STATIONARY PROCESSES

K.Grill
Institut für Statistik und Wahrscheinlichkeitstheorie, TU Wien, Austria

Keywords: Stationary process, weak sense stationary process, strict sense stationary process, correlation function, spectral process, spectral function, spectral density, mean square error prediction, singularity, regularity, linear singularity, linear regularity.

Contents

Summary

Stationary processes are stochastic processes whose probabilistic structure is unaffected by shifts in time. According to the interpretation of the term "probabilistic structure", one distinguishes weak sense stationary processes, where only the covariance structure is supposed to be invariant, and strict sense stationary processes, for which all finite-dimensional distributions have to remain the same under shifts of time. Some important basic properties are discussed, and the spectral representation of a stationary process and its relation to questions of linear prediction are studied.

1. Introduction

Stationary processes are defined by the property that their behavior is not affected by a shift of the time variable. Depending on how one chooses to interpret the word "behavior" above, one arrives at one of the following definitions:

1. A stochastic process $(\xi(t) t \in T)$, (see *Stochastic Processes and Random Fields*) with $T = \mathbb{R}$ or $T = \mathbb{Z}$ is called stationary in the weak sense, or in short "weakly stationary", if the expectation $\mathbf{E}(\xi^2(t))$ is finite for all t, and if

$$\mathbf{E}\left(\xi\left(t\right)\right) = m, \tag{1}$$

and

$$\mathbf{Cov}\left(\xi(s),\xi(t)\right)=\mathbf{Cov}\left(\xi(s+h),\xi(t+h)\right)=R(t-s), \qquad (2)$$

i.e., if the expectation of $\xi(t)$ is constant and the covariance of $\xi(s)$ and $\xi(t)$ depends only on $t-s$.

2. A stochastic process is $\xi(t)$ is called stationary in the strict sense, if the distribution of

$$\xi(t_1),\xi(t_2),...,\xi(t_n) \qquad (3)$$

is the same as that of

$$\xi(t_1+h),\xi(t_2+h),....,\xi(t_n+h) \qquad (4)$$

for any choice of $t_1,...,t_n$ and *h*.

It is clear that any strictly stationary stochastic process with finite variances is also weakly stationary, but the definition of a strict sense stationary process does not include the finiteness of the variance, not even the finiteness of the expectation. The function $R(t)=\mathbf{Cov}(\xi(s),\xi(s+t))$ is called the correlation function of the process $\xi(t)$. Furthermore, in most cases, the value of *m* = $\mathbf{E}(\xi(t))$ does not matter, so one can assume without loss of generality that *m* = 0.

Consider a few simple examples:

1. The simplest possible (and the least interesting) example is a constant process: $\xi(t)=\xi(0)$.

2. Almost as simple as the first example: A sequence of independent identically distributed random variables $...,\xi(-1),\xi(0),\xi(1),...$. Here the correlation function is $R(0)=\sigma^2, R(t)=0$ for $t\neq 0$.

3. A little more sophisticated is the following example: Let $\eta(n), n\in\mathbb{R}$ be a sequence of independent identically distributed random variables and $a_i, i=0,...,k$ some real constants. Then

$$\xi(t)=\sum_{i=0}^{k} a_i\eta(t-i) \qquad (5)$$

is a stationary process, the so-called moving average (MA) process.

4. One can reverse the roles of ξ and η in the last example; this gives rise to the following definition :

A stationary sequence $\xi(t)$ is called an autoregressive (AR) process if there are constants $b_0,...,b_m$ such that

$$\eta(t) = \sum_{j=0}^{m} b_j \xi(t-j) \tag{6}$$

is a sequence of uncorrelated random variables.

5. Combining the last two examples, one gets the so-called autoregressive moving average (ARMA) process. This is a sequence $\xi(t)$ such that

$$\sum_{j=0}^{m} b_j \xi(t-j) = \sum_{i=0}^{n} a_i \eta(t-i), \tag{7}$$

where $\eta(t)$ is again a sequence of uncorrelated variables.

In this and the preceding example, there is the question of whether an AR (or ARMA) process exists for a given sequence $b_0,...,b_m$. For example, if $b_0 = 1, b_1 = -1,$ then there exists no stationary solution of the autoregressive equation; it implies that $\xi(t) = \xi(t-k) + \sum_{j=0}^{k-1} \eta_{t-j},$ which contradicts the assumption that $\xi(t)$ has finite variance.

A more detailed analysis (the discussion of spectral densities later in this chapter can be used to prove this) shows that the AR equation has a stationary solution only if the polynomial with coefficients $b_0,...,b_m$ doesn't have roots on the unit circle (i.e., if $\sum b_i z^i \neq 0$ for all complex z with $|z| = 1$)

6. Let *A* and η be two random variables with arbitrary joint distribution on $[0,\infty)\times[0,\infty)$ and assume that ϕ is another random variable, independent of *A* and η, and uniformly distributed on. $[0,2\pi]$. Then

$$\xi(t) = A\cos(\eta t + \phi) \tag{8}$$

is a stationary process (this follows from the fact that $\eta h + \phi$ modulo 2π is again uniformly distributed on $[0,2\pi]$ and independent of *A* and η. The correlation function of this process (if *A* has finite variance) is readily calculated as

$$R(t) = \mathbf{E}\Big(A^2\cos(\eta t+\phi)\cos(\phi)\Big)$$
$$= \mathbf{E}\Big(A^2\cos(\eta t)\Big) + \mathbf{E}\Big(A^2\cos(\eta t+2\phi)\Big). \tag{9}$$

The second expectation is zero because the argument of the cosine is again uniformly distributed modulo 2π and independent of *A*. A simple transformation of the first expectation yields

$$R(t) = \int_0^\infty \cos(yt)\,\mu(dy), \tag{10}$$

Where μ is a finite measure on $[0,\infty)$ defined by

$$\mu(B) = \frac{1}{2}\mathbf{E}\Big(A^2 I_B(\eta)\Big). \tag{11}$$

Letting $v(B) = \frac{1}{2}(\mu(B\cap[0,\infty)) + \mu((-B)\cap[0,\infty)))$, this becomes

$$R(t) = \int_{-\infty}^{\infty} e^{iyt}v(dy). \tag{12}$$

This way, one can get a correlation function of the form (??) with any finite symmetric measure *v*. In order to get the same form with a general, not necessarily symmetric measure *v*, it is sufficient to consider $\xi(t) = A\exp(i(\eta t+\phi))$, letting η take negative as well as positive values.

7. If $\xi(t)$ is a Gaussian process (i.e., all finite–dimensional distributions are multivariate Gaussian), then weak and strict stationarity are equivalent.

A particularly interesting example is obtained by letting

$$\xi(t) = e^{-t/2}W\left(e^t\right), \tag{13}$$

where W is a Wiener process, i.e., Gaussian process with mean zero and covariance function

$$R(s,t) = \min(s,t).$$

The correlation function of this process is

$$\begin{aligned} R(t) &= \mathbf{E}\big(\xi(s)\xi(s+t)\big) \\ &= \mathbf{E}\left(e^{-(2s+t)/2} W\left(e^{s}\right) W\left(e^{s+t}\right)\right) = e^{-t/2} \end{aligned} \tag{14}$$

If $t \geq 0$, and for general $t, R(t) = e^{-|t|/2}$. This is a Gaussian stationary Markov process (the Markov property follows from the fact that the Wiener process is Markov), and is called the Ornstein-Uhlenbeck process (see *Markov Processes*).

2. Spaces and Operators related to stationary processes

2.1 Spaces of square-integrable functions

This section introduces a number of spaces of functions and random variables that play an important role in the investigation of stationary processes.

First observe that the definition of a weak sense stationary process only involves the first two moments of the distribution $\xi(t)$; this makes it desirable to have a notion of convergence that can be expressed in terms of the first two moments. This is given by convergence in mean square. A sequence of $\xi(n)$ of random variables converges in mean square to a random variable ξ if

$$\mathbf{E}\left(\left|\xi(\eta) - \xi\right|^{2}\right) \to 0. \tag{15}$$

If one identifies random variables that are almost surely equal, then $\|\xi\|_2 = \sqrt{\mathbf{E}(\xi^2)}$ is a norm on the space of all (equivalence classes of) random variables with finite variance. Denote this space by $\mathbf{L}_2 = \mathbf{L}_2(\Omega, \mathcal{F}, \mathbf{P})$, indicating the underlying probability space to avoid ambiguities. On this space, $\|.\|_2$ is not only a norm, but it is also complete (i.e., any Cauchy sequence with respect to $\|.\|_2$ has a limit). In addition, this norm is generated by the inner product $\mathbf{E}(\xi\eta)$ (or $\mathbf{E}(\xi\overline{\eta})$ if one considers complex-valued random variables), so $\mathbf{L}_2$ is a Hilbert space. Convergence with respect to the norm $\|.\|_2$ is of course equivalent to convergence in mean square.

In the sequel, various subspaces of $\mathbf{L}_2$ will be needed. These are usually the spaces generated by certain sets of random variables. The most important case is the one where a process $\xi(t), t \in T$ is given and one considers all random variables that can be defined in terms of the values of $\xi(t)$ for t in some subset S and T; thus, let $\mathbf{L}_2(\xi(.), S)$ denoted the space of all random variables with finite variance that are measurable with respect to the σ- algebra $\mathcal{F}_S$ generated by the random variables

$\xi(t),\, t \in S$. Another way of defining this subspace is as the closure of the set of all $n \in \mathbf{L}_2$ that can be expressed in the form $\eta = f(\xi_1, \ldots, \xi_n)$ with some measurable function $f(.)$.

A similar notation will be used for spaces of measurable functions on the real line. If a Lebesgue Stieltjes measures on the real line is given by its cumulative distribution function *F*, then $\mathbf{L}_2(dF)$ is used to denote the space of all measurable functions that are square integrable with respect to *dF* (and identifying functions that agree almost everywhere with (*dF*)), endowed with the norm $\| f \|_2 = \left(\int |f|^2 \, dF \right)^{1/2}$.

For a weak sense stationary process $\xi(t), t \in T$ the space $\mathbf{L}_2(\xi(.), T)$ is still too rich; in particular, many of the random variables in this space have means and variances that cannot be expressed in terms of the mean and covariance function of $\xi(.)$. A random variable for which this can be achieved is that of a linear combination

$$\lambda_1 \xi(t_1) + \ldots + \lambda_n \xi(t_n) + c, \tag{16}$$

where $\lambda_1, \ldots \lambda_n$ and *c* are real numbers and $t_1, \ldots, t_n \in T$. Of course, for limits (in square mean) of variables of the above form, their mean and variance can also be expressed in terms of the mean and covariance function of $\xi(.)$. So, for $S \subset T$, let $\mathbf{H}_S$ denote the closure in $\mathbf{L}_2$ of the set of all random variables of the form (??) with $t_1, \ldots, t_n \in S$; the subspace $\mathbf{H}_S^0$ is obtained by imposing the additional condition *c* = 0 in (??).

2.2. Shift operators

These operators, which will be denoted by θ_h, are models for a “ shift in time”, i.e., they map the trajectory $\xi(t)$ into $\theta_h \xi(t) = \xi(t+h)$.

This type of operators can be defined for any (not necessarily stationary) stochastic process if one assumes that for any $\omega \in \Omega$ there exists a $\omega_h \in \Omega$ such that $\xi(t, \omega_h) = \xi(t+h, \omega)$. This, however, will not be followed further, in particular because there are subtle problems about measurability- in particular, the action of θ_h on spaces like $\mathbf{L}_2(\xi(.), S)$ or $\mathbf{H}_S$ may not be well defined, because the images of equivalent random variables need not be equivalent again.

For strict sense stationary processes, there is a simpler way to obtain the shift operators; namely, if $\eta \in \mathbf{L}_2(\xi(.), T)$ can be expressed in the form $\eta = f(\xi(t_1), \ldots, \xi(t_n))$, then let

$$\theta_h\eta = f(\xi(t_1+h),\ldots,\xi(t_n+h)). \tag{17}$$

This is an isometric mapping into $\mathbf{L}_2$, since it is obvious that $\mathbf{E}(\theta_h\eta^2) = \mathbf{E}(\eta^2)$. This implies in particular that the definition of θ_h is correct for η of the above form (i.e., for two different representations of η, the corresponding expressions for $\theta_h\eta$ agree with probability one), and then the θ_h can be continuously (and isometrically) extended to the set of all limits in square mean of random variables η of the above form.

For weak sense stationary processes the above construction is not available for all of $\mathbf{L}_2$, but it works on the space $\mathbf{H}_T$. Clearly, for a linear combination

$$\eta = \lambda_1\xi(t_1) + \ldots + \lambda_n\xi(t_n) + c, \tag{18}$$

one lets

$$\theta_h\eta = \lambda_1\xi(t_1+h) + \ldots + \lambda_n\xi(t_n+h) + c. \tag{19}$$

Again, this is an isometric mapping, so it can be extended to all of $\mathbf{H}_T$

3. The Correlation function

The correlation function determines a number of properties of a stationary process $\xi(t)$. Questions of continuity and differentiability, for instance, can be answered, if one defines them using convergence in square mean. In particular, one has the following theorem.

Theorem 1: *A weak sense stationary process* $\xi(.)$ *is continuous in square mean if and only if its correlation function* $R(.)$ *is continuous, which is in turn equivalent to* $R(.)$ *being continuous at* 0.

$\xi(.)$ *is differentiable in square mean if* $R(.)$ *is twice continuously differentiable; the correlation function of* $\xi'(.)$ *is* $\tilde{R}(t) = -R''(t)$.

The first assertion is very simple to prove. Assume without loss of generality that $\mathbf{E}(\xi(t)) = 0$. Then $\mathbf{E}((\xi(s+t) - \xi(s))^2) = 2(R(0) - R(t))$, so the continuity of $R(.)$ at zero implies the continuity (even the uniform continuity) of $\xi(.)$ in square mean. On the other hand, from the continuity of $\xi(.)$ one finds that $R(.)$ is continuous by applying Cauchy's inequality:

$$\begin{aligned}\left|R(t+h)-R(t)\right|^2 &= \left(\mathbf{E}\left(\xi(0)\left(\xi(t+h)-\xi(t)\right)\right)\right)^2 \\ &\le \mathbf{E}\left(\xi(0)^2\right)\mathbf{E}\left(\left(\xi(t+h)-\xi(t)\right)^2\right)\end{aligned} \tag{20}$$

The proof of the second assertion proceeds in a similar manner with just a little more complexity and will be omitted.

One has to keep in mind that continuity in square mean has nothing to do with pathwise continuity; a stationary Markov chain (in continuous time of course) with finitely many states, for example, is continuous in square mean, but its trajectories obviously are far from being continuous.

Another important property of the correlation function is that it is positive definite: for any complex $z_1,....,z_k$ and $t_1,...,t_k \in T$, one has

$$\sum_{i=1}^{k}\sum_{j=1}^{k} z_i \overline{z}_j R\left(t_i - t_j\right) \ge 0. \tag{21}$$

For positive definite functions, the following characterizations are available:

Theorem 2: *The sequence* $R(n), n=-1,0,1,....$ *is positive definite if and only if it can be written as*

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} dv(\lambda), \tag{22}$$

where v is a (unique) finite measure on $[-\pi,\pi]$.

Theorem 3: (*Bochner-Khinchine*) *A continuous function* $R(t), t \in \mathbb{R}$ *is a positive definite if and only if it admits the representation*

$$R(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dv(\lambda), \tag{23}$$

where v is a uniquely determined finite measure on the real line.

The function $F(x) = v((-\infty, x])$ is called the spectral function. If *F* is absolutely continuous with derivative *f*, then *f* is called the spectral density of *F*.

The spectral function *F* can be interpreted as a cumulative distribution function describing the energy distribution of a random oscillation; the spectral density is the density of this energy distribution.

-

-
-

Bibliography

Cramér, H., and Leadbetter, M. (1966) Stationary and Related Stochastic Processes. New York: Wiley [This is one of the most important references on stationary processes].

Karlin, S., and Taylor, H.M. (1974) A First Course in Stochastic Processes. New York: Academic Press [This is an introduction to the theory of stochastic processes, and in particular, stationary processes].

Yaglom, A.M. (1962) An Introduction to the Theory Random Functions. Englewood Cliffs, N.J, USA: Prentice-Hall. [This is another basic reference, giving in particular a good introduction to spectral theory].

Biographical Sketch

Karl Grill received the Ph.D. degree from TU Wien in 1983. Since 1982 he is with TU Wien where he became an Associate Professor in 1988. He was a visiting Professor in the Department of Statistics, University of Arizona during 1991-92. From February to August 1994, he held NSERC Foreign Researcher Award, Carleton University, Ottawa, Canada