Chapter 22
Processes with Finite Second Moments. Gaussian Processes

Abstract The chapter is devoted to the classical "second-order theory" of time-homogeneous processes with finite second moments. Section 22.1 explores the relationships between the covariance function properties and those of the process itself and proves the ergodic theorem (in quadratic mean) for processes with covariance functions vanishing at the infinity. Section 22.2 is devoted to the special case of Gaussian processes, while Sect. 22.3 solves the best linear prediction problem.

22.1 Processes with Finite Second Moments

Let $\{\xi(t), \ -\infty < t < \infty\}$ be a random process for which there exist the moments $a(t) = \mathbf{E}\xi(t)$ and $R(t, u) = \mathbf{E}\xi(t)\xi(u)$. Since it is always possible to study the process $\xi(t) - a(t)$ instead of $\xi(t)$, we can assume without loss of generality that $a(t) \equiv 0$.

Definition 22.1.1 The function $R(t, u)$ is said to be the *covariance function* of the process $\xi(t)$.

Definition 22.1.2 A function $R(t, u)$ is said to be *nonnegative* (*positive*) *definite* if, for any $k; u_1, \ldots, u_k; a_1, \ldots, a_k \neq 0$,

$$\sum_{i,j} a_i a_j R(u_i, u_j) \geq 0 \quad (> 0).$$

It is evident that the covariance function $R(t, u)$ is nonnegative definite, because

$$\sum_{i,j} a_i a_j R(u_i, u_j) = \mathbf{E}\left(\sum_{i,j} a_j \xi(u_i)\right)^2 \geq 0.$$

Definition 22.1.3 A process $\xi(t)$ is said to be *unpredictable* if no linear combination of the variables $\xi(u_1), \ldots, \xi(u_k)$ is zero with probability 1, i.e. if there exist no $u_1, \ldots, u_k; a_1, \ldots, a_k$ such that

$$\mathbf{P}\left(\sum_i a_i \xi(u_i) = 0\right) = 1.$$

If $R(t, u)$ is the covariance function of an unpredictable process, then $R(t, u)$ is positive definite. We will see below that the converse assertion is also true in a certain sense.

Unpredictability means that we cannot represent $\xi(t_k)$ as a linear combination of $\xi(t_j)$, $j < k$.

Example 22.1.1 The process $\xi(t) = \sum_{k=1}^{N} \xi_k g_k(t)$, where $g_k(t)$ are linearly independent and ξ_k are independent, is not unpredictable, because from $\xi(t_1), \ldots, \xi(t_N)$ we can determine the values $\xi(t)$ for all other t.

Consider the Hilbert space L_2 of all random variables η on $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ having finite second moments, $\mathbf{E}\eta = 0$, endowed with the inner product $(\eta_1, \eta_2) = \mathbf{E}\eta_1\eta_2$ corresponding to the distance $\|\eta_1 - \eta_2\| = [\mathbf{E}(\eta_1 - \eta_2)^2]^{1/2}$. Convergence in L_2 is obviously convergence in mean quadratic.

A random process $\xi(t)$ may be thought of as a curve in L_2.

Definition 22.1.4 A random process $\xi(t)$ is said to be *wide sense stationary* if the function $R(t, u) =: R(t - u)$ depends on the difference $t - u$ only. The function $R(s)$ is called nonnegative (positive) definite if the function $R(t, t + s)$ is of the respective type. For brevity, we will often call wide sense stationary processes simply stationary.

For the Wiener process, $R(t, u) = \mathbf{E}w(t)w(u) = \min(t, u)$, so that $w(t)$ cannot be stationary. But the process $\xi(t) = w(t + 1) - w(t)$ will already be stationary.

It is obvious that, for a stationary process, the function $R(s)$ is even and $\mathbf{E}\xi^2(t) = R(0) = \text{const}$. For simplicity's sake, put $R(0) = 1$. Then, by the Cauchy–Bunjakovsky inequality,

$$\left| R(s) \right| = \left| \mathbf{E}\xi(t)\xi(t + s) \right| \leq \left[\mathbf{E}\xi^2(t)\mathbf{E}\xi^2(t + s) \right]^{1/2} = R(0) = 1.$$

Theorem 22.1.1

(1) *A process $\xi(t)$ is continuous in mean quadratic ($\xi(t + \Delta) \xrightarrow{(2)} \xi(t)$ as $\Delta \to 0$) if and only if the function $R(u)$ is continuous at zero.*

(2) *If the function $R(u)$ is continuous at zero, then it is continuous everywhere.*

Proof

(1) $\left\| \xi(t + \Delta) - \xi(t) \right\|^2 = \mathbf{E}\big(\xi(t + \Delta) - \xi(t)\big)^2 = 2R(0) - 2R(\Delta).$

(2) $R(t + \Delta) - R(t) = \mathbf{E}\big(\xi(t + \Delta)\xi(0) - \xi(t)\xi(0)\big)$

$$= \big(\xi(0), \xi(t + \Delta) - \xi(t)\big) \leq \left\| \xi(t + \Delta) - \xi(t) \right\|$$

$$= \sqrt{2\big(R(0) - R(\Delta)\big)}. \qquad\qquad (22.1.1)$$

The theorem is proved.   □

A process $\xi(t)$ continuous in mean quadratic will be stochastically continuous, as we can see from Chaps. 6 and 18. The continuity in mean quadratic does not,

however, imply path-wise continuity. The reader can verify this by considering the example of the process

$$\xi(t) = \eta(t+1) - \eta(t) - 1,$$

where $\eta(t)$ is the Poisson process with parameter 1. For that process, the covariance function

$$R(t) = \begin{cases} 0 & \text{for } t \geq 1, \\ 1 - t & \text{for } 0 \leq t \leq 1 \end{cases}$$

is continuous, although the trajectories of $\xi(t)$ are not. If

$$\left| R(\Delta) - R(0) \right| < c\Delta^{1+\varepsilon} \tag{22.1.2}$$

for some $\varepsilon > 0$ then, by the Kolmogorov theorem (see Theorem 18.2.1), $\xi(t)$ has a continuous modification. From this it follows, in particular, that if $R(t)$ is twice differentiable at the point $t = 0$, then the trajectories of $\xi(t)$ may be assumed continuous. Indeed, in that case, since $R(t)$ is even, one has

$$R'(0) = 0 \quad \text{and} \quad R(\Delta) - R(0) \sim \frac{1}{2} R''(0)\Delta^2.$$

As a whole, the smoother the covariance function is at zero, the smoother the trajectories of $\xi(t)$ are.

Assume that the trajectories of $\xi(t)$ are measurable (for example, belong to the space D).

Theorem 22.1.2 (The simplest ergodic theorem) *If*

$$R(s) \to 0 \quad as \; s \to \infty, \tag{22.1.3}$$

then

$$\zeta_T := \frac{1}{T} \int_0^T \xi(t)\,dt \xrightarrow{(2)} 0.$$

Proof Clearly,

$$\|\zeta_T\|^2 = \frac{1}{T^2} \int_0^T \int_0^T R(t-u)\,dt\,du.$$

Since $R(s)$ is even,

$$J := \int_0^T \int_0^T R(t-u)\,dt\,du = 2 \int_0^T \int_u^T R(t-u)\,dt\,du.$$

Making the orthogonal change of variables $v = (t-u)/\sqrt{2}$, $s = (t+u)/\sqrt{2}$, we obtain

$$J \leq 2 \int_{s=0}^{T/\sqrt{2}} \int_{v=0}^{T/\sqrt{2}} R(v\sqrt{2})\,dv\,ds \leq 2T \int_0^T R(v)\,dv,$$

$$\|\zeta_T\|^2 \leq \frac{2}{T} \int_0^T R(v)\,dv \to 0.$$

The theorem is proved. $\square$

Example 22.1.2 The stationary white noise process $\xi(t)$ is defined as a process with independent values, i.e. a process such that, for any $t_1, \ldots, t_n$, the variables $\xi(t_1), \ldots, \xi(t_n)$ are independent. For such a process,

$$R(t) = \begin{cases} 1 & \text{for } t = 0, \\ 0 & \text{for } t \neq 0, \end{cases}$$

and thus condition (22.1.3) is met. However, one cannot apply Theorem 22.1.2 here, for the trajectories of $\xi(t)$ will be non-measurable with probability 1 (for example, the set $B = \{t : \xi(t) > 0\}$ is non-measurable with probability 1).

Definition 22.1.5 A process $\xi(t)$ is said to be *strict sense stationary* if, for any $t_1, \ldots, t_k$, the distribution of $(\xi(t_1 + u), \xi(t_2 + u), \ldots, \xi(t_k + u))$ is independent of u.

It is obvious that if $\xi(t)$ is a strict sense stationary process then

$$\mathbf{E}\xi(t)\xi(u) = \mathbf{E}\xi(t - u)\xi(0) = R(t - u),$$

and $\xi(t)$ will be wide sense stationary. The converse is, of course, not true. However, there exists a class of processes for which both concepts of stationarity coincide.

22.2 Gaussian Processes

Definition 22.2.1 A process $\xi(t)$ is said to be *Gaussian* if its finite-dimensional distributions are normal.

We again assume that $\mathbf{E}\xi(t) = 0$ and $R(t, u) = \mathbf{E}\xi(t)\xi(u)$.

The finite-dimensional distributions are completely determined by the ch.f.s ($\lambda = (\lambda_1, \ldots, \lambda_k)$, $\xi = (\xi(t_1), \ldots, \xi(t_k))$)

$$\mathbf{E}e^{i(\lambda, \xi)} = \mathbf{E}e^{i \sum_j \lambda_j \xi(t_j)} = e^{-\frac{1}{2}\lambda R \lambda^T},$$

where $R = \|R(t_i, t_j)\|$ and the superscript T stands for transposition, so that

$$\lambda R \lambda^T = \sum_{i,j} \lambda_i \lambda_j R(t_i, t_j).$$

Thus for a Gaussian process the finite-dimensional distributions are completely determined by the covariance function $R(t, u)$.

We saw that for an unpredictable process $\xi(t)$, the function $R(t, u)$ is positive definite. A converse assertion may be stated in the following form.

Theorem 22.2.1 *If the function $R(t, u)$ is positive definite, then there exists an unpredictable Gaussian process with the covariance function $R(t, u)$.*

Proof For arbitrary $t_1, \ldots, t_k$, define the finite-dimensional distribution of the vector $\xi(t_1), \ldots, \xi(t_k)$ via the density

$$p_{t_1, \ldots, t_k}(x_1, \ldots, x_k) = \frac{\sqrt{|A|}}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} x A x^T\right\},$$

where A is the matrix inverse to the covariance matrix $R = \|R(t_i, t_j)\|$ (see Sect. 7.6) and $|A|$ is the determinant of A. These distributions will clearly be consistent, because the covariance matrices are consistent (the matrix for $\xi(t_1), \ldots, \xi(t_{k-1})$ is a submatrix of R). It remains to make use of the Kolmogorov theorem. The theorem is proved. $\square$

Example 22.2.1 Let $w(t)$ be the standard Wiener process. The process

$$w^0(t) = w(t) - tw(1), \quad t \in [0, 1],$$

is called the *Brownian bridge* (its "ends are fixed": $w^0(0) = w^0(1) = 0$). The covariance function of $w^0(t)$ is equal to

$$R(t, u) = \mathbf{E}\big(w(t) - tw(1)\big)\big(w(u) - uw(1)\big) = t(1 - u)$$

for $u \geq t$.

A Gaussian wide sense stationary process $\xi(t)$ is strict sense stationary. This immediately follows from the fact that for $R(t, u) = R(t - u)$ the finite-dimensional distributions of $\xi(t)$ become invariant with respect to time shift:

$$p_{t_1, \ldots, t_k}(x_1, \ldots, x_k) = p_{t_1 + u, \ldots, t_k + u}(x_1, \ldots, x_k)$$

since $\|R(t_i + u, t_j + u)\| = \|R(t_i, t_j)\|$.

If $\xi(t)$ is a Gaussian process, then conditions ensuring the smoothness of its trajectories can be substantially relaxed in comparison with (22.1.2).

Let for simplicity's sake the Gaussian process $\xi(t)$ be stationary.

Theorem 22.2.2 *If, for $h < 1$,*

$$\big|R(h) - R(0)\big| < c\left(\log \frac{1}{h}\right)^{-\alpha}, \quad \alpha > 3, \ c < \infty,$$

then the trajectories of $\xi(t)$ can be assumed continuous.

Proof We make use of Theorem 18.2.2 and put $\varepsilon(h) = (\log \frac{1}{h})^{-\beta}$ for $1 < \beta < (\alpha - 1)/2$ (we take logarithms to the base 2). Then

$$\sum_{n=1}^{\infty} \varepsilon(2^{-n}) = \sum_{n=1}^{\infty} n^{-\beta} < \infty,$$

and, by (22.1.1),

$$\mathbf{P}\big(\big|\xi(t+h)-\xi(t)\big| > \varepsilon(h)\big) = 2\left[1 - \Phi\left(\frac{\varepsilon(h)}{\sqrt{2(1-R(h))}}\right)\right]$$

$$\leq 2\left[1 - \Phi\left(c\varepsilon(h)\left(\log\frac{1}{h}\right)^{\alpha/2}\right)\right] = 2\left[1 - \Phi\left(c\left(\log\frac{1}{h}\right)^{\alpha/2-\beta}\right)\right].$$

(22.2.1)

Since the argument of Φ increases unboundedly as $h \to 0$, $\gamma = \alpha - 2\beta > 1$, and by (19.3.1)

$$1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}x}e^{-x^2/2} \quad \text{as } x \to \infty,$$

we see that the right-hand side of (22.2.1) does not exceed

$$q(h) := c_1\left(\log\frac{1}{h}\right)^{\beta-\alpha/2}\exp\left\{-c_2\left(\log\frac{1}{h}\right)^{\alpha-2\beta}\right\},$$

so that

$$\sum_{n=1}^{\infty} 2^n q\big(2^{-n}\big) = c_1 \sum_{n=1}^{\infty} n^{-\gamma/2}\exp\{-c_2 n^\gamma + n\ln 2\} < \infty,$$

because $c_2 > 0$ and $\gamma > 1$. The conditions of Theorem 18.2.2 are met, and so Theorem 22.2.2 is proved. □

22.3 Prediction Problem

Suppose the distribution of a process $\xi(t)$ is known, and one is given the trajectory of $\xi(t)$ on a set $B \subset (-\infty, t]$, B being either an interval or a finite collection of points. What could be said about the value $\xi(t + u)$? Our aim will be to find a random variable ζ, which is $\mathfrak{F}_B = \sigma(\xi(v), v \in B)$-measurable (and called a *prediction*) and such that $\mathbf{E}(\xi(t+u) - \zeta)^2$ assumes the smallest possible value. The answer to that problem is actually known (see Sect. 4.8):

$$\zeta = \mathbf{E}\big(\xi(t+u)\big|\mathfrak{F}_B\big).$$

Let $\xi(t)$ be a Gaussian process, $B = \{t_1, \dots, t_k\}$, $t_1 < t_2 < \cdots < t_k < t_0 = t + u$, $A = (\sigma^2)^{-1} = \|a_{ij}\|$ and $\sigma^2 = \|\mathbf{E}\xi(t_i)\xi(t_j)\|_{i,j=1,\dots,k,0}$. Then the distribution of the vector $(\xi(t_1), \dots, \xi(t_0))$ has the density

$$f(x_1, \dots, x_k, x_0) = \frac{\sqrt{|A|}}{(2\pi)^{(k+1)/2}}\exp\left\{-\frac{1}{2}\sum_{i,j}x_i x_j a_{ij}\right\},$$

and the conditional distribution of $\xi(t_0)$ given $\xi(t_1), \dots, \xi(t_k)$ has density equal to the ratio

$$\frac{f(x_1, \dots, x_k, x_0)}{\int_{-\infty}^{\infty} f(x_1, \dots, x_k, x_0)\, dx_0}.$$

The exponential part of this ratio has the form

$$\exp\left\{-\frac{a_{00}x_0^2}{2}-\sum_{j=1}^{k}x_0x_ja_{j0}\right\}.$$

This means that the conditional distribution under consideration is the normal law $\boldsymbol{\Phi}_{\alpha,d^2}$, where

$$\alpha=-\sum_{j}\frac{x_ja_{j0}}{a_{00}},\qquad d^2=\frac{1}{a_{00}}.$$

Thus, in our case the best prediction ζ is equal to

$$\zeta=-\sum_{j=1}^{k}\frac{\xi(t_j)a_{0j}}{a_{00}}.$$

The mean quadratic error of this prediction equals $\sqrt{1/a_{00}}$.

We have obtained a *linear prediction*. In the general case, the linearity property is usually violated.

Consider now the problem of the best linear prediction in the case of an arbitrary process $\xi(t)$ with finite second moments. For simplicity's sake we assume again that $B=\{t_1,\ldots,t_k\}$.

Denote by $H(\xi)$ the subspace of L_2 generated by the random variables $\xi(t)$, $-\infty<t<\infty$, and by $H_B(\xi)$ the subspace of $H(\xi)$ generated (or spanned by) $\xi(t_1),\ldots,\xi(t_k)$. Elements of $H_B(\xi)$ have the form

$$\sum_{j=1}^{k}a_j\xi(t_j).$$

The existence and the form of the best linear prediction in this case are established by the following assertion.

Theorem 22.3.1 *There exists a unique point $\zeta\in H_B(\xi)$ (the projection of $\xi(t+u)$ onto $H_B(\xi)$, see Fig. 22.1) such that*

$$\xi(t+u)-\zeta\perp H_B(\xi). \tag{22.3.1}$$

Relation (22.3.1) is equivalent to

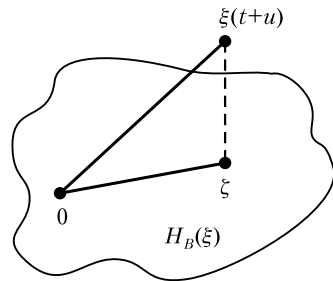$$\left\|\xi(t+u)-\zeta\right\|=\min_{\theta\in H_B(\xi)}\left\|\xi(t+u)-\theta\right\|. \tag{22.3.2}$$

Explicit formulas for the coefficients a_j in the representation $\zeta=\sum a_j\xi(t_j)$ are given in the proof.

Proof Relation (22.3.1) is equivalent to the equations

$$\bigl(\xi(t+u)-\zeta,\xi(t_j)\bigr)=0,\quad j=1,\ldots,k.$$

Fig. 22.1 Illustration to
Theorem 22.3.1: the point ζ
is the projection of $\xi(t + u)$
onto $H_B(\xi)$

Substituting here

$$\zeta = \sum_{l=1}^{k} a_l \xi(t_l) \in H_B(\xi),$$

we obtain

$$R(t + u, t_j) = \sum_{l=1}^{k} a_l R(t_j, t_l), \quad j = 1, \ldots, k, \tag{22.3.3}$$

or, in vector form, $R_{t+u} = aR$, where

$$a = (a_1, \ldots, a_k),$$
$$R_{t+u} = \big(R(t + u, t_1), \ldots, R(t + u, t_k)\big), \quad R = \big\| R(t_i, t_j) \big\|.$$

If the process $\xi(t)$ is unpredictable, then the matrix R is non-degenerate and
Eq. (22.3.3) has a unique solution:

$$a = R_{t+u} R^{-1}. \tag{22.3.4}$$

If $\xi(t)$ is not unpredictable, then either R^{-1} still exists and then (22.3.4) holds, or
R is degenerate. In that case, one has to choose from the collection $\xi(t_1), \ldots, \xi(t_k)$
only $l < k$ linearly independent elements for which all the above remains true after
replacing k with l.

The equivalence of (22.3.1) and (22.3.2) follows from the following considera-
tions. Let θ be any other element of $H_B(\xi)$. Then

$$\eta := \theta - \zeta \in H_B(\xi), \qquad \eta \perp \xi(t + u) - \zeta,$$

so that

$$\big\| \xi(t + u) - \theta \big\| = \big\| \xi(t + u) - \zeta \big\| + \|\eta\| \geq \big\| \xi(t + u) - \zeta \big\|.$$

The theorem is proved.   □

Remark 22.3.1 It can happen (in the case where the process $\xi(t)$ is not unpre-
dictable) that $\xi(t + u) \in H_B(\xi)$. Then the error of the prediction ζ will be equal
to zero.