Spectral representations and ergodic theorems for stationary stochastic processes

Stationary stochastic processes

Theory and methods for stochastic processes are considerably simplified under the assumption of (either strong or weak) stationarity, that imposes certain structure on the set of fdds (strong stationarity) or the mean function and the (auto)covariance function (weak stationarity). Stationarity also has deeper consequences, including spectral theorems and ergodic theorems.

A stochastic process $X$ is strongly stationary if its fdds are invariant under time shifts, that is, for any (finite) $n$, for any $t_0$ and for all $t_1, ..., t_n \in T$, $(X_{t_1}, ..., X_{t_n})$ and $(X_{t_1+t_0}, ..., X_{t_n+t_0})$ have the same distribution.

A stochastic process $X$ is weakly stationary if its mean function is constant and its covariance function is invariant under time shifts. That is, for all $t \in T$, $E(X_t) = \mu$ and for all $t_i, t_j \in T$, $Cov(X_{t_i}, X_{t_j}) = c(t_i - t_j)$, a function of $t_i - t_j$ only. (Note that the definition of weak stationarity implicitly assumes existence of first and second order moments of the process.)

Spectral theorems for stationary processes

From the theory of Fourier analysis, any function $f : \mathbb{R} \to \mathbb{R}$ with certain properties (including periodicity and continuity) has a unique Fourier expansion $f(x) = 0.5a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$, that expresses $f$ as a sum of varying proportions of regular oscillations. In some sense, (weakly) stationary processes are similar to periodic functions since their autocovariance functions are invariant under time shifts. The spectral theorem yields that, under certain conditions, stationary processes can be decomposed in terms of regular underlying oscillations whose magnitudes are random variables.

In spectral theory it is convenient to allow for stochastic processes that take values in the complex plane $C$. This provides the natural setting for the theory but does require extensions of the concepts and definitions we have seen for stochastic processes with state spaces $S \subseteq \mathbb{R}^k$. (See Appendix C for a discussion of complex-valued stochastic processes.)

Consider first weakly stationary continuous-time (with $T = \mathbb{R}$) stochastic processes $X = \{X_t : t \in \mathbb{R}\}$ (that take values in $C$).

By weak stationarity, we have that $E(X_t) = \mu$, for all $t \in \mathbb{R}$, and $Cov(X_s, X_{s+t}) = c(t)$ a function of $t$ only, for any $s, t \in \mathbb{R}$. Note that $\text{Var}(X_t) = Cov(X_t, X_t) = c(0) \equiv \sigma^2$, for all $t \in \mathbb{R}$, that is,
the variance is also constant. (Hence, we typically assume, without loss of generality, $\mu = 0$ and $\sigma^2 = 1$ for a weakly stationary process with strictly positive variance.)

The autocorrelation function of $X$ is given by

$$
\text{Corr}(X_s, X_{s+t}) = \frac{\text{Cov}(X_s, X_{s+t})}{\sqrt{\text{Var}(X_s)\text{Var}(X_{s+t})}} = \frac{c(t)}{c(0)} \equiv r(t),
$$

for all $s, t \in \mathbb{R}$ (again, a function of $t$ only), provided $\text{Var}(X_t) = c(0) > 0$.

The spectral theorem for autocorrelation functions describes regular oscillations within the random fluctuation of a weakly stationary stochastic process through such oscillations in its autocorrelation function.

**Spectral theorem for autocorrelation functions:** Consider a continuous-time weakly stationary stochastic process $X = \{X_t : t \in \mathbb{R}\}$ with strictly positive variance. If the autocorrelation function $r(t)$ of $X$ is continuous at $t = 0$, then $r(t)$ is the characteristic function of some distribution function $F$, that is,

$$
r(t) = \int_{-\infty}^{\infty} \exp(itu)dF(u).
$$

(Based on Bochner’s theorem from Appendix A, proving the theorem reduces essentially to checking uniform continuity for $r(t)$.)

The distribution function $F$ is called the *spectral distribution function* of the process. The uniqueness result for characteristic functions (see Appendix A) implies the uniqueness of the spectral distribution function. The *spectral density function* of the process is the density function that corresponds to $F$ whenever this density exists. The inversion techniques for characteristic functions (see Appendix A) yield expressions for the spectral distribution and density functions in terms of the autocorrelation function of the process. The *spectrum* of $X$ is the set of all real numbers $u$ with the property that $F(u + \epsilon) - F(u - \epsilon) > 0$, for all $\epsilon > 0$ (that is, the support of the spectral distribution function $F$).

To interpret the result, consider a random variable $U$ with distribution function $F$, so that $\exp(itU) = \cos(tU) + isin(tU)$ is a pure oscillation with a random frequency. Then, under the conditions of the theorem, $r(t)$ is the expectation of this random oscillation with respect to the spectral distribution of the process.

Turning to discrete-time (with say $T = \mathbb{Z}$, the set of integers) weakly stationary stochastic processes $X = \{X_n : n \in \mathbb{Z}\}$, results from characteristic functions are not directly applicable, since now the autocorrelation function is a function on $\mathbb{Z}$ (taking again values in $C$). Of course, continuity conditions for $r$ are not relevant here. Moreover, the representation in this case, $r(n) = \int_{-\infty}^{\infty} \exp(inu)dF(u)$ for some distribution function $F$, is not unique, since the function $\exp(inu)$ is periodic in $u$ (for all $n$, $\exp(in(u+2\pi)) = \exp(inu)$). Hence, the spectral theorem for discrete-time weakly stationary processes is typically given in the form

$$
r(n) = \int_{[-\pi, \pi]} \exp(inu)dF^*(u),
$$

for a distribution function $F^*$ that results from $F$ but is truncated in the interval $[-\pi, \pi]$ (thus, $F^*(-\pi) = 0$ and $F^*(\pi) = 1$). Inversion theorems can be used to obtain expressions for the spectral
distribution in terms of \( r \). For example, if \( F^* \) has density \( f \), then
\[
f(u) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \exp(-inu)r(n),
\]
at every point \( u \in [-\pi, \pi] \) at which \( f \) is differentiable.

The above results simplify further for discrete-time processes that are real-valued. In this case, \( r(n) = \int_{[-\pi,\pi]} \cos(nu)dF^*(u) \), since \( r(n) = r(-n) \). In addition, \( \cos(nu) = \cos(-nu) \) and therefore an expression for the autocorrelation function of a discrete-time weakly stationary real-valued process is
\[
r(n) = \int_{[-\pi,\pi]} \cos(nu)dG(u),
\]
where \( G \) is the distribution function of a symmetric distribution on \([-\pi, \pi]\). The expression for the spectral density function becomes
\[
f(u) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \cos(nu)r(n),
\]
for \( u \in [-\pi, \pi] \) at which \( f \) is differentiable.

Besides several applications in time series analysis, spectral representations for autocorrelation (or autocovariance) functions are also very important for spatial stochastic processes (here \( T \subseteq R^d \), \( d > 1 \)). For example, the theory is used to construct valid covariogram models in \( R^d \) for spatial data modeling (see, e.g., section 2.5 in Statistics for Spatial Data, 1993, by Cressie).

Note that the above results are essentially analytical providing representations for a deterministic function (the autocorrelation function) of the stationary process \( X \). Of more (probabilistic) interest is perhaps a spectral representation of the process \( X \) itself. Such a representation is possible under conditions (for example, \( X \) must have a continuous autocorrelation function if it is a continuous-time stationary process) and is the result of the spectral theorem for stationary processes. For a continuous-time stationary process \( X = \{X_t : t \in R\} \) (taking values in \( C \) as above) the spectral theorem yields the representation
\[
X_t = \int_{-\infty}^{\infty} \exp(itu)dS_u,
\]
where \( S = \{S_u : u \in R\} \) is a complex-valued stochastic process (the spectral process of \( X \)) that has orthogonal increments (that is, \( E((S_v - S_u)(\overline{S_s} - \overline{S_t})) = 0 \), for any \( u \leq v \leq s \leq t \) and is related with the spectral distribution function \( F \) through \( E(|S_v - S_u|^2) = F(v) - F(u) \), if \( u \leq v \).

A similar representation is available for discrete-time stationary processes, the main difference being that the index set of the spectral process can now be taken to be \((-\pi, \pi]\).

The integral above is a stochastic integral as it involves a stochastic process for its integrating function. (Stochastic integration is essential in modern probability theory, e.g., for the study of diffusion processes.) In fact, use of the familiar notation for integrals should not create any confusion here; the result of this stochastic integration is a random variable that is defined as the mean-square limit of finite approximating sums. (See section 9.4 of Probability and Random Processes, 2001,
by Grimmett and Stirzaker, for a discussion of stochastic integration and the proof of the spectral theorem.

Ergodic theorems for stationary processes

Given a (countable) sequence \( \{X_j : j \geq 1\} \) of random variables, the study of the asymptotic behavior of the resulting sequence \( \{S_n : n \geq 1\} \), where \( S_n = \sum_{j=1}^{n} X_j \), is of great importance in probability and statistics. This is a problem that has been studied since the early years of probability theory in several forms and under several conditions. (See Appendix B for some definitions of convergence for sequences of random variables.)

The standard related results (the various laws of large numbers) rely heavily on independence of the random variables \( X_j \). For example, a simple application of Chebyshev’s inequality yields the Weak law of large numbers: If the \( X_j \) are independent and identically distributed with finite mean \( \mu \) and finite variance, then \( n^{-1}S_n \to \mu \) in mean square (and hence also \( n^{-1}S_n \to^p \mu \)).

One version of the strong law of large numbers that is easy to prove (using the Kronecker lemma for series of real numbers) but requires a finite second moment is given by the following

**Theorem:** If the \( X_j \) are independent with finite means (say, without loss of generality, all equal to 0), \( E(X_j^2) < \infty \), for all \( j \), and \( \sum_{j=1}^{\infty} j^{-2}E(X_j^2) < \infty \), then \( n^{-1}S_n \to^{a.s.} 0 \).

As a corollary to the theorem, we obtain that if the \( X_j \) are independent and identically distributed with finite mean \( \mu \) and finite variance, then \( n^{-1}S_n \to^{a.s.} \mu \).

Finally, an improved version of the theorem above yields the

**(Kolmogorov) Strong law of large numbers:** If the \( X_j \) are independent and identically distributed with \( E(|X_1|) < \infty \), then \( n^{-1}S_n \to^{a.s.} E(X_1) \). Moreover, if \( E(|X_1|) = \infty \), then \( n^{-1}S_n \) diverges with probability one.

The ergodic theorems for stationary processes provide a very important generalization of the laws of large numbers by replacing the assumption of independence for the \( X_j \) with the assumption that they form a stationary process. Stated below are two versions of the ergodic theorem (for discrete-time processes), depending on the type of stationarity, weak or strong. Note, again, that under weak stationarity we implicitly assume existence of first and second order moments of the process.

**Ergodic theorem for weakly stationary processes:** If \( X = \{X_j : j \geq 1\} \) is a weakly stationary process, there exists a random variable \( Y \) such that \( E(Y) = E(X_1) \) and \( n^{-1}S_n \to Y \) in mean square.

**Ergodic theorem for strongly stationary processes:** If \( X = \{X_j : j \geq 1\} \) is a strongly stationary process such that \( E(|X_1|) < \infty \), then there exists a random variable \( Y \) with \( E(Y) = E(X_1) \) and \( n^{-1}S_n \to Y \) almost surely and in mean square.
Appendix A: Background on characteristic functions

The characteristic function of a random variable provides a very useful tool to study theoretical properties of the random variable and is a key function for the spectral representation results for stationary processes.

By definition, the characteristic function \( \phi \) (or \( \phi_X \)) of a random variable \( X \) is a function on \( \mathbb{R} \) taking values on the complex plane and given by

\[
\phi(t) = \mathbb{E} \left( e^{itX} \right),
\]

where \( i = \sqrt{-1} \).

Note that characteristic functions are related to Fourier transforms as

\[
\phi(t) = \int \exp(itx) dF(x),
\]

where \( F \) is the distribution function of \( X \). A key property of \( \phi \) is that it is always well defined and, in fact, finite, since

\[
\phi(t) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX)).
\]

This is an advantage over the moment generating function \( m(t) = \mathbb{E}(e^{tX}) \), as is the fact that, in general, \( \phi \) has better analytical properties than \( m \).

For instance, Bochner’s theorem, one of the important results for characteristic functions, yields that the following three conditions are necessary and sufficient for a function \( \phi \) to be the characteristic function of a random variable \( X \):

(a) \( \phi(0) = 1, \ |\phi(t)| \leq 1, \) for all \( t \).
(b) \( \phi \) is uniformly continuous on \( \mathbb{R} \) (that is, for all \( \epsilon > 0 \), there exists some \( \delta > 0 \) such that for all \( s,t \in \mathbb{R} \) with \( |s-t| < \delta, \ |\phi(s) - \phi(t)| < \epsilon \)).
(c) \( \phi \) is a non-negative definite function (that is, for all real \( t_1,...,t_n \) and complex \( z_1,...,z_n \),

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} z_i \overline{z_j} \phi(t_i - t_j) \geq 0.
\]

Moments of a random variable \( X \) are generated by its characteristic function \( \phi \). The extension of Taylor’s theorem for complex-valued functions yields

\[
\phi(t) \approx \sum_{j=0}^{k} \frac{\mathbb{E}(X^j)}{j!} (it)^j,
\]

provided \( \mathbb{E}|X^j| < \infty \). Hence the \( k \)th order derivative of \( \phi \) evaluated at 0, \( \phi^{(k)}(0) = i^k \mathbb{E}(X^k) \).

Other useful properties include results for sums of independent random variables, \( \phi_{X+Y}(t) = \phi_X(t) \phi_Y(t) \) for independent random variables \( X \) and \( Y \), and linear combinations of random variables, \( \phi_{aX+b}(t) = \exp(itb)\phi_X(at) \) for constants \( a,b \in \mathbb{R} \).

Arguably, the most important property of a characteristic function is the fact that knowledge of \( \phi \) suffices to recapture the distribution of the corresponding random variable (not just moments as we have seen above). This is the inversion theorem for characteristic functions, a special case of which follows.

**Theorem:** Assume that \( X \) is a continuous random variable with density function \( f \) and characteristic function \( \phi \). Then for any point \( x \) at which \( f \) is differentiable,

\[
f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi(t)dt.
\]

This is essentially the Fourier inversion theorem. (Note that, even if the random variable is continuous, its density is not necessarily differentiable at any point.)

The general version of the inversion theorem for characteristic functions is more technical.
**Inversion Theorem:** Assume that $X$ is a random variable with distribution function $F$ and characteristic function $\phi$. Define $F^*: \mathbb{R} \to [0,1]$ by $F^*(x) = 0.5(F(x) + \lim_{y \to x} F(y))$. Then

$$F^*(b) - F^*(a) = \lim_{N \to \infty} \int_{-N}^{N} \frac{\exp(-iat) - \exp(-ibt)}{2\pi it} \phi(t) dt.$$ 

A corollary to the inversion theorem yields the familiar result on the characterization of a random variable in terms of its characteristic function, that is, random variables $X$ and $Y$ have the same distribution function if and only if they have the same characteristic function.

Another important result relates convergence of a sequence of distribution functions with convergence of the corresponding sequence of characteristic functions. This result is used in the proof of the standard version of the central limit theorem and certain laws of large numbers. Consider a sequence $\{F_n : n \geq 1\}$ of distribution functions (corresponding to a sequence $\{X_n : n \geq 1\}$ of random variables). We say that the sequence $\{F_n : n \geq 1\}$ converges to the distribution function $F$ (notation $F_n \to F$) if $F(x) = \lim_{n \to \infty} F_n(x)$ at any point $x$ where $F$ is continuous. This definition essentially yields the definition of one mode of convergence for sequences of random variables, namely convergence in distribution (see Appendix B).

**Continuity Theorem:** Consider a sequence of distribution functions $\{F_n : n \geq 1\}$ and the corresponding sequence of characteristic functions $\{\phi_n : n \geq 1\}$.

(i) If $F_n \to F$ for some distribution function $F$ with characteristic function $\phi$, then $\lim_{n \to \infty} \phi_n(t) = \phi(t)$ for all $t \in \mathbb{R}$.

(ii) If $\phi(t) = \lim_{n \to \infty} \phi_n(t)$ exists for all $t \in \mathbb{R}$ and $\phi$ is continuous at $t = 0$, then $\phi$ is the characteristic function of some distribution function $F$, and $F_n \to F$.

The definition of the characteristic function can be extended to (possibly dependent) collections of random variables. For example, the joint characteristic function of two random variables $X$ and $Y$ is defined by $\phi_{X,Y}(s,t) = E(\exp(isX)\exp(itY))$, for $s, t \in \mathbb{R}$.

Joint moments of $X$ and $Y$ can be obtained from their joint characteristic function $\phi_{X,Y}$. In particular, under appropriate conditions of differentiability,

$$E(X^mY^n) = \frac{\partial^{m+n} \phi_{X,Y}}{\partial s^m \partial t^n} |_{s=t=0},$$

for any positive integers $m$ and $n$.

Moreover, random variables $X$ and $Y$ are independent if and only if $\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t)$, for all $s, t \in \mathbb{R}$.

Finally, the inversion theorem can be extended to jointly distributed random variables. For instance, if $X$ and $Y$ are continuous random variables with joint density function $f_{X,Y}$ and joint characteristic function $\phi_{X,Y}$,

$$f_{X,Y}(x,y) = \frac{1}{4\pi^2} \int \int_{R^2} \exp(-isx)\exp(-ity)\phi_{X,Y}(s,t) ds dt,$$

for all $(x,y)$ at which $f_{X,Y}$ is differentiable.
Appendix B: Modes of convergence for sequences of random variables

Given a sequence of random variables \( \{X_n : n \geq 1\} \) and some limiting random variable \( X \), there are several ways to formulate convergence “\( X_n \rightarrow X \) as \( n \rightarrow \infty \)”. The following four definitions are commonly employed to study various limiting results for random variables and stochastic processes.

**Almost sure convergence** \( (X_n \rightarrow a.s. \ X) \).
Let \( \{X_n : n \geq 1\} \) and \( X \) be random variables defined on some probability space \((\Omega, \mathcal{F}, P)\). \( \{X_n : n \geq 1\} \) converges almost surely to \( X \) if
\[
P(\left\{ \omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega) \right\}) = 1.
\]

**Convergence in \( r \)th mean** \( (X_n \rightarrow r-mean \ X) \).
Let \( \{X_n : n \geq 1\} \) and \( X \) be random variables defined on some probability space \((\Omega, \mathcal{F}, P)\). \( \{X_n : n \geq 1\} \) converges in mean of order \( r \geq 1 \) (or in \( r \)th mean) to \( X \) if
\[
E(|X_n - X|^r) < \infty \quad \text{for all} \quad n, \quad \text{and} \quad \lim_{n \to \infty} E(|X_n - X|^r) = 0.
\]

**Convergence in probability** \( (X_n \rightarrow p \ X) \).
Let \( \{X_n : n \geq 1\} \) and \( X \) be random variables defined on some probability space \((\Omega, \mathcal{F}, P)\). \( \{X_n : n \geq 1\} \) converges in probability to \( X \) if for any \( \epsilon > 0 \),
\[
\lim_{n \to \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0.
\]

**Convergence in distribution** \( (X_n \rightarrow d \ X) \).
Let \( \{X_n : n \geq 1\} \) and \( X \) be random variables with distribution functions \( \{F_n : n \geq 1\} \) and \( F \), respectively. \( \{X_n : n \geq 1\} \) converges in distribution to \( X \) if
\[
\lim_{n \to \infty} F_n(x) = F(x),
\]
for all points \( x \) at which \( F \) is continuous.

Note that the first three types of convergence require that \( X_n \) and \( X \) are all defined on the same underlying probability space, as they include statements involving the (common) probability measure \( P \). However, convergence in distribution applies to random variables defined possibly on different probability spaces, as it only involves the corresponding distribution functions.

It can be shown that:
- Almost sure convergence implies convergence in probability.
- Convergence in \( r \)th mean implies convergence in probability, for any \( r \geq 1 \).
- Convergence in probability implies convergence in distribution.
- Convergence in \( r \)th mean implies convergence in \( s \)th mean, for \( r > s \geq 1 \).

No other implications hold without further assumptions on \( \{X_n : n \geq 1\} \) and/or \( X \). If we do impose further structure, there are several results that can be obtained.
Appendix C: Complex-valued stochastic processes

First, recall standard operations on the complex plane $C = \{(a, b) : a, b \in R\}$ where to each pair $(a, b)$ we associate a complex number $z = a + ib$ with $i^2 = -1$. We have $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \times (c, d) = (ac - bd, ad + bc)$. The most commonly used norm on $C$ is defined by $|a + ib| = \sqrt{a^2 + b^2}$. The complex conjugate $\overline{z}$ of $z = a + ib$ is given by $\overline{z} = a - ib$, whence $z\overline{z} = |z|^2 = a^2 + b^2$.

For any two real-valued random variables $X$ and $Y$ defined on the same probability space, we can define a complex-valued random variable by $Z = X + iY$.

Statistical properties of $Z$ are studied through the joint distribution of $(X, Y)$. For example, assuming the expectations of $X$ and $Y$ exist, $E(Z) = E(X) + iE(Y)$. The covariance between two complex random variables $Z_1$ and $Z_2$ is defined by

$$\text{Cov}(Z_1, Z_2) = E((Z_1 - E(Z_1))(\overline{Z_2} - E(Z_2))) = E(Z_1\overline{Z_2}) - E(Z_1)E(\overline{Z_2}),$$

assuming again that all required expectations exist. Note that the covariance operator is not symmetric for complex random variables, since $\text{Cov}(Z_2, Z_1) \neq \text{Cov}(Z_1, Z_2)$. Complex random variables $Z_1$ and $Z_2$ are called orthogonal if $\text{Cov}(Z_1, Z_2) = 0$.

A collection $Z = (Z_1, ..., Z_n)$ of complex random variables $Z_j = X_j + iY_j$ defines a complex random vector. Statistical properties of $Z$ result from the joint distribution of $(X_1, ..., X_n, Y_1, ..., Y_n)$. For example, we say that the complex random variables $Z_j, j = 1, ..., n$, are independent if $f(x_1, ..., x_m, y_1, ..., y_m) = \prod_{j=1}^{m} f(x_j, y_j)$ for any $2 \leq m \leq n$.

Now a complex stochastic process $Z = \{Z(\omega, t) : \omega \in \Omega, t \in T\}$ is defined in the same fashion with real-valued stochastic processes, the difference being that for any fixed index point $t$ we have a complex random variable $Z_t = X_t + iY_t$ (and, in general, for any finite collection of fixed index points we have a complex random vector). It is useful to think about $Z$ in terms of two underlying real-valued stochastic processes $X = \{X(\omega, t) : \omega \in \Omega, t \in T\}$ and $Y = \{Y(\omega, t) : \omega \in \Omega, t \in T\}$. As above, to extend the standard definitions for stochastic processes, we need to take into account the fact that the distribution of a complex random variable arises from the joint distribution of two real random variables. Hence now the fdds of $Z$ will be defined through the joint fdds of $X$ and $Y$.

The definitions for the mean function and autocovariance function of $Z$ arise using the definitions of the mean and covariance for complex random variables given above. Again, the autocovariance function is a non-negative definite function, that is,

$$\sum_{r=1}^{k} \sum_{j=1}^{k} z_r \overline{z_j} \text{Cov}(Z_{t_r}, Z_{t_j}) \geq 0,$$

for all (finite) $k$, for any $t_1, ..., t_k \in T$ and for any complex constants $z_1, ..., z_k$.

Two real-valued stochastic processes $X$ and $Y$ are jointly strongly stationary if for any (finite) $n$, for any $t_0$ and for all $t_1, ..., t_n \in T$, $(X_{t_1}, ..., X_{t_n}, Y_{t_1}, ..., Y_{t_n})$ and $(X_{t_1+t_0}, ..., X_{t_n+t_0}, Y_{t_1+t_0}, ..., Y_{t_n+t_0})$ have the same distribution. The complex-valued stochastic process $Z$ is strongly stationary if its associated real-valued stochastic processes $X$ and $Y$ are jointly strongly stationary.

The definition of weak stationarity is the same with before, where now the mean and covariance function for $Z$ result using the more general definitions for complex-valued random variables.