Given that $\Theta = \mathbb{Z}$, we can shift time just by adding a constant to each element, obtaining again a subset of Θ. Suppose that $m \in \mathbb{Z}$ and consider the set $\{t_1 + m, t_2 + m, \dots, t_\kappa + m\}$, call this as the m-time translation of $\{t_1, t_2, \dots, t_\kappa\}$ and the fidi that corresponds to the m-time translation is $\mathbb{P}_{x_{t_1+m}, x_{t_2+m}, \dots, x_{t_\kappa+m}}$. For a general stochastic process this two fidis won't coincide.

Definition 5. If for all $m \in \mathbb{Z}$, $\mathbb{P}_{x_{t_1}, x_{t_2}, \dots, x_{t_\kappa}} = \mathbb{P}_{x_{t_1+m}, x_{t_2+m}, \dots, x_{t_\kappa+m}}$ then we say that the particular fidi remains invariant with respect to time shifts.

[2] For another simple example see https://eclass.aueb.gr/modules/document/file.php/OIK230/ExGauss.pdf
[3] This holds when Θ does not have an initial element.

Definition 6. x is (strictly) stationary iff every fidi remains invariant w.r.t. time shifts.

Hence, strict stationarity is essentially invariance of the fidis w.r.t. time shifts, and it constitutes of a strong collection of restrictions on the behavior of the process. We immediately obtain the following *necessary* condition for stationarity.

Corollary 1. *If x is (strictly) stationary then $\mathbb{P}_{x_t} = \mathbb{P}_{x_{t^\star}}$ for all $t, t^\star \in \mathbb{Z}$, i.e. the marginal coincide.*

Proof. Consider $\{t\}$ as a singleton subset of $\mathbb{Z}$ and apply the definition. □

Example 6. The process in example 4 is stationary since

$$\mathbb{P}_{x_{t_1},x_{t_2},\ldots,x_{t_\kappa}} \overset{\text{ind.}}{=} \mathbb{P}_{x_{t_1}} \cdot \ldots \cdot \mathbb{P}_{x_{t_\kappa}} \overset{\text{hom.}}{=} \mathbb{P}_{x_{t_1+m}} \cdot \ldots \cdot \mathbb{P}_{x_{t_\kappa+m}} \overset{\text{ind.}}{=} \mathbb{P}_{x_{t_1+m},x_{t_2+m},\ldots,x_{t_\kappa+m}},$$

for κ, m arbitrary.

Example 7. The process in example 5 is not stationary since it violates Corollary (1) (explain!).

3 Properties: Weak Stationarity

Definition 7. x is weakly stationary (or covariance stationary, or second order stationarity) iff it satisfies the following conditions:

a. $\mathbb{E}\left(x_t\right) \in \mathbb{R}$ and is independent of t

b. $\mathbb{V}ar\left(x_t\right) < +\infty$ and is independent of t, and

c. $Cov\left(x_t, x_{t-\kappa}\right)$ is independent of t (but may depend on κ), $\forall \kappa \in \mathbb{Z}$.

Notice that the notion involves moment existence conditions, while the covariances in the third condition exist due to condition 2 and the Cauchy-Schwarz inequality. A primal example of a weakly stationary process is that of a white noise process.

Example 8. (White noise process-WN (σ^2)). Let $\sigma^2 > 0$. Then $\varepsilon = \left(\varepsilon_t\right)_{t\in\mathbb{Z}} \sim$ WN(σ^2) iff $\mathbb{E}\left(\varepsilon_t\right) = 0$, $\forall t$, $\mathbb{V}ar\left(\varepsilon_t\right) = \sigma^2$, $\forall t$ and $Cov\left(\varepsilon_t, \varepsilon_{t-\kappa}\right) = 0$, $\forall t, \kappa$. White noise processes exist (See example 4 along with $\varepsilon_t \sim N\left(0, \sigma^2\right)$). They are obviously weakly stationary (why?).

Notice that conditions 2, 3 in the definition of weak stationarity can be equivalently stated by restricting $\kappa \in \mathbb{N}$ (why?). Using this remark, we readily obtain the following definitions.

Definition 8. Suppose that x is weakly stationary. Then the function $\gamma_\kappa : \mathbb{N} \to \mathbb{R}$ with $\gamma_\kappa \triangleq Cov\left(x_t, x_{t-\kappa}\right)$ is called autocovariance function of x. Analogously $\rho_\kappa \triangleq \frac{\gamma_\kappa}{\gamma_0} = \frac{Cov(x_t,x_{t-\kappa})}{Var(x_t)}$ is called autocorrelation function of x.

Those codify the covariance structure (and hence part of the dependence structure) of the elements of the process. If we restrict this structure to asymptotic uncorrelateness then we obtain the following chain of definitions.

Definition 9. Suppose that x is weakly stationary. If $\gamma_\kappa \to 0$ as $\kappa \to \infty$ then x is called *regular*.

Definition 10. Suppose that x is weakly stationary. $\sum\limits_{\kappa=0}^{\infty} |\gamma_\kappa| < +\infty$, i.e. the autocovariance function is absolutely summable, then x is called *short memory*.

Remark 5. Regularity implies asymptotic uncorrelateness. Short memory implies that the elements of the process become uncorrelated asymptotically at an appropriately fast rate. This is due to that $\sum\limits_{\kappa=0}^{\infty} |\gamma_\kappa| < +\infty$ implies that $\gamma_\kappa \to 0$ as $\kappa \to \infty$, thus short memory implies regularity. However, the converse does not hold (e.g. suppose that $\gamma_\kappa = \frac{1}{\kappa+1}$, hence $\gamma_\kappa \to 0$ as $\kappa \to \infty$, yet $\sum\limits_{\kappa=0}^{\infty} \frac{1}{\kappa+1} = +\infty$-harmonic series, in the linear processes paragraph we will see that there exist processes with this autocovariance function).

Example 9. Suppose that $\varepsilon \sim$ WN(σ^2), then $\gamma_\kappa = \begin{cases} \sigma^2 & \kappa = 0 \\ 0 & \kappa > 0 \end{cases}$, $\rho_\kappa = \begin{cases} 1 & \kappa = 0 \\ 0 & \kappa > 0 \end{cases}$ hence a white noise process is a regular short memory process. Furthermore, $\sum\limits_{\kappa=0}^{\infty} |\gamma_\kappa| = \sigma^2 < +\infty$ hence it is also short memory.

4 Comparison: Strict and Weak Stationarity in the Independence Framework

In what follows suppose that x is a well defined time series comprised by **independent random variables. In this framework** we will provide with examples which imply that any combination of those processes is possible, hence the linguistic interpretation of the terms can be misleading. During the course we will examine more complex analogous examples in the framework of dependence.

Example 10. [Neither Strictly, Nor Weakly Stationary]. See example 5 (heterogeneity and moments depending on t imply that both properties fail to hold).

Example 11. [Both Strictly and Weakly Stationary] See example 4 along with $x_t \sim N\left(0, \sigma^2\right)$ (homogeneity and enough moment existence conditions imply that both properties hold).

Example 12. [Strictly but not Weakly Stationary] $x_t \sim t_v$, where $v \leq 2$ (homogeneity implies strict stationarity, while insufficient moment existence conditions imply weak stationarity does not hold due to the failure of condition 2 (and 1 when $v \leq 1$)).

Example 13. [Not Strictly but Weakly Stationary] $x_t \sim N(0,1)\,,\ t \neq 0,\ \sqrt{\frac{v}{v-2}}x_0 \sim t_v$, where $v > 2$ (heterogeneity implies that strict stationarity does not hold, yet independence and sufficient moment existence conditions that are independent of t, imply that weak stationarity holds).

5 Transformations and Hereditarity

Our setup is incomplete, yet it is sufficient to study the issues of hereditarity of the previous properties under particular transformations. This will be useful when for example, we study whether such properties are valid for processes defined either by linear transformations, and/or as solutions of stochastic difference equations (or stochastic recurrence equations-**SRE**s), w.r.t. processes that we already know that they have them. We first define an extension of the property of stationarity.

Definition 11. The processes $x_1 = \left(x_{1_t}\right)_{t\in\mathbb{Z}}$, $x_2 = \left(x_{2_t}\right)_{t\in\mathbb{Z}}$, ..., $x_\rho = \left(x_{\rho_t}\right)_{t\in\mathbb{Z}}$ are jointly strictly stationary iff the joint distribution of every finite collection of random variables from $x_1, x_2, \ldots, x_\rho$ is invariant w.r.t. time shifts.

Hence stationarity is a subcase of joint stationarity for $\rho = 1$. Notice that joint stationarity implies stationarity for every of the involved processes but the converse does not necessarily hold.

Lemma 1. *Suppose that* $x_1 = \left(x_{1_t}\right)_{t\in\mathbb{Z}}$, $x_2 = \left(x_{2_t}\right)_{t\in\mathbb{Z}}$, ..., $x_\rho = \left(x_{\rho_t}\right)_{t\in\mathbb{Z}}$ *are jointly strictly stationary. Let* $f : \mathbb{R}^\rho \to \mathbb{R}$ *be measurable (appropriately). Then* $y = \left(f\left(x_{1_t}, \ldots, x_{\rho_t}\right)\right)_{t\in\mathbb{Z}}$ *is strictly stationary.*

Proof. Out of the scope of the course (although not that difficult! Try it as an optional exercise.). $\square$

Example 14. $\rho = 1$, $f(z) = z^4$. $y = \left(x_t^4\right)_{t\in\mathbb{Z}}$ is strictly stationary if x is. $\rho = 2$ $f(z_1, z_2) = z_1 + z_2 + z_1 z_2$, then $y = \left(x_{1_t}^4 + x_{2_t} + x_{1_t} x_{2_t}^4\right)_{t\in\mathbb{Z}}$ is strictly stationary, if x_1, x_2 are jointly strictly stationary.

Notice that the results of Lemma 1 do not *generally* hold for weak stationarity since f may destroy moment existence conditions (find examples!).

Remember the lag operator L, and via this, consider the operator $L_\star : \mathbb{R}^\mathbb{Z} \to \mathbb{R}^\mathbb{Z}$ defined by $L_\star x \triangleq (Lx_t)_{t\in\mathbb{Z}} = (x_{t-1})_{t\in\mathbb{Z}}$. It is easy to see $L_\star$ is linear, i.e. if $\alpha, \beta \in \mathbb{R}$, then $L_\star(\alpha x + \beta y) = \alpha L_\star x + \beta L_\star y$, and invertible, i.e. $L_\star^{-1} x = \left(L^{-1} x_t\right)_{t\in\mathbb{Z}} = (x_{t+1})_{t\in\mathbb{Z}}$, due to the fact that $L_\star$ is defined via pointwise application of L and the latter has those properties. Then for any $m \in \mathbb{Z}$, we can extend it to $L_\star^m \triangleq \begin{cases} \overbrace{L_\star \circ L_\star \circ \ldots \circ L_\star}^{m-times} & \kappa \geq 0 \\ \underbrace{L_\star^{-1} \circ L_\star^{-1} \circ \ldots \circ L_\star^{-1}}_{|m|-times} & \kappa < 0 \end{cases}$

i.e. $L_\star^m x = (x_{t-m})_{t\in\mathbb{Z}}$.

Lemma 2. *If x is strictly stationary then $L_\star^m x$ is strictly stationary. If x is weakly stationary then $L_\star^m x$ is weakly stationary, and $\gamma_\kappa(x) = \gamma_\kappa(L_\star^m x)\ \forall \kappa \geq 0$.*

Proof. Exercise! □

Example 15. If x is strictly stationary then $y = (x_t^3 x_{t-m}^2)$ is strictly stationary. Simply combine Lemmata 1-2.

Lemma 3. *If x is strictly stationary and f_m:$\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ measurable and $f_m(y_1, y_2, \ldots) \to f(y_1, y_2, \ldots)$ for $\mathbb{P}$ almost every element of $\mathbb{R}^{\mathbb{N}}$ as $m \to \infty$ (see 6). Then $y = (f(x_t, x_{t-1}, \ldots))_{t\in\mathbb{Z}}$ exists as a stochastic process and is strictly stationary.*

Proof. Out of the scope of the course. □

Example 16. Suppose that $f_m(y_1, y_2, \ldots) = f(y_1, y_2, \ldots) = \max_{0\leq i\leq 100} y_i$. Then if x is stationary, $y = (\max_{0\leq i\leq 100} y_{t-i})_{t\in\mathbb{Z}}$ is stationary due to Lemma 3.

6 Linear Transformations: Causal Linear Processes

One important example of processes obtained by some transformation (as the above) of a given building block process is the one of *linear processes*. A particular instance is that of a causal linear process based on a white noise and a sequence of coefficients that converge to zero in a sufficiently high rate. Suppose that $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise process($\varepsilon \sim$ WN (σ^2)). Furthermore, for the real sequence $(\beta_j)_{j\in\mathbb{N}}$, suppose that

$$\sum_{j=0}^{\infty} |\beta_j| < +\infty,$$

i.e. the coefficient sequence is absolute summable, and in order to avoid degeneracies, suppose that $\beta_j \neq 0$ for some $j \in \mathbb{N}$.

Remark 6. It can be proven (using the completeness of relevant space of random variables endowed with the L_1-norm, and the assumed properties of ε and $(\beta_j)_{j\in\mathbb{N}}$) that $\forall t \in \mathbb{Z}$ the random variable $y_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}$, obtained as a stochastic series based on elements of ε and the coefficients sequence $(\beta_j)_{j\in\mathbb{N}}$, is well defined. Furthermore, and via a dominated convergence argument it can be proven that in this framework, and when the relevant expectations exist, then the $\mathbb{E}$ and the $\sum_{j=0}^{\infty}$ operators *commute*.

Definition 12. The process $y = (y_t)_{t\in\mathbb{Z}}$ is called *linear causal* on $(\varepsilon_t)_{t\in\mathbb{Z}}$ and the absolutely summable $(\beta_j)_{j\in\mathbb{N}}$.

Remark 7. Given the time interpretation of $\Theta = \mathbb{Z}$, the term *causal*, refers to the fact that each element y_t of y, is determined by the elements of the constructing process ε that correspond to time instances $s \le t$ (Justify the term linear!).

Proposition 1. *y is weakly stationary.*

Proof. Given Remark 6 we readily perform the following computations:

a. $\mathbb{E}(y_t) = \mathbb{E}\left(\sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \beta_j \mathbb{E}\left(\varepsilon_{t-j}\right) = \mathbb{E}(\varepsilon_0) \sum_{j=0}^{\infty} \beta_j = 0$, since the series converges and $\mathbb{E}(\varepsilon_0) = 0$. Hence the first moments exist and are independent of t.

b. Due to the previous,
$$\mathbb{V}ar(y_t) = \mathbb{E}(y_t^2) = \mathbb{E}\left(\sum_{j=0}^{\infty}\sum_{j^\star=0}^{\infty} \beta_j \beta_{j^\star} \varepsilon_{t-j}\varepsilon_{t-j^\star}\right) = \sum_{j,j^\star=0}^{\infty} \beta_j \beta_{j^\star} \mathbb{E}\left(\varepsilon_{t-j}\varepsilon_{t-j^\star}\right).$$
Notice that since ε is white noise $\mathbb{E}\left(\varepsilon_{t-j}\varepsilon_{t-j^\star}\right) = \begin{cases} 0, & j \neq j^\star \\ \sigma^2, & j = j^\star \end{cases}$ hence $Var(y_t) = \sigma^2 \sum_{j=0}^{\infty} \beta_j^2$, which converges since $\sum_{j=0}^{\infty} \left|\beta_j\right|$ converges. Hence the variances exist and are independent of t.

c. Analogously, for any $\kappa \in \mathbb{N}^\star$,
$$Cov(y_t, y_{t-\kappa}) = \mathbb{E}(y_t, y_{t-\kappa}) = \mathbb{E}\left(\sum_{j=0}^{\infty}\sum_{j^\star=0}^{\infty} \beta_j \beta_{j^\star} \varepsilon_{t-j}\varepsilon_{t-\kappa-j^\star}\right)$$
$$= \sum_{j,j^\star=0}^{\infty} \beta_j \beta_{j^\star} E\left(\varepsilon_{t-j}\varepsilon_{t-\kappa-j^\star}\right).$$
Notice that $\mathbb{E}\left(\varepsilon_{t-j}\varepsilon_{t-\kappa-j^\star}\right) = \begin{cases} 0 & j \neq j^\star + \kappa \\ \sigma^2 & j = j^\star + \kappa \end{cases}$. Hence
$$Cov(y_t, y_{t-\kappa}) = \sigma^2 \sum_{j^\star=0, j=j^\star+\kappa}^{\infty} \beta_j \beta_{j^\star} = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+\kappa}$$
which converges since $\sum_{j=0}^{\infty} \left|\beta_j\right|$ converges. Notice that $\sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+\kappa}$ is independent of t (yet it may depend on κ).

$\square$

Remark 8. By an argument involving partial sum sequences, it is easy to see that, since $\mathbb{E}\left(\varepsilon_t\right) = 0\ \forall t$, then $\mathbb{E}\left(\sum\limits_{j=0}^{\infty} \beta_j \varepsilon_{t-j}\right) = 0$, even if $\sum\limits_{j=0}^{\infty} \beta_j$ diverges. Given this it is easy to see that the results in this paragraph would hold under the weaker condition $\sum\limits_{j=0}^{\infty} \beta_j^2 < +\infty$, with the appropriate modifications.

Remark 9. Notice that for any $\kappa \geq 0$, due to the Cauchy-Schwarz inequality $\sum\limits_{j=0}^{\infty} \beta_j \beta_{j+\kappa} \leq \sqrt{\sum\limits_{j=0}^{\infty} \beta_j^2 \sum\limits_{j=0}^{\infty} \beta_{j+\kappa}^2}\ (*)$. Also, there exists some $C > 0$ such that $\sum\limits_{j=0}^{\infty} \beta_j^2 \leq C \sum\limits_{j=0}^{\infty} \left|\beta_j\right|$ (why?), while $\sum\limits_{j=0}^{\infty} \left|\beta_{j+\kappa}\right| \underset{i=j+\kappa}{=} \sum\limits_{i=\kappa}^{\infty} \left|\beta_i\right| \leq \sum\limits_{i=0}^{\infty} \left|\beta_i\right| < +\infty$. Hence, the right hand side of $(*)$ is less than or equal to $\sum\limits_{j=0}^{\infty} \left|\beta_j\right| < +\infty$.

Remark 10. y has autocovariance function $\gamma_\kappa = \sigma^2 \sum\limits_{j=0}^{\infty} \beta_j \beta_{j+\kappa}$, $\kappa \in \mathbb{N}$, and autocorrelation function $\rho_\kappa = \frac{\sum_{j=0}^{\infty} \beta_j \beta_{j+\kappa}}{\sum_{j=0}^{\infty} \beta_j^2}$, $\kappa \in \mathbb{N}$.

Example 17. Suppose that $\beta_j = \beta^j$ for some $\beta \in (-1, 1)$. Then $\gamma_\kappa = \sigma^2 \sum\limits_{j=0}^{\infty} \beta^j \beta^{j+\kappa} = \sigma^2 \beta^\kappa \sum\limits_{j=0}^{\infty} \beta^{2j} = \sigma^2 \frac{\beta^\kappa}{1-\beta^2}$, and $\rho_\kappa = \beta^\kappa$. Notice that $\gamma_\kappa \to 0$, $\kappa \to \infty$ and $\sum\limits_{\kappa=0}^{\infty} \gamma_\kappa = \frac{\sigma^2}{\mathbf{1-\beta^2}} \sum\limits_{\kappa=0}^{\infty} \beta^\kappa = \frac{\sigma^2}{(1-\beta^2)(1-\beta)}$. Hence, it is a regular short memory process, termed as an AR(1) process.

Example 18. Suppose as a further example that for some $m \in \mathbb{N}$, $\beta_m \neq 0$, $\beta_{m+j} = 0,\ \forall j > 0$. Then we have that $\gamma_\kappa = \sigma^2 \sum\limits_{j=0}^{\infty} \beta_j \beta_{j+\kappa} = \begin{cases} \sigma^2 \sum\limits_{j=0}^{m-\kappa} \beta_j \beta_{j+\kappa}, & \text{if } \kappa \leq m \\ 0, & \text{if } \kappa > m \end{cases}$. Hence, $\gamma_0 = \sigma^2 \sum\limits_{j=0}^{m} \beta_j^2$ and thereby $\rho_\kappa = \begin{cases} \frac{\sum_{j=0}^{m-\kappa} \beta_j \beta_{j+\kappa}}{\sum_{j=0}^{m} \beta_j^2}, & \text{if } \kappa \leq m \\ 0, & \text{if } \kappa > m \end{cases}$. Obviously, this is also a regular short memory process termed as an MA(m) process (why?).

Remark 11. In the proof of general linear causal case the notation $\sum\limits_{j \in A, j^* \in B}$ denotes the double sum $\sum\limits_{j \in A} \sum\limits_{j^* \in B}$. Furthermore, when a functional relation is present between the two indices, e.g. $\sum\limits_{j^*=0, j=f(j^*)}^{\infty}$ then this is actually a single sum (why?).

Proposition 2. *If ε is also strictly stationary then y is strictly stationary.*

Proof. Use Lemma 3 with $f_m(x_0, x_1, ...) = \sum_{\kappa=0}^{m} \beta_\kappa x_\kappa$ and $f(x_0, x_1, ...) = \sum_{\kappa=0}^{\infty} \beta_\kappa x_\kappa$ along with Remark 6. □

Remark 12. We can extend the notion of a linear causal process, by allowing ε to be arbitrary, and/or the sequence of coefficients to possess weaker properties. In doing so it is possible that several of the aforementioned properties would cease to hold.

The notes are in state of "perpetual" correction. They do not substitute the lectures. Please report any typos to stelios@aueb.gr on the course's e-class.