

# LECTURES ON STATIONARY STOCHASTIC PROCESSES

A COURSE FOR PHD STUDENTS IN  
MATHEMATICAL STATISTICS AND OTHER FIELDS

GEORG LINDGREN

October 2006



LUND UNIVERSITY

Faculty of Engineering  
Centre for Mathematical Sciences  
Mathematical Statistics



# Contents

<b>Foreword</b>	<b>vii</b>
<b>1 Some probability and process background</b>	<b>1</b>
1.1 Probability measures on sample spaces . . . . .	1
1.2 Events, probabilities and random variables . . . . .	3
1.2.1 Events and families of events . . . . .	3
1.2.2 Probabilities . . . . .	4
1.2.3 Random variables and random sequences . . . . .	5
1.2.4 Conditional expectation . . . . .	7
1.3 Stochastic processes . . . . .	7
1.3.1 Stochastic processes and finite-dimensional distributions . . . . .	7
1.3.2 The distribution of a random sequence . . . . .	9
1.3.3 The continuous parameter case . . . . .	10
1.4 Stationary processes and fields . . . . .	13
1.4.1 Stationary processes . . . . .	13
1.4.2 Random fields . . . . .	14
1.5 Gaussian processes . . . . .	15
1.5.1 Multivariate normal distributions and Gaussian processes . . . . .	15
1.5.2 Linear prediction and reconstruction . . . . .	18
1.5.3 Some useful inequalities . . . . .	18
1.6 Some historical landmarks . . . . .	19
1.6.1 Brownian motion and the Wiener process . . . . .	19
1.6.2 Rice and electronic noise . . . . .	21
1.6.3 Gaussian random wave models . . . . .	22
1.6.4 Detection theory and statistical inference . . . . .	24
Exercises . . . . .	25
<b>2 Stochastic analysis</b>	<b>27</b>
2.1 Quadratic mean properties . . . . .	27
2.2 Sample function continuity . . . . .	28
2.2.1 Countable and uncountable events . . . . .	28
2.2.2 Conditions for sample function continuity . . . . .	30
2.2.3 Probability measures on $\mathbb{C}[0, 1]$ . . . . .	37
2.3 Derivatives, tangents, and other characteristics . . . . .	37

2.3.1	Differentiability . . . . .	37
2.3.2	Jump discontinuities and Hölder conditions . . . . .	40
2.4	Quadratic mean properties a second time . . . . .	45
2.4.1	Quadratic mean continuity . . . . .	45
2.4.2	Quadratic mean differentiability . . . . .	46
2.4.3	Higher order derivatives and their correlations . . . . .	48
2.5	Summary of smoothness conditions . . . . .	49
2.6	Stochastic integration . . . . .	49
2.7	An ergodic result . . . . .	52
	Exercises . . . . .	55
<b>3</b>	<b>Crossings</b>	<b>57</b>
3.1	Level crossings and Rice's formula . . . . .	57
3.1.1	Level crossings . . . . .	57
3.1.2	Rice's formula for absolutely continuous processes . . . . .	58
3.1.3	Alternative proof of Rice's formula . . . . .	60
3.1.4	Rice's formula for differentiable Gaussian processes . . . . .	62
3.2	Prediction from a random crossing time . . . . .	63
3.2.1	Prediction from upcrossings . . . . .	64
3.2.2	The Slepian model . . . . .	67
3.2.3	Excursions and related distributions . . . . .	73
	Exercises . . . . .	78
<b>4</b>	<b>Spectral- and other representations</b>	<b>79</b>
4.1	Complex processes and their covariance functions . . . . .	79
4.1.1	Stationary processes . . . . .	79
4.1.2	Non-negative definite functions . . . . .	80
4.1.3	Strict and weak stationarity . . . . .	82
4.2	Bochner's theorem and the spectral distribution . . . . .	82
4.2.1	The spectral distribution . . . . .	82
4.2.2	Properties of the spectral distribution . . . . .	86
4.2.3	Spectrum for stationary sequences . . . . .	88
4.3	Spectral representation of a stationary process . . . . .	89
4.3.1	The spectral process . . . . .	89
4.3.2	The spectral theorem . . . . .	90
4.3.3	More on the spectral representation . . . . .	93
4.3.4	Spectral representation of stationary sequences . . . . .	100
4.4	Linear filters . . . . .	100
4.4.1	Projection and the linear prediction problem . . . . .	100
4.4.2	Linear filters and the spectral representation . . . . .	101
4.4.3	Linear filters and differential equations . . . . .	106
4.4.4	White noise in linear systems . . . . .	112
4.4.5	The Hilbert transform and the envelope . . . . .	115
4.4.6	The sampling theorem . . . . .	119

4.5	Karhunen-Loève expansion . . . . .	121
4.5.1	Principal components . . . . .	121
4.5.2	Expansion of a stationary process along eigenfunctions . .	122
4.5.3	The Karhunen-Loève theorem . . . . .	124
	Exercises . . . . .	130
<b>5</b>	<b>Ergodic theory and mixing</b>	<b>133</b>
5.1	The basic Ergodic theorem in $L^2$ . . . . .	133
5.2	Stationarity and transformations . . . . .	134
5.2.1	Pseudo randomness and transformation of sample space .	134
5.2.2	Strict stationarity and measure preserving transformations	135
5.3	The Ergodic theorem, transformation view . . . . .	137
5.3.1	Invariant sets and invariant random variables . . . . .	138
5.3.2	Ergodicity . . . . .	139
5.3.3	The Birkhoff Ergodic theorem . . . . .	141
5.4	The Ergodic theorem, process view . . . . .	144
5.5	Ergodic Gaussian sequences and processes . . . . .	148
5.6	Mixing and asymptotic independence . . . . .	150
5.6.1	Singularity and regularity . . . . .	150
5.6.2	Asymptotic independence, regularity and singularity . . .	152
5.6.3	Uniform, strong, and weak mixing . . . . .	156
	Exercises . . . . .	159
<b>6</b>	<b>Vector processes and random fields</b>	<b>161</b>
6.1	Cross-spectrum and spectral representation . . . . .	161
6.1.1	Spectral distribution . . . . .	162
6.1.2	Spectral representation of $\mathbf{x}(t)$ . . . . .	163
6.2	Some random field theory . . . . .	165
6.2.1	Homogeneous fields . . . . .	166
6.2.2	Isotropic fields . . . . .	167
6.2.3	Randomly moving surfaces . . . . .	169
6.2.4	Stochastic water waves . . . . .	170
	Exercises . . . . .	173
<b>A</b>	<b>The axioms of probability</b>	<b>175</b>
A.1	The axioms . . . . .	175
A.2	Extension of a probability from field to $\sigma$ -field . . . . .	176
A.3	Kolmogorov's extension to $\mathbb{R}^\infty$ . . . . .	178
	Exercises . . . . .	179
<b>B</b>	<b>Stochastic convergence</b>	<b>181</b>
B.1	Criteria for convergence almost surely . . . . .	181
B.2	Criteria for convergence in quadratic mean . . . . .	183
B.3	Criteria for convergence in probability . . . . .	184
	Exercises . . . . .	184

---

<b>C Hilbert space and random variables</b>	<b>187</b>
C.1 Hilbert space and scalar products . . . . .	187
C.2 Projections in Hilbert space . . . . .	189
C.3 Stochastic processes and Hilbert spaces . . . . .	190
<b>D Spectral simulation of random processes</b>	<b>193</b>
D.1 The Fast Fourier Transform, FFT . . . . .	193
D.2 Random phase and amplitude . . . . .	194
D.3 Aliasing . . . . .	194
D.4 Simulation scheme . . . . .	195
D.5 Difficulties and details . . . . .	196
D.6 Simulation of the envelope . . . . .	197
D.7 Summary . . . . .	197
<b>Literature</b>	<b>199</b>
<b>Index</b>	<b>202</b>





# Foreword

The book *Stationary and Related Stochastic Processes* [9] appeared in 1967. Written by Harald Cramér and M.R. Leadbetter, it drastically changed the life of PhD students in Mathematical statistics with an interest in stochastic processes and their applications, as well as that of students in many other fields of science and engineering. By that book, they got access to tools and results for stationary stochastic processes that until then had been available only in rather advanced mathematical textbooks, or through specialized statistical journals. The impact of the book can be judged from the fact that still in 1999, after more than thirty years, it is a standard reference to stationary processes in PhD theses and research articles.

Unfortunately, the book only appeared in a first edition and it is since long out of print. Even if many of the more specialized results in the book now have been superseded by more general results, and simpler proofs have been found for some of the statements, the general attitude in the book makes it enjoyable reading both for the student and for the teacher. It will remain a definite source of reference for many standard results on sample function and crossings properties of continuous time processes, in particular in the Gaussian case.

These lecture notes are the results of a series of PhD courses on *Stationary stochastic processes* which have been held at the Department of Mathematical Statistics, Lund University, during a sequence of years, all based on and inspired by the book by Cramér and Leadbetter. The aim of the notes is to provide a reasonably condensed presentation of sample function properties, limit theorems, and representation theorems for stationary processes, in the spirit of [9]. It must be said, however, that it represents only a selection of the material, and the reader who has found interest in the present course, should take the time to read the original.

Even if the Cramér and Leadbetter book is the basic source of inspiration, other texts have influenced these notes. The most important of these is the now reprinted book on *Probability* [5] by Leo Breiman. The Ergodic chapter is a mixture of the two approaches. The Karhunen-Loève expansion follows the book by Wong [35]. Finally, the classical memoirs by S.O. Rice [27], have also been a source of inspiration.

Some knowledge of the mathematical foundations of probability helps while

reading the text; I have included most of it in Appendices on the probability axioms, together with the existence and basic convergence properties, as well as some Hilbert space concepts. There is also an Appendix on how to simulate stationary stochastic processes by spectral methods and the FFT algorithm.

I am grateful to the PhD-students Rikard Berthilsson, Jonas Brunskog, Halfdan Grage, Peter Gustafsson, Pär Johannesson, Finn Lindgren, Karl-Ola Lundberg, Dan Mattsson, Tord Rikte, Jesper Rydén, Martin Sköld, Martin Svensson, Magnus Wiktorsson, in the 1998/99 course for many detailed comments on the text and pertinent questions during the lectures, which hopefully helped clarify some of the obscurities. They also helped to remove many of the misprints.

Lund in May, 1999

Georg Lindgren

### **Printing September 2002 and April 2004**

These printings of the Lecture notes differ considerably from the printing of May 1999. Many misprints have been corrected and there are also many additions, mostly of previously left out details, but there is also some new aspects. I am grateful to comments by Bengt Ringnér, which helped to remove unclear statements and errors, the PhD students at the Lund 2000/01 course, Anastasia Baxevanni, Torgny Lindström, Ulla Machado, Anders Malmberg, Sebastian Rasmus, Mikael Signahl, and to Lars Holst and PhD student Henrik Hult at KTH who used the material and made several suggestions.

Some new references have been added, in particular some standard textbooks on ergodic theory (K. Petersen: *Ergodic Theory*), real analysis (H.L. Royden: *Real Analysis*), and the new probability book, *Weighting the Odds* by D. Williams [25, 29, 37].

In the 2004 printing several more changes were made, as the result of numerous valuable comment by Oskar Hagberg and Linda Werner.

Lund in September, 2002 and April 2004

Georg Lindgren

### **Printing October 2006**

In this printing, several changes and additions have been made. I have included some more elementary facts, together with historical and general examples in Chapter 1, and expanded the section on Linear filters in Chapter 4 in order not to rely too much on specific Lund courses.

To provide some background to the theory I have, in Chapter 1, highlighted four remarkable research achievements that have helped to shape the theory of stationary processes in general. The examples are specialized and based on real

demands, namely Albert Einsteins derivation of the physics and mathematics behind the Brownian motion from 1905, Steve Rice 1944-45 broad invitation to stochastic Gaussian noise, so important in communication theory, and the equally important introduction of stochastic thinking in naval engineering by StDenis and Pierson from 1953, and finally the link between stochastic processes theory and statistical inference, by Ulf Grenander from 1950, which forms the basis for present day signal processing.

The old Chapter 3, on prediction, has been transformed into a chapter on crossing-related problems, including the form and use of the Slepian model process. What remains of prediction has been moved to the chapter on Ergodic theory. The section on random fields has been slightly expanded. The notation  $\lambda$  for frequency has been changed to the standard  $\omega$ .

Besides the additions, several errors and misprints have been corrected efter comments from the PhD students in the 2004 course, Klas Bogsjö, Johan Lindström, and Sofia Åberg. Timo Koski, who used the material in Linköping, has also given me several suggestions and pointed at unclear points and misprints.

Last, but not least, the Cramér & Leadbetter book has finally reappeared, reprinted by Dover Publications, 2004.

Lund in October, 2006

Georg Lindgren



# Chapter 1

## Some probability and process background

This introductory chapter gives a brief summary of the probability theory needed for the study of stochastic processes with discrete or continuous time. It concentrates on the finite-dimensional distribution functions, which uniquely define probabilities for a sufficiently rich family of events, namely events that can be identified through process values at discrete sets of times. In particular, they allow us to find conditions for sample function continuity, at least if we restrict ourselves to a study of the process at a dense discrete time set.

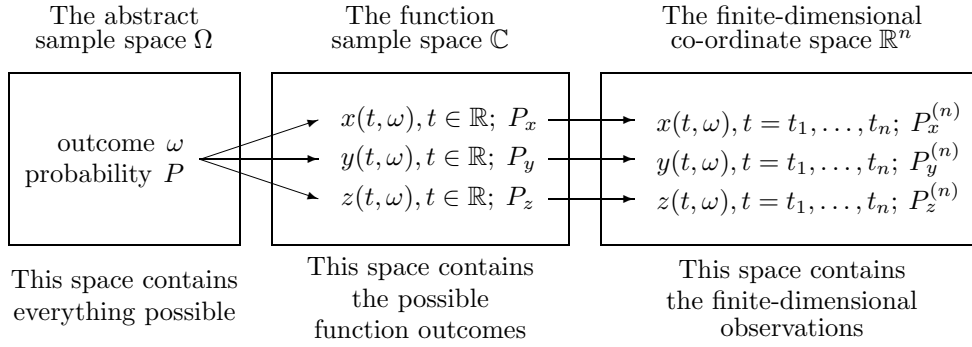
### 1.1 Probability measures on sample spaces

Stochastic processes are often called random functions; these two notions put emphasis on two different aspects of the theory, namely

- stochastic processes as families of infinitely many random variables on the same sample space, usually equipped with a fixed probability measure,
- stochastic processes as a means to assign probabilities to sets of functions, for example some specified sets of continuous functions, or sets of piecewise constant functions with unit jumps.

These two aspects of stochastic processes can be illustrated as in Figure 1.1, corresponding to an experiment where the outcomes are continuous functions.

The figure illustrates the three levels of abstraction and observability for a random experiment. To be concrete, think of an experiment controlling the steering of a ship. The general sample space  $\Omega$  is an abstract set that contains all the possible outcomes of the experiment that can conceivably happen – and it may contain more. A probability measure  $P$  is defined on  $\Omega$  that assigns probabilities to all interesting subsets – we need only one single probability measure to describe our whole world.



**Figure 1.1:** Overview of the three types of worlds in which our processes live.

During the experiment one can record the time evolution of a number of things, such as rudder angle, which we call  $\{x(t), t \in \mathbb{R}\}$ , ship head angle, called  $\{y(t), t \in \mathbb{R}\}$ , and roll angle  $\{z(t), t \in \mathbb{R}\}$ . Each observed function is an observation of a continuous random process. In the figure, the randomness is indicated by the dependence of the experiment outcome  $\omega$ . The distributions of the different processes are  $P_x, P_y, P_z$  – we need one probability measure for each of the phenomena we have chosen to observe.<sup>1</sup>

In practice, the continuous functions are sampled in discrete time steps,  $t = t_1, \dots, t_n$ , resulting in a finite-dimensional observation vector,  $(x_1, \dots, x_n)$ , with an  $n$ -dimensional distribution,  $P_x^{(n)}$ , etc. This is illustrated in the third box in the figure.

Since we do not always want to specify a finite value for  $n$ , the natural mathematical model for the practical situation is to replace the middle box, the sample space  $\mathbb{C}$  of continuous functions, by the sample space  $\mathbb{R}^\infty$  of infinite sequences of real numbers  $(x_0, x_1, \dots)$ . This is close, as we shall see later, really very close, to the finite-dimensional space  $\mathbb{R}^n$ , and mathematically not much more complicated.

**WARNING:** Taking the set  $\mathbb{C}$  of continuous functions as a sample space and assigning probabilities  $P_x$ , etc, on it, is not as innocent as it may sound from the description above. Chapter 2 deals with conditions that guarantee that a stochastic process is continuous, i.e. has continuous sample functions. In fact, these conditions are all on the finite-dimensional distributions.

**SUMMARY:** The abstract sample space  $\Omega$  contains everything that can conceivably happen and is therefore very complex and detailed. Each outcome  $\omega \in \Omega$  is unique, and we need only one comprehensive probability measure  $P$  to describe every outcome of experiment we can do. An experiment is a way to “observe the world”.

<sup>1</sup>The symbol  $\omega$  is here used to represent then elementary experimental outcome, a practice that is standard in probability theory. In most part of this book,  $\omega$  will stand for (angular) frequency; no confusion should arise from this.

The function (sequence) sample space  $\mathbb{C}$  ( $\mathbb{R}^\infty$ ) is simple. It can be used as sample space for a specified experiment for which the result is a function or sequence of numbers. We have to define a unique probability measure for each experiment.

## 1.2 Events, probabilities and random variables

### 1.2.1 Events and families of events

A probability measure  $P$  assigns probabilities to certain events, i.e. subsets, in the sample space  $\Omega$ , in such a way that Kolmogorov's probability axioms are satisfied.<sup>2</sup> Thus, if a subset  $A$  has a probability, then also its complement  $A^*$  has a probability, and  $P(A^*) = 1 - P(A)$ , and further, if  $A$  is disjoint with  $B$ , i.e.  $A \cap B = \emptyset$ , and  $B$  has probability  $P(B)$ , then also  $A \cup B$  has a probability, and  $P(A \cup B) = P(A) + P(B)$ . These requirements lead to the conclusion that probabilities have to be defined at least on a certain minimal family of subsets of  $\Omega$ .

**Definition 1:1** A family of subsets  $\mathcal{F}_0$  to an arbitrary space  $\Omega$  is called a *field* if it contains the whole set  $\Omega$  and is closed<sup>3</sup> under the set operations complement,  $A^*$ , union,  $A \cup B$ , and intersection,  $A \cap B$ . It then also contains all unions of finitely many sets  $A_1, \dots, A_n$  in  $\mathcal{F}_0$ . A field  $\mathcal{F}$  of subsets is called a  $\sigma$ -field if it contains all countable unions and intersections of its sets. The terms *algebra* and  $\sigma$ -algebra are also used, instead of field and  $\sigma$ -field.

To every collection  $\mathcal{A}$  of subsets of  $\Omega$  there is always a unique *smallest* field  $\mathcal{F}_0$  that contains all the sets in  $\mathcal{A}$ . Similarly, there exists a (unique) smallest  $\sigma$ -field  $\mathcal{F}$  that contains all  $\mathcal{A}$ -sets. That  $\sigma$ -field  $\mathcal{F}$  is said to be *generated by*  $\mathcal{A}$ , and it is denoted  $\mathcal{F} = \sigma(\mathcal{A})$ .

**Example 1:1 (Fields and  $\sigma$ -fields in  $\mathbb{R}$ )** The simplest useful field  $\mathcal{F}_0$  of subsets of the real line  $\mathbb{R}$  consists of all finite half-open intervals,  $a < x \leq b$ , together with unions of a finite number of such intervals. In order for  $\mathcal{F}_0$  to be a field, it is required that it also contains the complement of such unions. To anticipate the introduction of probabilities and random variables, we can remark here that  $\mathcal{F}_0$  is a natural family of sets, since distribution functions can be used to assign probabilities to intervals.

The *smallest*  $\sigma$ -field that contains all half-open intervals is called the *Borel field* in  $\mathbb{R}$ . It is denoted  $\mathcal{B}$ , and its sets are called *Borel sets*.

---

<sup>2</sup>See Appendix A.

<sup>3</sup>That is, if a set  $A$  is in the family  $\mathcal{F}_0$ , then also the complement  $A^*$  belongs to  $\mathcal{F}_0$ , etc.

**Example 1:2 (Fields and  $\sigma$ -fields in  $\mathbb{R}^n$ )** Intervals in  $\mathbb{R}$  correspond to  $n$ -dimensional rectangles in  $\mathbb{R}^n$ , and the smallest interesting field in  $\mathbb{R}^n$  consists of unions of finitely many half-open rectangles, with sides  $(a_i, b_i]$ ,  $i = 1, \dots, n$ , and the complements of such unions. As in  $\mathbb{R}$ , the  $\sigma$ -field generated by  $\mathcal{F}_0$  is called the Borel field in  $\mathbb{R}^n$ . It is denoted by  $\mathcal{B}_n$  and its sets are the Borel sets in  $\mathbb{R}^n$ .

One could note here that it is possible to start with more general "rectangles", where the "sides" are real Borel sets instead of intervals, i.e. sets  $B_1 \times B_2 \times \dots \times B_n$ , where the  $B_j$  are one-dimensional Borel sets. However, even if these generalized rectangles form a richer class than the simple rectangles, the smallest  $\sigma$ -field that contains all such generalized rectangles is exactly equal to  $\mathcal{B}_n$ .

### 1.2.2 Probabilities

Probabilities are defined for events, i.e. subsets of a sample space  $\Omega$ . By a probability measure is meant any function  $P$  defined for every event in a field  $\mathcal{F}_0$ , such that

$$0 \leq P(A) \leq 1, \quad P(\emptyset) = 0, \quad P(\Omega) = 1,$$

and such that, first of all, for any finite number of disjoint events  $A_1, \dots, A_n$  in  $\mathcal{F}_0$  one has

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n). \quad (1.1)$$

That is, probabilities are *finitely additive*. In order to deal with limiting events and the infinity, they are also required to be *countably additive*, i.e. equation (1.1) holds for infinitely many disjoint events, i.e. it holds with  $n = \infty$ ,

$$P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

for all disjoint events  $A_k \in \mathcal{F}_0$  such that  $\cup_1^{\infty} A_k \in \mathcal{F}_0$ .

As remarked, it is easy to assign probabilities to intervals, and unions of intervals, simply by taking

$$P((a, b]) = F(b) - F(a),$$

for some distribution function  $F$ . By additivity and the property of fields, one then also assigns probability to the field  $\mathcal{F}_0$  of finite unions of intervals.

A natural question to ask is, whether this also produces probabilities to the events in the  $\sigma$ -field  $\mathcal{F}$  generated by  $\mathcal{F}_0$ . In fact, it does, and that in a unique way:

**Extension of probability measures:** Every probability measure  $P$ , defined and countably additive on a field  $\mathcal{F}_0$ , can be extended to be defined for every event in the  $\sigma$ -field  $\mathcal{F}$  generated by  $\mathcal{F}_0$ . This can be done in one way only. This means that a probability measure on the real Borel sets is uniquely determined by its values on the half-open intervals, i.e. it depends only of the values of the function

$$F(x) = P((-\infty, x]), \quad F(b) - F(a) = P((a, b]).$$

Probabilities on the Borel sets in  $\mathbb{R}^n$  are similarly uniquely determined by the  $n$ -dimensional distribution function

$$F(x_1, \dots, x_n) = P((-\infty, x_1] \times \dots \times (-\infty, x_n]). \quad (1.2)$$

For example, for a bivariate variable  $(x, x_2)$ ,

$$P((a_1, b_1] \times (a_2, b_2]) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)).$$

The probability measure  $P$  is defined on the *measurable space*  $(\Omega, \mathcal{F})$ , and sets in the  $\sigma$ -field  $\mathcal{F}$  are called the *measurable sets*. For a proof of the existence of a unique extension, see Appendix A.

**Remark 1:1** *The completion of a probability measure  $P$  is obtained as follows. Suppose  $P$  is defined on  $(\Omega, \mathcal{F})$ , i.e. it assigns a probability  $P(A)$  to every  $A$  in  $\mathcal{F}$ . Now, if there is an event  $B$  with  $P(B) = 0$ , then it seems natural to assign probability 0 to any smaller set  $B' \subset B$ . Unfortunately, subsets of measurable sets, are not necessarily measurable, so one can not immediately conclude that  $P(B') = 0$ . However, no other choice is possible, and it is also easy to create the  $\sigma$ -field that also contains all subsets of  $\mathcal{F}$ -sets  $B$  with  $P(B) = 0$ . The extended probability measure is called a complete probability measure.*

### 1.2.3 Random variables and random sequences

#### 1.2.3.1 A random variable and its distribution

A random variable is just a real-valued function  $x(\omega), \omega \in \Omega$ , on a probability space  $(\Omega, \mathcal{F}, P)$ , such that it is possible to talk about its distribution, i.e. the probability

$$P(x \leq a) = P(\{\omega; x(\omega) \leq a\})$$

is defined for all real  $a$ . This means that the set (event)

$$A_a = x^{-1}((-\infty, a]) = \{\omega; x(\omega) \leq a\}$$

is a member of the family  $\mathcal{F}$ , for all  $a \in \mathbb{R}$ . This is equivalent to the seemingly more general statement that

$$x^{-1}(B) \in \mathcal{F} \quad \text{for all Borel sets } B \in \mathcal{B}, \quad (1.3)$$

and of course it holds that

$$P(x^{-1}(B)) = \text{Prob}(x \in B).$$

The requirement (1.3) is the formal definition of a random variable: a random variable is a *Borel measurable function*.

If  $x$  is a random variable on  $(\Omega, \mathcal{F}, P)$ , then we write  $P_x$  for the probability measure on  $(\mathbb{R}, \mathcal{B})$  that is defined by

$$P_x(B) = P(x^{-1}(B)).$$

It is possible to define several random variables  $x_1, x_2, \dots, x_n$  on the same probability space  $(\Omega, \mathcal{F}, P)$ , and there is no difficulty to let  $n = \infty$ . In that case we call the sequence  $\{x_n\}_{n=1}^{\infty}$  a *stochastic process with discrete time*, or a *random sequence*.

### 1.2.3.2 The $\sigma$ -field generated by random variables

When  $x$  is a random variable on  $(\Omega, \mathcal{F}, P)$ , the  $\omega$ -set  $\{\omega \in \Omega; x(\omega) \leq a\}$  belongs to  $\mathcal{F}$  and hence, it has a probability  $\text{Prob}(x \leq a)$ . Furthermore, all sets of the type  $x^{-1}(B)$ , where  $B$  is a Borel set, belong to  $\mathcal{F}$ . In fact, the family of such  $\Omega$ -sets is a  $\sigma$ -field, and it is denoted  $\mathcal{F}(x)$  or  $\sigma(x)$ . It is obvious that  $\mathcal{F}(x) \subset \mathcal{F}$  and we already know that  $P$  assigns a probability to these sets. If  $x$  were the only random variable of interest to us, we could have worked on the probability space  $(\Omega, \mathcal{F}(x), P)$ . The reason for using a general, usually larger  $\sigma$ -field  $\mathcal{F}$ , is that it allows us perfect freedom to include any further random variable without changing neither the  $\sigma$ -field, nor the probability measure.

Another characterization of  $\mathcal{F}(x)$  is that it is the smallest  $\sigma$ -field on  $\Omega$  that makes the function  $x$  measurable, i.e. a random variable. The  $\sigma$ -field  $\mathcal{F}(x)$  is called the  *$\sigma$ -field generated by the random variable  $x$* .

When there are several random variables  $x_1, \dots, x_n$ ,  $n \leq \infty$ , there will be a *smallest  $\sigma$ -field*, denoted  $\mathcal{F}(x_1, \dots, x_n)$ , that contains all the sub- $\sigma$ -fields  $\mathcal{F}(x_j)$ . It is the smallest  $\sigma$ -field that makes all the  $x_j$  random variables.

**Remark 1:2** *When we have the  $\sigma$ -field generated by a random variable  $x$  we have got our first opportunity to really construct a probability measure, in the sense that we can define the values of  $P(A)$  for certain events  $A \in \mathcal{F}$ . If  $F$  is a distribution function<sup>4</sup> on  $\mathbb{R}$  and  $x$  is a Borel measurable function, i.e. a random variable, then*

$$P(A_a) = P(\{\omega \in \Omega; x(\omega) \leq a\}) = F(a)$$

*defines probabilities on the sub-class of events  $A_a$ , and that can be extended to a probability measure on the  $\sigma$ -field  $\mathcal{F}(x)$ .*

<sup>4</sup>i.e.  $F$  is non-decreasing, right-continuous, with  $0 \leq F(x) \leq 1$ , and  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$ .

### 1.2.4 Conditional expectation

Here we will give an elementary definition of the important concept *conditional expectation*. A more general definition will be introduced in Section 5.6.1, but here the simple definition is sufficient.

If  $x, y$  are two random variables, where  $y$  may be multivariate, with joint density  $f(x, y)$ , and with marginal  $y$ -density  $f(y) = \int_u f(u, y) du$ , the conditional expectation of  $x$  given  $y = v$ , is a random variable defined as a function of  $y$ , for  $\omega$  such that  $y(\omega) = v$ , and  $f(v) \neq 0$ , as

$$\varphi(v) = E(x | y = v) = \int_u u \frac{f(u, v)}{f(v)} du. \quad (1.4)$$

For outcomes such that  $f(y) = 0$ ,  $\varphi(y)$  can be defined arbitrarily. We write  $E(x | y) = \varphi(y)$ .

It satisfies

$$E(x) = E(\varphi(y)) = E(E(x | y)) = \int_y \varphi(y) f(y) dy, \quad (1.5)$$

$$V(x) = E(V(x | y)) + V(E(x | y)), \quad (1.6)$$

where  $V(x | y) = \int_x (x - \varphi(y))^2 f(x | y) dx$ . The reader should show this, and the following important theorem:

**Theorem 1:1** *The best predictor of  $x$  given  $y$  in least squares sense is given by  $\varphi(y)$ , i.e.*

$$E((x - \varphi(y))^2) \leq E((x - \psi(y))^2)$$

for every function  $\psi(y)$ .

## 1.3 Stochastic processes

### 1.3.1 Stochastic processes and finite-dimensional distributions

We are now ready to define stochastic processes in general. Remember that we have already defined infinite sequences of random variables,  $\mathbf{y} = \{x_n\}_{n=1}^{\infty}$ , defined on the same probability space  $(\Omega, \mathcal{F}, P)$ . Here, each  $x_j$  is a real-valued function on  $\Omega$ , i.e.  $x_j(\omega) \in \mathbb{R}$ .

There is no further difficulty in considering more than countably many random variables at the same time, and letting  $t$  denote a general *parameter* taking values in a parameter space  $T$ . Thus we can consider a family of functions,  $\{x(t, \omega) \in \mathbb{R}\}_{t \in T}$ , where each  $x(t) = x(t, \cdot)$  is a random variable, i.e. a measurable function from  $\Omega$  to  $\mathbb{R}$ . Hence it has a distribution with a distribution function on  $\mathbb{R}$ , which we denote  $F(\cdot; t)$ , i.e.

$$F(a; t) = \text{Prob}(x(t) \leq a).$$

Taking several variables, at times  $t_1, \dots, t_n$ , one gets an  $n$ -variate random variable

$$(x(t_1), \dots, x(t_n))$$

with an  $n$ -variate distribution in  $\mathbb{R}^n$ ,

$$F(a_1, \dots, a_n; t_1, \dots, t_n) = \text{Prob}(x(t_1) \leq a_1, \dots, x(t_n) \leq a_n).$$

We write  $F_{\mathbf{t}^n}$  for the  $n$ -dimensional distribution function of any vector

$$(x(t_1), \dots, x(t_n)).$$

We summarize the terminology in a formal, but simple, definition.

**Definition 1:2** *Let  $T$  be a parameter set. A stochastic process  $\{x(t)\}_{t \in T}$  indexed by the parameter  $t \in T$  is a family of random variables  $x(t)$  defined on one and the same probability space  $(\Omega, \mathcal{F}, P)$ . In other words, a stochastic process is a function*

$$T \times \Omega \ni (t, \omega) \mapsto x(t, \omega) \in \mathbb{R},$$

such that for fixed  $t = t_0$ ,  $x(t_0, \cdot)$  is a random variable, i.e. a Borel measurable function,  $\Omega \ni \omega \mapsto x(t_0, \omega) \in \mathbb{R}$ , and for fixed  $\omega = \omega_0$ ,  $x(\cdot, \omega_0)$  is a function  $T \ni t \mapsto x(t, \omega_0) \in \mathbb{R}$ .

The family  $\{F_{\mathbf{t}^n}\}_{n=1}^{\infty}$  of finite-dimensional distributions is the family of distribution functions

$$F(a_1, \dots, a_n; t_1, \dots, t_n) = \text{Prob}(x_1 \leq a_1, \dots, x_n \leq a_n); n = 1, 2, \dots; t_j \in T.$$

The finite-dimensional distributions in  $\{F_{\mathbf{t}^n}\}_{n=1}^{\infty}$  of a stochastic process satisfy some trivial conditions to make sure they are *consistent* with each other, of the type

$$\begin{aligned} F(a_1, a_2; t_1, t_2) &= F(a_2, a_1; t_2, t_1) \\ F(a_1, \infty; t_1, t_2) &= F(a_1, t_1). \end{aligned}$$

By this definition we have the following concepts at our disposal in the three scenes from Section 1.1:

sample space	events	probability
abstract space: $\Omega$	$\sigma$ -field $\mathcal{F}$	$P$
continuous functions: $\mathcal{C}$	???	???
real sequences: $\mathbb{R}^{\infty}$	???	???
real vectors: $\mathbb{R}^n$	Borel sets: $\mathcal{B}_n$	$P_{\mathbf{t}^n}$ from finite-dimensional distribution functions $F_{\mathbf{t}^n}$
real line: $\mathbb{R}$	Borel sets: $\mathcal{B}$	$P$ from a distribution function $F$

In the table, the ??? indicate what we yet have to define – or even show existence of – to reach beyond the elementary probability theory, and into the world of stochastic processes.

### 1.3.2 The distribution of a random sequence

Our aim now is to find the events in  $\mathbb{R}^\infty$  (= all real sequences) and see how one can define a probability measure for these events. When this is done, it is legitimate to talk about *the distribution of an infinite random sequence*. It was this step, from probabilities for one-dimensional or finite-dimensional real sets and events, to probabilities in  $\mathbb{R}^\infty$  and probabilistic statements about infinite sequences, that was made axiomatic in Kolmogorov's celebrated *Grundbegriffe der Wahrscheinlichkeitsrechnung* from 1933, [20].

#### Generalized rectangles, intervals, and the field $\mathcal{I}$

The basic requirement on the events in  $\mathbb{R}^\infty$  is that they should not be simpler than the events in the finite-dimensional spaces  $\mathbb{R}^n$ , which means that if an event  $B_n \in \mathcal{B}_n$  is expressed by means of a finite set of random variables  $x_1, \dots, x_n$ , then it should be an event also in the space  $\mathbb{R}^\infty$ . Now, it can be written

$$\{\mathbf{y} = (x_1, x_2, \dots) \in \mathbb{R}^\infty; (x_1, x_2, \dots, x_n) \in B_n\} = B_n \times \mathbb{R} \times \mathbb{R} \times \dots = B_n \times \mathbb{R}^\infty.$$

A set of this form is called a *generalized rectangle in  $\mathbb{R}^\infty$* . Hence, we have to require that the  $\sigma$ -field of events in  $\mathbb{R}^\infty$  contains at least all generalized rectangles. The natural event field is exactly the smallest  $\sigma$ -field which contains all such sets; cf. Example 1:2. This  $\sigma$ -field is denoted  $\mathcal{B}_\infty$  and is called the Borel field. Symbolically, we can write

$$\mathcal{B}_\infty = \sigma(\cup_{n=1}^\infty (\mathcal{B}_n \times \mathbb{R}^\infty)).$$

A particularly simple form of rectangles are the *intervals*, which are sets of the form

$$I = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_n, b_n] \times \mathbb{R}^\infty,$$

where each  $(a_j, b_j]$  is a half-open interval. Thus, the sequence  $\mathbf{x} = (x_1, x_2, \dots)$  belongs to the interval  $I$  if

$$a_1 < x_1 \leq b_1, a_2 < x_2 \leq b_2, \dots, a_n < x_n \leq b_n. \quad (1.7)$$

Sets which are unions of a finite number of intervals will be important later; they form a field, which we denote  $\mathcal{I}$ . The  $\sigma$ -field generated by  $\mathcal{I}$  is exactly  $\mathcal{B}_\infty$ , i.e.

$$\sigma(\mathcal{I}) = \mathcal{B}_\infty.$$

#### Probabilities on $\mathbb{R}^\infty$

The next step is to assign probabilities to the events in  $\mathcal{B}_\infty$ , and this can be done in either of two ways, from the abstract side or from the observable, finite-dimensional side:

**from a random sequence:** if the probability space  $(\Omega, \mathcal{F}, P)$  is given a priori, and

$$\mathbf{y} = \{x_n\}_{n=1}^{\infty}$$

is a random sequence, i.e. a function from  $\Omega$  to  $\mathbb{R}^{\infty}$ , then a probability measure  $P_{\mathbf{y}}$  is defined on  $(\mathbb{R}^{\infty}, \mathcal{B}_{\infty})$  by

$$P_{\mathbf{y}}(B) = P(\mathbf{y}^{-1}(B)), \quad \text{for } B \in \mathcal{B}_{\infty}.$$

Thus, each random sequence  $\mathbf{y}$  produces a probability measure  $P_{\mathbf{y}}$  on  $(\mathbb{R}^{\infty}, \mathcal{B}_{\infty})$ .

**from a family of finite-dimensional distributions:** if a consistent family of finite-dimensional distributions

$$\mathbf{F} = \{F_{t^n}\}_{n=1}^{\infty}$$

is given a priori, then one can define probabilities  $P_{\mathbf{F}}$  for all half-open  $n$ -dimensional intervals in  $\mathbb{R}^{\infty}$ , by, for  $n = 1, 2, \dots$ , taking (cf. (1.2))

$$P_{\mathbf{F}}((a_1, b_1] \times \dots \times (a_n, b_n] \times \mathbb{R}^{\infty}) = P_n((a_1, b_1] \times \dots \times (a_n, b_n]).$$

Here the probability measure  $P_n$  on  $(\mathbb{R}^n, \mathcal{B}_n)$  is uniquely defined by the distribution functions in  $\mathbf{F}$ . Now, it remains to show that this will give us a countably additive probability measure on the field  $\mathcal{I}$  of finite unions of intervals. By the extension property of probability measures on fields, one can then conclude that there is a unique probability measure  $P$  on  $(\mathbb{R}^{\infty}, \mathcal{B}_{\infty})$  that has  $\mathbf{F}$  as finite-dimensional distributions. The proof of the countable additivity is a significant part of Kolmogorov's existence theorem for stochastic processes; see Appendix A.

By this, we have defined events in  $\mathbb{R}^{\infty}$  and know how to define probability measures in  $(\mathbb{R}^{\infty}, \mathcal{B}_{\infty})$ . What to remember here is in particular, that every probability measure on  $(\mathbb{R}^{\infty}, \mathcal{B}_{\infty})$  is uniquely determined by its finite-dimensional distributions.

### 1.3.3 The continuous parameter case

Now we shall investigate stochastic processes with continuous, one-dimensional time parameter,  $t$  in a real interval  $T$ . By definition, it is a family of random variables  $\{x(t)\}_{t \in T}$ , defined on the same probability space  $(\Omega, \mathcal{F}, P)$ , i.e. it is a function of time  $t \in T$  and outcome  $\omega \in \Omega$ , measurable as a function of  $\omega$  for fixed  $t$ .

Even if this definition is simple and innocent – obviously such processes exist – the practical application needs some care. The sample space  $\Omega$  is an abstract space and a mathematical construction, and the link to reality is provided by the random variables. In an experiment, one can observe the values of one or

more random variables,  $x_1, x_2$ , etc. and also find their distribution, by some statistical procedure. There is no serious difficulty to allow the outcome to be any real number, and to define probability distributions on  $\mathbb{R}$ .

When the result of an experiment is a function with continuous parameter, the situation is more complicated. In principle, *all functions* of  $t \in T$  are potential outcomes, and the sample space of all functions on  $T$  is simply too big to allow any sensible probabilistic structure. There are too many possible realizations that ask for probability.

Here practice comes to our assistance. In an experiment one can only observe the values of  $x(t)$  at a finite number of times,  $t_1, t_2, \dots, t_n$ ; with  $n = \infty$  we allow an unlimited series of observations. The construction of processes with continuous time is built on exactly this fact: the observable events are those which can be defined by countably many  $x(t_j)$ ,  $j \in m\mathbb{N}$ , and the probability measure shall assign probabilities to only such events.

Write  $\mathbb{R}^T$  for the set of all real-valued functions of  $t \in T$ . By an *interval* in  $\mathbb{R}^T$  is meant any set of functions  $x(t)$  which are characterized by finitely many inequalities of the same type as (1.7),

$$a_1 < x(t_1) \leq b_1, a_2 < x(t_2) \leq b_2, \dots, a_n < x(t_n) \leq b_n,$$

the only difference being that now  $t_1, \dots, t_n$  are any  $n$  time points in  $T$ . The *Borel field* in  $\mathbb{R}^T$  is the smallest  $\sigma$ -field that contains all intervals,

$$\mathcal{B}_T = \sigma(\mathcal{I}).$$

### 1.3.3.1 Sets with countable basis

One may wonder how far the Borel sets in  $\mathbb{R}^T$  are from the intervals. The intervals were characterized by some restriction on function values at a finite number of times. A set  $C \subseteq \mathbb{R}^T$  which is characterized by function values at a countable set of times,  $T' = (t_1, t_2, \dots)$  is said to have a *countable basis*. More precisely,  $C \subseteq \mathbb{R}^T$  has a countable basis  $T'$  if there is a Borel set  $B \subset \mathbb{R}^\infty$ , (with  $B \in \mathcal{B}_\infty$ ), such that

$$x \in C \quad \text{if and only if} \quad (x(t_1), x(t_2), \dots) \in B.$$

The Borel sets in  $\mathbb{R}^T$  are exactly those sets which have a countable basis, i.e.

$$\mathcal{B}_T = \{C \subset \mathbb{R}^T; C \text{ has a countable basis}\}.$$

We show this, as an example of a typical  $\sigma$ -field argument.

First, it is clear that if  $B$  is a Borel set in  $\mathbb{R}^\infty$ , then

$$C = \{x \in \mathbb{R}^T; (x(t_1), x(t_2), \dots) \in B\}$$

is a Borel set in  $\mathbb{R}^T$ , since  $\mathcal{B}_T$  contains all intervals with base in  $T'$ , and hence all sets in the  $\sigma$ -field generated by those intervals. This shows that

$$\{C \subset \mathbb{R}^T; C \text{ has a countable basis}\} \subseteq \mathcal{B}_T.$$

To show the other inclusion we show that the family of sets with countable basis is a  $\sigma$ -field which contains the intervals, and then it must be at least as large as the smallest  $\sigma$ -field that contains all interval, namely  $\mathcal{B}_T$ . First, we note that taking complements still gives a set with countable basis. Then, take a sequence  $C_1, C_2, \dots$ , of sets, all with countable basis, and let  $T_1, T_2, \dots, T_j = \{t_1^1, t_2^1, \dots\}, \dots, \{t_1^j, t_2^j, \dots\}$  be the corresponding countable sets of time points, so that

$$C_j = \{x \in \mathbb{R}^T; (x(t_1^{(j)}), x(t_2^{(j)}), \dots) \in B_j\}, \quad \text{with } B_j \in \mathcal{B}_\infty.$$

Then  $T' = \cup_j T_j$  is a countable set,  $T' = (t'_1, t'_2, \dots)$ , and  $\cup_{j=1}^\infty C_j$  is characterized by its values on  $T'$ .

**Example 1:3** Here are some examples of function sets with and without countable basis, when  $T = [0, 1]$ :

- $\{x \in \mathbb{R}^T; \lim_{n \rightarrow \infty} x(1/n) \text{ exists}\} \in \mathcal{B}_T$ ,
- $\{x \in \mathbb{R}^T; \lim_{t \rightarrow 0} x(t) \text{ exists}\} \notin \mathcal{B}_T$ ,
- $\{x \in \mathbb{R}^T; x \text{ is a continuous function}\} \notin \mathcal{B}_T$ ,
- $\{x \in \mathbb{R}^T; x(t) \leq 2 \text{ for all rational } t\} \in \mathcal{B}_T$ ,

### 1.3.3.2 Approximation by finite-dimensional events

The events in the  $\sigma$ -field  $\mathcal{B}_\infty$  in  $\mathbb{R}^\infty$  can be approximated in probability by finite-dimensional sets. If  $(\mathbb{R}^\infty, \mathcal{B}_\infty, P)$  is a probability space, and  $B \in \mathcal{B}_\infty$ , then for every  $\epsilon > 0$ , there is a finite  $n$  and an event  $B_n \in \mathcal{B}_n$  such that

$$P(B \Delta \tilde{B}_n) \leq \epsilon,$$

where  $\tilde{B}_n = \{x \in \mathbb{R}^\infty; (x_1, \dots, x_n) \in B_n\}$  and  $A \Delta B = (A - B) \cup (B - A)$ .

Similarly, events in  $\mathcal{B}_T$  in  $\mathbb{R}^T$  can be approximated arbitrarily close by events defined by the values of  $x(t)$  for a finite number of  $t$ -values:  $P(B \Delta \tilde{B}_n) \leq \epsilon$ , with

$$\tilde{B}_n = \{x \in \mathbb{R}^\infty; (x(t_1), \dots, x(t_n)) \in B_n\}.$$

Remember that every probability measure on  $(\mathbb{R}^\infty, \mathcal{B}_\infty)$  is uniquely determined by its finite-dimensional distributions, which implies that also every probability measure  $P$  on  $(\mathbb{R}^T, \mathcal{B}_T)$  is determined by the finite-dimensional distributions,  $\{F_{t^n}\}_{n=1}^\infty$ . In particular, the probability

$$P(\lim_{n \rightarrow \infty} x(t_0 + 1/n) \text{ exists and is equal to } x(t_0))$$

is determined by the finite-dimensional distributions. Unfortunately,  $x(t_0 + 1/n) \rightarrow x(t_0)$  as  $n \rightarrow \infty$ , is almost, but not quite, the same as  $x(t) \rightarrow x(t_0)$  as  $t \rightarrow t_0$ . To deal with sample function continuity we need more refined construction of the probability measure from the finite-dimensional distributions.

## 1.4 Stationary processes and fields

This section summarizes some elementary notation for stationary processes. More details, properties, and proofs of the most important facts, will be given in Chapter 4.

To avoid too cumbersome notation we from now on allow ourselves to talk about “the process  $x(t)$ ”, when we should have used the notation  $\{x(t), t \in \mathbb{R}\}$  for the process. If we mean “the random variable  $x(t)$ ” we will say so explicitly.

### 1.4.1 Stationary processes

A stochastic process  $x(t)$  is *strictly stationary* if all  $n$ -dimensional distributions of

$$x(t_1 + \tau), \dots, x(t_n + \tau)$$

are independent of  $\tau$ . It is called *weakly stationary* (the term *second order stationary* is also used) if its mean is constant,  $E(x(t)) = m$ , and if its *covariance function*

$$r(t) = \text{Cov}(x(s+t), x(s)),$$

is a function only of the time lag  $t$ . Every continuous covariance function has a representation as a Fourier integral,

$$r(t) = \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega), \quad (1.8)$$

where the function  $F(\omega)$  is called the *spectral distribution function*. It is characterized by the properties:

- symmetry:  $dF(-\omega) = dF(\omega)$ ,
- monotonicity:  $\omega \leq \omega'$  implies  $F(\omega) \leq F(\omega')$ ,
- boundedness:  $F(+\infty) - F(-\infty) = r(0) < \infty$ .

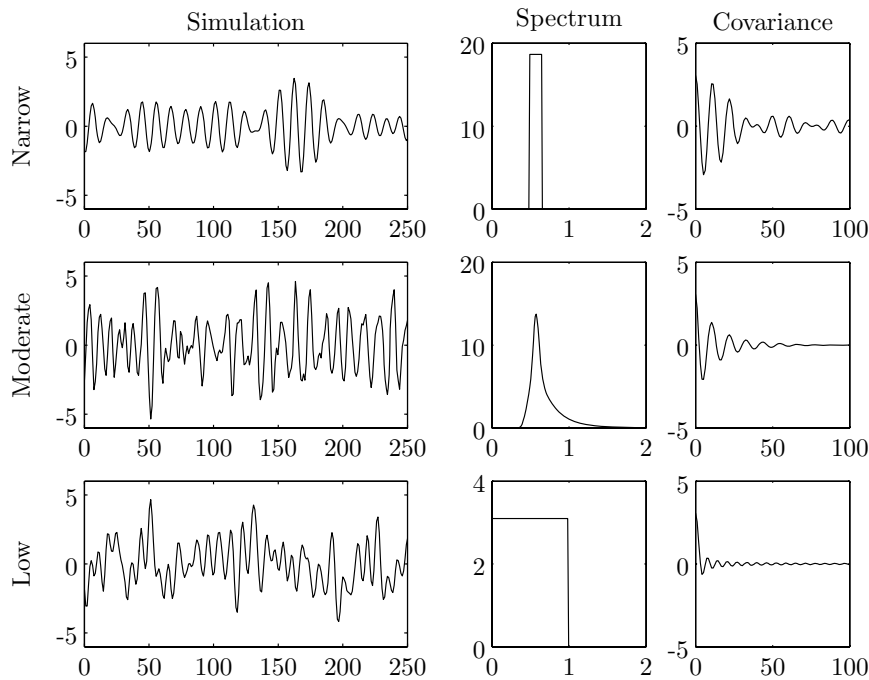
As indicated by the way we write the three properties,  $F(\omega)$  is defined only up to an additive constant, and we usually take  $F(-\infty) = 0$ . The spectral distribution function is then equal to a cumulative distribution function multiplied by a positive constant, equal to the variance of the process.

If  $F(\omega)$  is absolutely continuous with  $F(\omega) = \int_{s=-\infty}^{\omega} f(s) ds$ , then the spectrum is said to be (absolutely) continuous, and  $f(\omega)$  is the *spectral density function*; see Section 5.6.2.1 for more discussion of absolute continuity.

The *spectral moments* are defined as

$$\omega_k = \int_{-\infty}^{\infty} |\omega|^k dF(\omega).$$

Note that the odd spectral moments are defined as absolute moments. Since  $F$  is symmetric around 0 the signed odd moments are always 0. Spectral



**Figure 1.2:** Processes with narrow band spectrum, moderate width JONSWAP wave spectrum, and low frequency white noise spectrum.

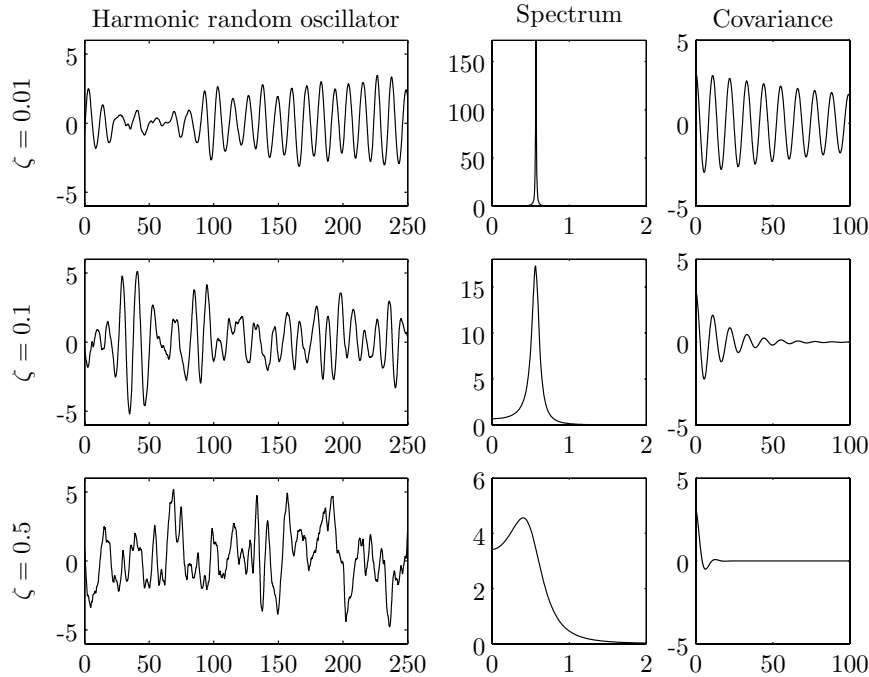
moments may be finite or infinite. As we shall see in the next chapter, the finiteness of the spectral moments are coupled to the smoothness properties of the process  $x(t)$ . For example, the process is differentiable (in quadratic mean), see Section 2.1, if  $\omega_2 = -r''(0) < \infty$ , and similarly for higher order derivatives.

As we shall see in later sections,  $\omega$  is in a natural way interpreted as an *angular frequency*, not to be confused with the elementary event  $\omega$  in basic probability theory.

**Example 1:4** Here is a first example on the visual characteristics of spectrum, covariance function, and sample function. Figure 1.2 illustrates one very narrow spectrum, one realistic water wave spectrum, and one “low frequency white noise” spectrum. Figure 1.3 show the output of a linear oscillator driven by white noise, and with different relative damping; see Section 4.4.3.3 and Example 4:10.

### 1.4.2 Random fields

A random field is a stochastic process  $x(\mathbf{t})$  with multi-dimensional parameter  $\mathbf{t} = (t_1, \dots, t_p) \in \mathbf{T}$ , which can be discrete or continuous. For example, if



**Figure 1.3:** Harmonic oscillator with different relative damping  $\zeta$ .

$\mathbf{t} = (t_1, t_2)$  is two-dimensional we can think of  $(t_1, t_2, x(\mathbf{t}))$ ,  $(t_1, t_2) \in \mathbb{R}^2$ , as a random surface. The mean value and covariance functions are defined in the natural way,  $m(\mathbf{t}) = E(x(\mathbf{t}))$  and  $r(\mathbf{t}, \mathbf{u}) = C(x(\mathbf{t}), x(\mathbf{u}))$ .

A random field is called *homogeneous* if it has constant mean value  $m(\mathbf{t}) = m$  and the covariance function  $r(\mathbf{t}, \mathbf{u})$  depends only on the vector  $\mathbf{t} - \mathbf{u}$  between the two observation points, i.e. assuming  $m = 0$ ,

$$r(\mathbf{t}) = r(\mathbf{u} + \mathbf{t}, \mathbf{u}) = E(x(\mathbf{u} + \mathbf{t}) \cdot x(\mathbf{u})).$$

The covariance of the process values at two parameter points depends on *distance* as well as on *direction* of the vector between the two points.

If the covariance between  $x(\mathbf{u})$  and  $x(\mathbf{v})$  depends only on the distance  $\tau = \|\mathbf{u} - \mathbf{v}\|$  between the observation points and not on the direction, the field is called *isotropic*. This requirement poses severe restrictions on the covariance function, as we shall see in Chapter 6, where random fields are treated in more detail.

## 1.5 Gaussian processes

### 1.5.1 Multivariate normal distributions and Gaussian processes

**Definition 1.3** A vector  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)'$  of  $p$  random variables is said to have a  $p$ -variate Gaussian (normal) distribution if every linear combination

of its components  $\mathbf{a}' \cdot \boldsymbol{\xi} = \sum_k a_k \xi_k$  has a normal distribution. The variables  $\xi_1, \dots, \xi_p$  are then said to be “jointly normal”.

With mean vector  $\mathbf{m} = E(\boldsymbol{\xi})$  and covariance matrix

$$\boldsymbol{\Sigma} = \text{Cov}(\boldsymbol{\xi}; \boldsymbol{\xi}) = E((\boldsymbol{\xi} - \mathbf{m}) \cdot (\boldsymbol{\xi} - \mathbf{m})'),$$

the variance of  $\mathbf{a}' \cdot \boldsymbol{\xi}$  is

$$V(\mathbf{a}' \cdot \boldsymbol{\xi}) = \mathbf{a}' \boldsymbol{\Sigma} \mathbf{a}.$$

If the determinant of  $\boldsymbol{\Sigma}$  is positive, the distribution of  $\boldsymbol{\xi}$  is non-singular and has a density

$$f_{\boldsymbol{\xi}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \sqrt{\boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})}.$$

If the determinant is zero, the distribution of  $\boldsymbol{\xi}$  is concentrated to linear subspace of  $\mathbb{R}^n$  and there exists at least one linear relationship between the components, i.e. there is at least one  $\mathbf{a}$  for which  $\mathbf{a}' \cdot \boldsymbol{\xi}$  is a constant.

**Definition 1:4** A stochastic process  $\{x(t), t \in \mathbb{R}\}$  is a Gaussian process if every linear combination

$$S = \sum_k a_k x(t_k)$$

for real  $a_k$  and  $t_k \in \mathbb{R}$  has a Gaussian distribution.

It is an easy consequence of the definition that the derivative of a Gaussian process is also Gaussian (when it exists), since it is the limit of the Gaussian variable  $z_h = (x(t+h) - x(t))/h$  as  $h \rightarrow 0$ . For a stationary Gaussian process  $\{x(t), t \in \mathbb{R}\}$  the mean of  $z_h$  is 0 and it has variance  $V(z_h) = 2(r(0) - r(h))/h^2$ . As we shall prove in Section 2.4.2 this converges to  $\omega_2 = \int \omega^2 dF(\omega \leq \infty)$ . The derivative exists only if this limit is finite.

Also the integral of a Gaussian process is a Gaussian variable; conditions for the existence will be given in Section 2.6.

### 1.5.1.1 Conditional normal distributions

The multivariate normal distribution has the very useful property that conditioned on observations of a subset of variables, the unobserved variables are also normal. Further, the conditional mean is a linear in the observations while variances and covariances are independent of observations.

Let  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)'$  and  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_m)'$  be two jointly Gaussian vectors with mean values

$$E(\boldsymbol{\xi}) = m_{\boldsymbol{\xi}}, \quad E(\boldsymbol{\eta}) = m_{\boldsymbol{\eta}},$$

and with covariance matrix (with  $\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}} = \boldsymbol{\Sigma}'_{\boldsymbol{\eta}\boldsymbol{\xi}}$ )

$$\Sigma = \text{Cov}((\boldsymbol{\xi}, \boldsymbol{\eta}); (\boldsymbol{\xi}, \boldsymbol{\eta})) = \begin{pmatrix} \Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}} & \Sigma_{\boldsymbol{\xi}\boldsymbol{\eta}} \\ \Sigma_{\boldsymbol{\eta}\boldsymbol{\xi}} & \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{pmatrix}.$$

If the determinant of the covariance matrix  $\Sigma$  is positive, then the distribution of  $(\boldsymbol{\xi}, \boldsymbol{\eta})$  has a non-singular density

$$f_{\boldsymbol{\xi}\boldsymbol{\eta}}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{(m+n)/2} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{x}-m_{\boldsymbol{\xi}}, \mathbf{y}-m_{\boldsymbol{\eta}})\Sigma^{-1}(\mathbf{x}-m_{\boldsymbol{\xi}}, \mathbf{y}-m_{\boldsymbol{\eta}})'}$$

The density of  $\boldsymbol{\eta}$  is

$$f_{\boldsymbol{\eta}}(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}}} e^{-\frac{1}{2}(\mathbf{y}-m_{\boldsymbol{\eta}})\Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\mathbf{y}-m_{\boldsymbol{\eta}})'}$$

and the conditional density  $f_{\boldsymbol{\xi}|\boldsymbol{\eta}}(\mathbf{x}|\mathbf{y})$ , defined as

$$f_{\boldsymbol{\xi}|\boldsymbol{\eta}}(\mathbf{x} | \mathbf{y}) = \frac{f_{\boldsymbol{\xi}\boldsymbol{\eta}}(\mathbf{y}, \mathbf{x})}{f_{\boldsymbol{\eta}}(\mathbf{y})},$$

is also Gaussian with conditional mean matrix

$$\begin{aligned} E(\boldsymbol{\xi} | \boldsymbol{\eta} = \mathbf{y}) &= \widehat{\boldsymbol{\xi}}(\mathbf{y}) = E(\boldsymbol{\xi}) + \text{Cov}(\boldsymbol{\xi}, \boldsymbol{\eta}) \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} (\mathbf{y} - E(\boldsymbol{\eta}))' \\ &= m_{\boldsymbol{\xi}} + \Sigma_{\boldsymbol{\xi}\boldsymbol{\eta}} \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} (\mathbf{y} - m_{\boldsymbol{\eta}})'. \end{aligned} \quad (1.9)$$

$$(1.10)$$

The conditional covariance is

$$\Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}|\boldsymbol{\eta}} = E((\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}(\boldsymbol{\eta})) \cdot (\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}(\boldsymbol{\eta}))') = \Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\xi}\boldsymbol{\eta}} \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} \Sigma_{\boldsymbol{\eta}\boldsymbol{\xi}}. \quad (1.11)$$

In two dimensions the formulas read

$$\begin{aligned} m_{x|y} &= m_x + \sigma_x \sigma_y \rho_{xy} \cdot \frac{y - m_y}{\sigma_y}, \\ \sigma_{x|y}^2 &= \sigma_x^2 (1 - \rho_{xy}^2), \end{aligned}$$

with  $\rho_{xy} = \text{Cov}(x, y) / \sqrt{V(x)V(y)}$ ; thus the squared correlation  $\rho_{xy}^2$  gives the relative reduction of the variability (uncertainty) in the random variable  $x$  gained by observation of  $y$ .

Observe the mnemotechnical friendliness of these formulas. For example, the covariance matrix  $\Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}|\boldsymbol{\eta}}$  has dimension  $n \times n$  and the configuration on the right hand side of (1.11) is the only way to combine the matrices involved that matches their dimensions – of course, you have to remember the general structure.

### 1.5.2 Linear prediction and reconstruction

Prediction and reconstruction are two of the most important applications of stationary process theory. Even though these problems are not main topics in this work, we present one of the basic concepts here; in Section 5.6 we will deal with the more philosophical sides of the prediction problem.

Suppose we have observed the outcomes of a set of random variables,  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_m)$  and that we want to give a statement  $\widehat{\xi}$  about the outcome of some other variable  $\xi$ , either to be observed sometimes in the future, or perhaps a missing observation in a time series. These two cases constitute the framework of *prediction* and *reconstruction*, respectively. Also suppose that we want to make the statement in the best possible way in the mean square sense, i.e. we want  $E((\xi - \widehat{\xi})^2)$  to be as small as possible.

Now we know from Theorem 1:1 that the best solution in mean square sense is given by the conditional expectation,  $\widehat{\xi} = E(\xi | \boldsymbol{\eta}) = \phi(\boldsymbol{\eta})$ . On the other hand, if the variables are jointly Gaussian, then we know from Section 1.5, formula 1.9 that the conditional expectation of  $\xi$  given  $\boldsymbol{\eta}$  is linear in  $\boldsymbol{\eta}$ , so that for Gaussian variables the optimal solution is

$$\widehat{\xi} = E(\xi | \boldsymbol{\eta}) = m_\xi + \boldsymbol{\Sigma}_{\xi\boldsymbol{\eta}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta} - m_{\boldsymbol{\eta}}), \quad (1.12)$$

and expression that depends only on the the mean values and the second order moments, i.e. variances and covariances.

We now look at the general case, without assuming normality, and restrict ourselves to solutions that are linear functions of the observed variables. It is clear that the solution that is optimal in the mean square sense only depends on the mean values and variances/covariances of the variables. It therefore has the same form for all variables with the same first and second order moments. Thus, (1.12) gives the best linear predictor in mean square sense.

### 1.5.3 Some useful inequalities

We shall in the next chapter need the following inequality for the normal density and distribution functions,  $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ ,  $\Phi(x) = \int_{-\infty}^x \phi(y) dy$ :

$$\phi(x) \left( \frac{1}{x} - \frac{1}{x^3} \right) \leq 1 - \Phi(x) \leq \phi(x) \frac{1}{x}, \quad (1.13)$$

for  $x > 0$ .

The following asymptotic expansion is useful as  $x \rightarrow \infty$ ,

$$1 - \Phi(x) \sim \phi(x) \left( \frac{1}{x} - \frac{1}{x^3} + \frac{1 \cdot 3}{x^5} - \frac{1 \cdot 3 \cdot 5}{x^7} + \dots (-1)^k \frac{1 \cdot 3 \cdots (2k-1)}{x^{2k+1}} \right).$$

Here the right hand side overestimates  $1 - \Phi(x)$  for  $x > 0$  if  $k$  is even and underestimates it if  $k$  is odd. More precisely, the difference between the left and right hand side is of the same order as  $\phi(x)/x^{2k+3}$  as  $x \rightarrow \infty$ . (The sign  $\sim$  means that the ratio between the left and the right hand side goes to 1.)

## 1.6 Some historical landmarks

This section contains a personal selection of research achievements that have shaped the theory of stationary processes in general. The examples are chosen, not only because they are important in their respective fields, but also because they illustrate the necessity of exchange of ideas between probability and statistics theory and applications.

### 1.6.1 Brownian motion and the Wiener process

There is no other Gaussian process with as wide applicability as the Wiener process. Even if it is a non-stationary process it appears repeatedly in the theory of stationary processes, and we spend this section to describe some of its properties and applications.

**Definition 1:5** *The Wiener process  $\{w(t); t \geq 0\}$  is a Gaussian process with  $w(0) = 0$  such that  $E(w(t)) = 0$ , and the variance of the increment  $w(t+h) - w(t)$  over any interval  $[t, t+h]$ ,  $h > 0$ , is proportional to the interval length,*

$$V(w(t+h) - w(t)) = h\sigma^2.$$

*A Wiener process  $\{w(t), t \in \mathbb{R}\}$  over the whole real line is a combination of two independent Wiener processes  $w_1$  and  $w_2$ , so that  $w(t) = w_1(t)$  for  $t \geq 0$  and  $w(t) = w_2(-t)$  for  $t < 0$ .*

It is an easy consequence of the definition that the increment  $w(t) - w(s)$  is uncorrelated with  $w(s)$  for  $s < t$ ,

$$V(w(t)) = V(w(s)) + V(w(t) - w(s)) + 2Cov(w(s), w(t) - w(s)),$$

and that therefore  $Cov(w(s), w(t)) = \sigma^2 \min(s, t)$ . Since the increments over disjoint intervals are normal, by the definition of a normal process, they are also independent.

A characteristic feature of the Wiener process is that its future changes are statistically independent of its actual and previous values. It is intuitively clear that a process with this property cannot be differentiable. The increment over a small time interval from  $t$  to  $t+h$  is of the order  $\sqrt{h}$ , which is small enough to make the process continuous, but it is too large to give a differentiable process.<sup>5</sup> The sample functions are in fact objects that have *fractal dimension*, and the process is *self similar* in the sense that when magnified with proper scales it retains its statistical geometrical properties. More precisely, for each  $a > 0$ , the process  $\sqrt{a}w(t/a)$  has the same distributions as the original process  $w(t)$ .

The Wiener process is commonly used to model phenomena where the local changes are virtually independent. Symbolically, one uses to write  $dw(t)$  for the

<sup>5</sup>Chapter 2 gives conditions for continuity and differentiability of sample functions.

infinitesimal independent increments, or simply as a “derivative”  $w'(t)$ . The Brownian motion is a good example of how one can use the Wiener process to get models with more or less physical realism.

The Brownian motion, first described 1828 by the Scottish biologist Robert Brown, is an erratic movement by small particles immersed in a fluid, for example pollen particles in water as in Brown’s original experiment. Albert Einstein’s presented 1905 a quantitative model for the Brownian movement in his paper *On the movements of small particles in a stationary liquid demanded by the molecular-kinetic theory of heat*, reprinted in [13], based on the assumption that the movements are caused by independent impacts on the particle by the molecules of the surrounding fluid medium.

In Einstein’s model the changes in location due to collisions over separate time intervals are supposed to be independent. This requires however that the particles have no mass, which physically wrong, but the model is still sufficiently accurate for microscopic purposes. According to Einstein, the change in location in any of the three directions,  $(x, y, z)$  over a time interval of length  $t$  is random and normal with mean zero, which is not surprising, since they are the results of a very large number of independent collisions. What made Einstein’s contribution conclusive was that he derived an expression for the variance in terms of other physical parameters, namely

$$V(x(t)) = V(y(t)) = V(z(t)) = t \frac{4RT}{Nf} = t \sigma^2, \quad (1.14)$$

where  $T$  is the absolute temperature,  $R$  is the Boltzmann constant, and  $N$  is Avogadro’s number, i.e. the number of molecules per mole of an ideal gas, and the friction coefficient  $f$  depends on the shape and size of the particle and on the viscosity of the fluid. Each coordinate are here independent Wiener processes.

Observations of the Brownian movement and estimation of its variance makes it possible to calculate any of the factors in  $\sigma^2$ , for example  $N$ , from the other ones. The French physicist J.B. Perrin estimated in a series of experiments 1908-1911 Avogadro’s number in this way by observing suspended rubber particles and found an estimate correct within about 10%.

In a more realistic model, one takes also particle mass and velocity into account. If  $v(t)$  denotes the velocity at time  $t$ , the fluid offers a resistance from the friction force, which is equal to  $fv(t)$ , with the friction coefficient as in (1.14). Further, the particle offers a resistance to changes in velocity proportional to its mass  $m$ . Finally, one needs to model the independent collisions by the fluid molecules, and here the Wiener process can be used, more precisely its increments  $dw(t)$ . This gives the *Langevin equation* for the particle velocity,

$$dv(t) + \alpha v(t) dt = \frac{1}{m} dw(t), \quad (1.15)$$

where  $\alpha = f/m$ , and  $w(t)$  is a standardized Wiener process. It is usually written

$$\frac{dv(t)}{dt} + \alpha v(t) = \frac{1}{m} w'(t). \quad (1.16)$$

We will meet the Langevin equation in Example 4:3 on the Ornstein-Uhlenbeck process in Section 4.3.3.6.

### 1.6.2 Rice and electronic noise

The two papers *Mathematical analysis of random noise*, by S.O. Rice, appeared in Bell System Technical Journal, 1944-1945, [27]. They represent a landmark in the history of stochastic processes in that they bring together and exhibit the wide applicability of the spectral formulation of a stationary process as a sum, or asymptotically an integral, of harmonic cosine functions with random amplitudes and phases. Correlation functions and their Fourier transforms had been studied at least since the early 1900s, and Rice's work brought together these results in a systematic way. But it also contained many new results, in particular pertaining crossing related properties, on the statistical properties of stationary processes "obtained by passing random noise through physical devices".

Rice uses the *spectral representation* of a stationary Gaussian process as a sum over discrete positive frequencies  $\omega_n > 0$ ,

$$x(t) = \sum_n a_n \cos \omega_n t + b_n \sin \omega_n t = \sum_n c_n \cos(\omega_n t + \phi_n) \quad (1.17)$$

where the amplitudes  $a_n, b_n$  and normal and independent with mean 0 and  $E(a_n^2) = E(b_n^2) = \sigma_n^2$ , and  $\phi_n$  uniformly distributed over  $(0, 2\pi)$ , independent of the amplitudes. As we shall see in Chapter 2 such a process has covariance

$$r(t) = \sum_n \sigma_n^2 \cos \omega_n t.$$

The spectral distributions function is a discrete distribution with point mass  $\sigma_n^2/2$  at the symmetrically located frequencies  $\pm\omega_n$ .

The absolutely continuous, integral form of the spectral representation is presented as limiting cases in Rice's work. At about the same time, Cramér gave a probabilistic formulation of the continuous spectral representation, in a mathematically impeccable way; [6, 7].

Besides the previously known "Rice's formula" for the expected number of level crossings, Rice's 1945 paper also analyzed crossing and excursion distributions and investigated the joint occurrence of crossings of a fixed level at two distinct points, necessary for calculation of the variance of the number of crossings.

The flexibility and generality of Rice's methods and examples, made correlation and spectral theory fundamental ingredients in communication theory and signal processing for decades to come. An example by Rice himself, is the ingenious explanation of the intriguing click noise in analogue FM-radio, [28].

### 1.6.3 Gaussian random wave models

Steve Rice's analysis of time dependent stationary processes had, as mentioned, great influence on signal processing in the information sciences. Less well known in the statistical world is the effect his work had in oceanography and naval architecture.

It is well worth to cite in extenso (references deleted) the first two paragraphs in Manley St. Denis and Willard J. Pierson's paper: *On the motion of ships in confused seas*, which came out 1954, [32].

#### HISTORY

Three years ago to first co-author of the present work collaborated with Weinblum in the writing of a paper entitled "On the motion of ships at sea". In that paper Lord Rayleigh was quoted saying: "The basic law of the seaway is the apparent lack of any law". Having made this quotation, however, the authors then proceed to consider the seaway as being composed of "a regular train of waves defined by simple equations". This artificial substitution of pattern for chaos was dictated by the necessity of reducing the utterly confused reality to a simple form amenable to mathematical treatment.

Yet at the same time and in other fields the challenging study of confusion was being actively pursued. Thus in 1945 Rice was writing on the mathematical analysis of random noise and in 1949 Tukey and Hamming were writing on the properties of stationary time series and their power spectra in connection with colored noise. In the same year Wiener published his now famous book on time series. These works were written as contributions to the theory of communication. Nevertheless the fundamental mathematical discipline expounded therein can readily be extended to other fields of scientific endeavor. Thus in 1952 the second co-author, inspired by a contribution of Tukey, was able to apply the foregoing theories to the study of actual ocean waves. As the result of analyses of actual wave records, he succeeded in giving not only a logical explanation as to why waves are irregular, but a statement as well of the laws underlying the behavior of a seaway. There is indeed a basic law of the seaway. Contrary to the obvious inference from the quotation of Lord Rayleigh, the seaway can be described mathematically and precisely, albeit in a statistical way.

If Rice's work had been in the vein of generally accepted ideas in communication theory, the St Denis and Pierson paper represented a complete revolution in common naval practice. Nevertheless, its treatment of irregular water waves as, what now is called, a random field was almost immediately accepted, and set a standard for much of naval architecture.

One possible reason for this can be that the authors succeeded to formulate and analyze the motions of a ship that moved with constant speed through the field in a rational way. The random sea could directly be used as input to a linear (later also non-linear) filter representing the ship.

St. Denis and Pierson extended the one-dimensional description of a time dependent process  $\{x(t), t \in \mathbb{R}\}$ , useful for example to model the waves measured at a single point, to a random field  $x(t, (s_1, s_2))$  with time and location parameter  $(s_1, s_2)$ . They generalized the sum (1.17) to be a sum of a packet of directed waves, with  $\omega = (\omega, \kappa_1, \kappa_2)$ ,

$$\sum_{\omega} A_{\omega} \cos(\omega t - \kappa_1 s_1 - \kappa_2 s_2 + \phi_{\omega}). \quad (1.18)$$

with random amplitude and phase.

For fixed  $t$  each element in (1.18) is a cosine-function in the plane, which is zero along lines  $\omega t - \kappa_1 s_1 - \kappa_2 s_2 + \phi_{\omega} = \pi/2 + k\pi$ ,  $k$  integer. The parameters  $\kappa_1$  and  $\kappa_2$  are called the *wave numbers*. For fixed  $(s_1, s_2)$  it is a cosine wave with (*angular*) *frequency*  $\omega$ .

Water waves are special cases of homogeneous random fields, for which there is a special relation between time and space frequencies (wave numbers). For a one-dimensional time dependent Gaussian wave  $x(t, s)$ , where  $s$  is distance along an axis, the elementary waves have the form

$$A_{\omega} \cos(\omega t - \kappa s + \phi_{\omega}).$$

By physical considerations one can derive an explicit relation, called the *dispersion relation*, between wave number  $\kappa$  and frequency  $\omega$ . If  $h$  is the water depth, then

$$\omega^2 = \kappa g \tanh(h\kappa),$$

which for infinite depth reduces to  $\omega^2 = \kappa g$ . Here  $g$  is the constant of gravity.

The case of a two-dimensional time dependent Gaussian wave  $x(t, s_1, s_2)$ , the elementary waves with frequency  $\omega$  and direction  $\theta$  becomes

$$A_{\omega} \cos(\omega t - \kappa(s_1 \cos \theta + s_2 \sin \theta) + \phi_{\omega}),$$

where  $\kappa$  is given by the dispersion relation.

The spectral distribution is often written in polar form, with spectral density

$$f(\omega, \theta) = f(\omega)g(\omega, \theta),$$

where the *spreading function*  $g(\omega, \theta)$  has  $\int_0^{2\pi} g(\omega, \theta) d\theta = 1$ .

In their paper St. Denis and Pierson also laid out the theory for how the wave spectrum should be transformed to a response spectrum for the motion of a ship, and they also described how the spectrum is changed to an *encountered spectrum* when a ship sails with constant speed through the waves.

### 1.6.4 Detection theory and statistical inference

The first three landmarks illustrated the relation between stochastic model building and physical knowledge, in particular how the concepts of statistical independence and dependence between signal and functions relate to the physical world. About the same time as Rice and StDenis & Pierson advanced physically based stochastic modeling, the statistical inference methodology was placed firmly into a theoretical mathematical framework, as documented by the classical book by Harald Cramér, *Mathematical methods of Statistics*, 1945, [8].

A few years later, the connection between the theoretical basis for statistical inference and important engineering questions related to signal detection was elegantly exploited by Ulf Grenander in his PhD thesis from Stockholm, *Stochastic processes and statistical inference*, [15]. The classical problem in signal processing of deciding whether a deterministic signal of known shape  $s(t)$  is present in an environment of Gaussian dependent, colored as opposed to white, random noise,  $x(t)$  can be treated as an infinite dimensional decision problem, testing an infinite dimensional statistical hypothesis; see also the classical book on detection theory [33]

Suppose one observes a Gaussian stochastic process  $x(t)$ ,  $a \leq t \leq b$ , with known correlation structure, but with unknown mean value function  $m(t)$ . If no signal is present, the mean value is 0, but with signal, the mean is equal to the known function  $s(t)$ . In statistical terms one has to test the following two hypotheses against each other:

$$H_0 : m(t) = 0,$$

$$H_1 : m(t) = s(t).$$

Grenander introduced a series of independent Gaussian *observables*,  $y_k = \int h_k(t)x(t) dt$ , by choosing the filter functions  $h_k$  as solutions to the integral equation

$$\int r(s, t)h_k(t) dt = c_k h_k(s),$$

with  $c_1 \geq c_2 \geq \dots$ ,  $c_k \rightarrow 0$  as  $k \rightarrow \infty$ . Under  $H_1$  the observables will have mean  $a_k = \int h_k(t)m(t) dt$  and variance  $c_k$ , while under  $H_0$  they will have mean 0, and the same variance. So instead of a continuous problem, we have gotten a denumerable problem, in which one can make a Likelihood-Ratio test of the two alternatives. We will return to this problem in Section 4.5, Example 4:12.

## Exercises

1:1. Consider the sample space  $\Omega = [0, 1]$  with uniform probability  $P$ , i.e.  $P([a, b]) = b - a$ ,  $0 \leq a \leq b \leq 1$ . Construct a stochastic process  $\mathbf{y} = (x_1, x_2, \dots)$  on  $\Omega$  such that the components are independent zero-one variables, with  $P(x_k = 0) = P(x_k = 1)$ . What is the distribution of  $\sum_{k=1}^{\infty} x_k/2^k$ ?

1:2. Show that  $\mathcal{B}(\mathbb{R}) =$  the class of Borel sets in  $\mathbb{R}$  is generated by

- a) the open intervals,
- b) the closed intervals.

1:3. A set  $A \subset \{1, 2, \dots\}$  is said to have asymptotic density  $\theta$  if

$$\lim_{n \rightarrow \infty} n^{-1} |A \cap \{1, 2, \dots, n\}| = \theta.$$

(Note,  $|B|$  denotes the number of elements in  $B$ .) Let  $\mathcal{A}$  be the family of sets for which the asymptotic density exists. Is  $\mathcal{A}$  a field? A  $\sigma$ -field?

1:4. Let  $x_1, x_2, \dots$  be random variables with values in a countable set  $\mathcal{F}$ , and suppose there are real constants  $a_k$  such that

$$\sum_{k=1}^{\infty} P(x_k \neq a_k) < \infty, \quad \sum_{k=1}^{\infty} a_k < \infty.$$

Prove that the sum  $x = \sum_{k=1}^{\infty} x_k$  has a discrete distribution, i.e. there exists a countable set  $D$  such that  $P(x \in D) = 1$ . (Hint: Use the Borel-Cantelli lemma, that says that if  $\sum_k P(A_k) < \infty$  then, with probability one, only a finite number of the events  $A_k$  occur.)

Show by example that it is possible for independent random variables  $x_k$  to have a sum  $\sum_{k=1}^{\infty} x_k$  with a continuous distribution, although all  $x_k$  are discrete variables with a common value space – obviously they can not be identically distributed.

1:5. Take  $\mathbb{R}^n$  and motivate that the family  $\mathcal{F}_0$  whose elements are unions of finitely many rectangles  $(a_i, b_j]$  (with possibly infinite end points) is a field.

Let  $T$  be an interval and convince yourself that the finite dimensional rectangles in  $\mathbb{R}^T$  and unions of finitely many such rectangles, is a field.

1:6. Take  $T = [0, 1]$ , and consider the set of functions which are continuous on the rational numbers, i.e.

$$C_Q = \{x \in \mathbb{R}^T; x(q) \rightarrow x(q_0) \text{ for all rational numbers } q_0\},$$

where the limit is taken as  $q$  tends to  $q_0$  through the rational numbers. Show that  $C_Q \in \mathcal{B}_T$ .

- 1:7. Prove Theorem 1:1.
- 1:8. Prove that the increments of a Wiener process, as defined as in Definition 1:5, are independent and normal.

## Chapter 2

# Stochastic analysis

This chapter is the stochastic equivalent of real analysis and integration. As in its deterministic counterpart limiting concepts and conditions for the existence of limits are fundamental. We repeat the basic stochastic limit definitions; a summary of basic concepts and results on stochastic analysis is given in Appendix B.

**Definition 2:1** *Let  $\{x_n\}_{n=1}^{\infty}$  be a random sequence, with the random variables  $x_1(\omega)$ ,  $x_2(\omega)$ ,  $\dots$  defined on the same probability space as a random variable  $x = x(\omega)$ . Then, the convergence  $x_n \rightarrow x$  as  $n \rightarrow \infty$  can be defined in three ways:*

- **almost surely, with probability one** ( $x_n \xrightarrow{a.s.} x$ ):  $P(\{\omega; x_n \rightarrow x\}) = 1$ ;
- **in quadratic mean** ( $x_n \xrightarrow{q.m.} x$ ):  $E(|x_n - x|^2) \rightarrow 0$ ;
- **in probability** ( $x_n \xrightarrow{P} x$ ): for every  $\epsilon > 0$ ,  $P(|x_n - x| > \epsilon) \rightarrow 0$ .

In Appendix B we give several conditions, necessary and sufficient, as well as only sufficient, for convergence of a random sequence  $x_n$ . The most useful of these involve only conditions on the bivariate distributions of  $x_m$  and  $x_n$ . We shall in this chapter examine such conditions for sample function continuity, differentiability, and integrability. We shall also give conditions which guarantee that only simple discontinuities occur. In particular, we shall formulate conditions in terms of bivariate distributions, which are easily checked for most standard processes, such as the normal and the Poisson process.

### 2.1 Quadratic mean properties

We first repair some concepts and properties that may be well known from previous courses in stochastic processes. We return to proofs and more details in Section 2.4.

A stochastic process  $\{x(t), t \in \mathbb{R}\}$  is said to be *continuous in quadratic mean* (or  $L^2$ -continuous) at time  $t$  if

$$x(t+h) \xrightarrow{q.m.} x(t)$$

as  $h \rightarrow 0$ , i.e. if  $E((x(t+h) - x(t))^2) \rightarrow 0$ . It is called *differentiable in quadratic mean* with derivative  $y(t)$  if

$$\frac{x(t+h) - x(t)}{h} \xrightarrow{q.m.} y(t)$$

as  $h \rightarrow 0$ . Of course, the process  $\{y(t), t \in \mathbb{R}\}$  is called the (quadratic mean) derivative of  $\{x(t), t \in \mathbb{R}\}$  and is denoted  $x'(t)$ . A stationary process  $x(t)$  is continuous in quadratic mean if its covariance function  $r(t)$  is continuous at  $t = 0$ . It is differentiable if  $r(t)$  is twice differentiable and then the derivative has covariance function

$$r_{x'}(t) = -r''(t).$$

Second and higher order derivatives are defined recursively:  $\{x(t), t \in \mathbb{R}\}$  is twice differentiable in quadratic mean if and only if its (quadratic mean) derivative is (quadratic mean) differentiable, i.e. if  $r^{iv}(t)$  exists. etc. The covariance function of  $\{x''(t), t \in \mathbb{R}\}$  is  $r_{x''}(t) = r^{iv}(t)$ .

Expressed in terms of the spectral distribution function  $F(\omega)$ , the process is differentiable in quadratic mean if and only if the second spectral moment is finite, i.e.

$$\omega_2 = \int_{-\infty}^{\infty} \omega^2 dF(\omega) < \infty;$$

for a proof, see Lemma 2.3, page 47. Since  $\omega_2 = -r''(0) = V(x'(t))$ , the finiteness of  $\omega_2$  is necessary and sufficient for the existence of a quadratic mean derivative. Analogous relations hold for higher derivatives of order  $k$  and the spectral moments  $\omega_{2k} = \int \omega^{2k} dF(\omega)$ . We will give some more details on quadratic mean properties in Section 2.4.

## 2.2 Sample function continuity

### 2.2.1 Countable and uncountable events

The first problem that we encounter with sample function continuity is that the sample event of interest, namely the set of continuous functions,

$$\mathbb{C} = \{x \in \mathbb{R}^T; x(\cdot) \text{ is a continuous function}\},$$

does not have a countable basis, and is not a Borel set, i.e.  $\mathbb{C} \notin \mathcal{B}_T$ . If  $\{x(t); t \in T\}$  is a stochastic process on a probability space  $(\Omega, \mathcal{F}, P)$ , then the probability  $P(\mathbb{C})$  need not be defined – it depends on the structure of  $(\Omega, \mathcal{F})$  and on how complicated  $x$  is in itself, as a function on  $\Omega$ . In particular, even if  $P(\mathbb{C})$  is

defined, it is not uniquely determined by the finite-dimensional distributions. To see this, take a process on a sufficiently rich sample space, i.e. one that contains enough sample points  $\omega$ , and suppose we have defined a stochastic process  $\{x(t); t \in \mathbb{R}\}$ , which has, with probability one, continuous sample paths. Then  $x$  has a certain family of finite-dimensional distributions. Now, take a random time,  $\tau$ , independent of  $x$ , and with continuous distribution, for example an exponential distribution.<sup>1</sup> Then define a new process  $\{y(t); t \in \mathbb{R}\}$ , such that

$$\begin{aligned} y(t) &= x(t), & \text{for } t \neq \tau, \\ y(\tau) &= x(\tau) + 1. \end{aligned}$$

Then  $y$  has the same finite-dimensional distributions as  $x$  but its sample functions are always discontinuous at  $\tau$ .

### 2.2.1.1 Equivalence

In the constructed example, the two processes  $x$  and  $y$  differ only at a single point  $\tau$ , and as we constructed  $\tau$  to be random with continuous distribution, we have

$$P(x(t) = y(t)) = 1, \quad \text{for all } t. \quad (2.1)$$

Two processes  $x$  and  $y$  which satisfy (2.1) are called *equivalent*. The sample paths of two equivalent process always coincide, with probability one, when looked at a fixed, pre-determined time point. (In the example above, the time  $\tau$  where they differed was random.)

### 2.2.1.2 Separability

The annoying fact that a stochastic process can fail to fulfill some natural regularity condition, such as continuity, even if it by all natural standards should be regular, can be partly neutralized by the concept of *separability*, introduced by Doob. It uses the approximation by sets with countable basis mentioned in Section 1.3.3. Loosely speaking, a process  $\{x(t), t \in \mathbb{R}\}$  is separable in an interval  $I$  if there exists a countable set of  $t$ -values  $T = \{t_k\} \subset I$  such that the process, with probability one, does not behave more irregularly on  $I$  than it does already on  $T$ . An important consequence is that for all  $t$  in the interior of  $I$ , there are sequences  $\tau_1 < \tau_2 < \dots < \tau_n \uparrow t$  and  $\tau'_1 > \tau'_2 > \dots > \tau'_n \downarrow t$  such that, with probability one,

$$\liminf_{n \rightarrow \infty} x(\tau_n) = \liminf_{\tau \uparrow t} x(\tau) \leq \limsup_{\tau \uparrow t} x(\tau) = \limsup_{n \rightarrow \infty} x(\tau_n),$$

with a similar set of relations for the sequence  $\tau'_n$ . Hence, if the process is continuous on any discrete set of points then it is continuous. Every process has an equivalent separable version; see [11].

<sup>1</sup>This is where it is necessary that  $\Omega$  is rich enough so we can define an independent  $\tau$ .

### 2.2.2 Conditions for sample function continuity

The finite-dimensional distributions of any two equivalent processes  $x(t)$  and  $y(t)$  are always the same – show that as an exercise. We shall now see under what conditions, on the finite-dimensional distribution functions, we can assume a stochastic process to have continuous sample paths. Conditions will be given both in terms of the bivariate distributions directly, and in terms of probabilistic bounds on the process increments. As we have seen, one has to be satisfied if, among all equivalent processes, one can find one which has continuous sample paths.

**Theorem 2:1** *Let  $\{x(t); 0 \leq t \leq 1\}$  be a given stochastic process. If there exist two non-decreasing functions,  $g(h)$  and  $q(h)$ ,  $0 \leq h \leq 1$ , such that*

$$\sum_{n=1}^{\infty} g(2^{-n}) < \infty \quad \sum_{n=1}^{\infty} 2^n q(2^{-n}) < \infty,$$

and, for all  $t < t+h$  in  $[0, 1]$ ,

$$P(|x(t+h) - x(t)| \geq g(h)) \leq q(h), \quad (2.2)$$

then there exists an equivalent stochastic process  $y(t)$  whose sample paths are, with probability one, continuous on  $[0, 1]$ .

**Proof:** Start with the process  $x(t)$  with given finite-dimensional distributions. Such a process exists, and what is questioned is whether it has continuous sample functions if its bivariate distributions satisfy the conditions in the theorem. We shall now explicitly construct a process  $y(t)$ , equivalent to  $x(t)$ , and with continuous sample paths. Then  $y(t)$  will automatically have the same finite-dimensional distributions as  $x(t)$ . The process  $y(t)$  shall be constructed as the limit of a sequence of piecewise linear functions  $x_n(t)$ , which have the correct distribution at the dyadic time points of order  $n$ ,

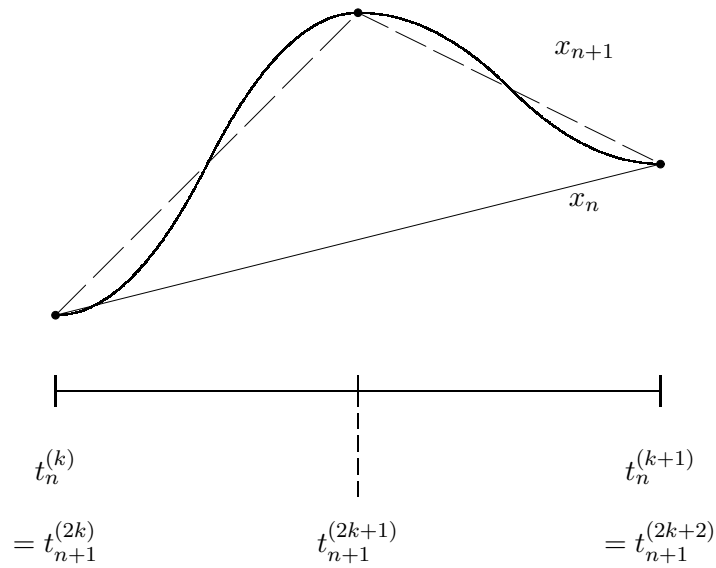
$$t_n^{(k)} = k/2^n, \quad k = 0, 1, \dots, 2^n; \quad n = 1, 2, \dots$$

Define the process  $x_n$  equal to  $x$  at the dyadic points,

$$x_n(t) = x(t), \quad \text{for } t = t_n^{(k)}, \quad k = 0, 1, \dots, 2^n,$$

and let it be linear between these points; see Figure 2.1.

Then we can estimate the maximal distance between two successive approximations. As is obvious from the figure, the maximal difference between two successive approximations for  $t$  between  $t_n^{(k)}$  and  $t_n^{(k+1)}$ , occurs in the middle



**Figure 2.1:** *Successive approximations with piecewise linear functions.*

of the interval, and hence

$$\begin{aligned}
 |x_{n+1}(t) - x_n(t)| &\leq \left| x(t_{n+1}^{(2k+1)}) - \frac{1}{2} \left( x(t_n^{(k)}) + x(t_n^{(k+1)}) \right) \right| \\
 &\leq \frac{1}{2} \left| x(t_{n+1}^{(2k+1)}) - x(t_{n+1}^{(2k)}) \right| + \frac{1}{2} \left| x(t_{n+1}^{(2k+2)}) - x(t_{n+1}^{(2k+1)}) \right| \\
 &= \frac{1}{2}A + \frac{1}{2}B, \text{ say.}
 \end{aligned}$$

The tail distribution of the maximal difference between two successive approximations,

$$M_n^{(k)} = \max_{t_n^{(k)} \leq t \leq t_n^{(k+1)}} |x_{n+1}(t) - x_n(t)| \leq \frac{1}{2}A + \frac{1}{2}B,$$

can therefore be estimated by

$$P(M_n^{(k)} \geq c) \leq P(A \geq c) + P(B \geq c),$$

since if  $M_n^{(k)} \geq c$ , then either  $A \geq c$  or  $B \geq c$ , or both.

Now take  $c = g(2^{-n-1})$  and use the bound (2.2), to get

$$P(M_n^{(k)} \geq g(2^{-n-1})) \leq 2q(2^{-n-1}),$$

for each  $k = 0, 1, \dots, 2^n - 1$ . By Boole's inequality<sup>2</sup> we get, since there are  $2^n$  intervals,

$$\begin{aligned} P\left(\max_{0 \leq t \leq 1} |x_{n+1}(t) - x_n(t)| \geq g(2^{-n-1})\right) \\ = P\left(\bigcup_{k=0}^{2^n-1} M_n^{(k)} \geq g(2^{-n-1})\right) \leq 2^{n+1} q(2^{-n-1}). \end{aligned}$$

Now  $\sum_n 2^{n+1} q(2^{-n-1}) < \infty$  by assumption, and then the Borel-Cantelli lemma (see Exercises in Appendix B), gives that, with probability one, only finitely many of the events

$$\max_{0 \leq t \leq 1} |x_{n+1}(t) - x_n(t)| \geq g(2^{-n})$$

occur. This means that there is a set  $\Omega_0$  with  $P(\Omega_0) = 1$ , such that for every outcome  $\omega \in \Omega_0$ , from some integer  $N$  (depending of the outcome,  $N = N(\omega)$ ) and onwards, ( $n \geq N$ ),

$$\max_{0 \leq t \leq 1} |x_{n+1}(t) - x_n(t)| < g(2^{-n}).$$

First of all, this shows that there exists a limiting function  $y(t)$  for all  $\omega \in \Omega_0$ ; the condition (B.4) for almost sure convergence, given in Appendix B, says that  $\lim_{n \rightarrow \infty} x_n(t)$  exists with probability one.

It also shows that the convergence is uniform: for  $\omega \in \Omega_0$  and  $n \geq N$ ,  $m > 0$ ,

$$\begin{aligned} |x_{n+m}(t) - x_n(t)| \\ \leq |x_{n+1}(t) - x_n(t)| + |x_{n+2}(t) - x_{n+1}(t)| + \dots + |x_{n+m}(t) - x_{n+m-1}(t)| \\ \leq \sum_{j=0}^{m-1} g(2^{-n-j}) \leq \sum_{j=0}^{\infty} g(2^{-n-j}). \end{aligned}$$

Letting  $m \rightarrow \infty$ , so that  $x_{n+m}(t) \rightarrow y(t)$ , and observing that the inequalities hold for all  $t \in [0, 1]$ , we get that

$$\max_{0 \leq t \leq 1} |y(t) - x_n(t)| \leq \sum_{j=0}^{\infty} g(2^{-n-j}) = \sum_{j=n}^{\infty} g(2^{-j}).$$

Since this bound tends to 0 as  $n \rightarrow \infty$ , we have the uniform convergence, and since all  $x_n$  are continuous functions, we also have that  $y$  is continuous for all  $\omega \in \Omega_0$ . For  $\omega \notin \Omega_0$ , define  $y(t) \equiv 0$ , making  $y$  a continuous function for all  $\omega \in \Omega$ .

---

<sup>2</sup>  $P(\cup A_k) \leq \sum P(A_k)$ .

It remains to prove that  $x$  and  $y$  are equivalent, i.e.  $P(x(t) = y(t)) = 1$ , for all  $t \in [0, 1]$ . For that sake, take any  $t \in [0, 1]$  and find a sequence of dyadic numbers  $t_n^{(k_n)} \rightarrow t$  such that

$$t_n^{(k_n)} \leq t < t_n^{(k_n)} + 2^{-n}.$$

Since both  $g(h)$  and  $q(h)$  are non-decreasing, we have from (2.2),

$$\begin{aligned} P\left(\left|x(t_n^{(k_n)}) - x(t)\right| \geq g(2^{-n})\right) &\leq P\left(\left|x(t_n^{(k_n)}) - x(t)\right| \geq g(t - t_n^{(k_n)})\right) \\ &\leq q(t - t_n^{(k_n)}) \leq q(2^{-n}). \end{aligned}$$

Adding over  $n$  gives

$$\sum_{n=1}^{\infty} P\left(\left|x(t_n^{(k_n)}) - x(t)\right| \geq g(2^{-n})\right) \leq \sum_{n=1}^{\infty} q(2^{-n}) < \infty,$$

and it follows from the Borel-Cantelli lemma that it can happen only finitely many times that  $\left|x(t_n^{(k_n)}) - x(t)\right| \geq g(2^{-n})$ . Since  $g(2^{-n}) \rightarrow 0$  as  $n \rightarrow \infty$ , we have proved that  $x(t_n^{(k_n)}) \rightarrow x(t)$  with probability one. Further, since  $y(t)$  is continuous,  $y(t_n^{(k_n)}) \rightarrow y(t)$ . But  $x(t_n^{(k_n)}) = y(t_n^{(k_n)})$ , and therefore the two limits are equal, with probability one, as was to be proved.  $\square$

The theorem says that for each process  $x(t)$  that satisfies the conditions there exists at least one other equivalent process  $y(t)$  with continuous sample paths, and with exactly the same finite-dimensional distributions. Of course it seems unnecessary to start with  $x(t)$  and immediately change to an equivalent continuous process  $y(t)$ . In the future we assume that we only have the continuous version, whenever the sufficient conditions for sample function continuity are satisfied.

### 2.2.2.1 Special conditions for continuity

Theorem 2:1 is simple to use, since it depends only on the distribution of the increments of the process, and involves only bivariate distributions. For special processes conditions that put bounds on the moments of the increments are even simpler to use. One such is the following.

**Corollary 2.1** *If there exist constants  $C$ , and  $r > p > 0$ , such that for all small enough  $h > 0$ ,*

$$E(|x(t+h) - x(t)|^p) \leq C \frac{|h|}{|\log |h||^{1+r}} \quad (2.3)$$

*then the condition in Theorem 2:1 is satisfied and the process has, with probability one, continuous sample paths.*

Note, that many processes satisfy a stronger inequality than (2.3), namely

$$E(|x(t+h) - x(t)|^p) \leq C|h|^{1+c} \quad (2.4)$$

for some constants  $C$ , and  $c > 0, p > 0$ . Then (2.3) is automatically satisfied with any  $r > p$ , and the process has, with probability one, continuous sample paths.

**Proof:** Markov's inequality, a generalization of Chebysjev's inequality, states that for all random variables  $U$ ,  $P(|U| \geq \lambda) \leq E(|U|^p)/\lambda^p$ . Apply the theorem with  $g(h) = |\log |h||^{-b}$ ,  $1 < b < r/p$ . One gets,

$$P(|x(t+h) - x(t)| > g(h)) \leq \frac{C|h|}{|\log |h||^{1+r-bp}}.$$

Since  $b > 1$ , one has  $\sum g(2^{-n}) = \sum \frac{1}{(n \log 2)^b} < \infty$ , and, with

$$q(h) = C|h|/|\log |h||^{1+r-bp},$$

and  $1 + r - bp > 1$ ,

$$\sum 2^n q(2^{-n}) = \sum \frac{C}{(n \log 2)^{1+r-bp}} < \infty,$$

which proves the assertion.  $\square$

**Example 2:1** We show that the Wiener process  $W(t)$  has, with probability one, continuous sample paths. In the standard Wiener process, the increment  $W(t+h) - W(t), h > 0$ , is Gaussian with mean 0 and variance  $h$ . Thus,

$$E(|W(t+h) - W(t)|^p) = C|h|^{p/2},$$

with  $C = E(|U|^p) < \infty$ , for a standard normal variable  $U$ , giving the moment bound

$$E(|W(t+h) - W(t)|^4) = C|h|^2 < \frac{|h|}{|\log |h||^6},$$

for small  $h$ . We see that condition (2.3) in the corollary is satisfied with  $r = 5 > 4 = p$ . Condition (2.4) is satisfied with  $p = 3$ ,  $c = 1/2$ .

### 2.2.2.2 Continuity of stationary processes

Stationary processes have constant mean and a covariance function

$$r(t) = \text{Cov}(x(s+t), x(s)),$$

which is a function only of the time lag  $t$ . Since the increments have variance

$$E((x(t+h) - x(t))^2) = 2(r(0) - r(h)), \quad (2.5)$$

it is clear that continuity conditions can be formulated in terms of the covariance function. Equivalent conditions can be formulated by means of the spectral distribution function  $F(\omega)$ , which is such that

$$r(t) = \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega).$$

A first immediate consequence of (2.5) is that  $x(t+h) \xrightarrow{q.m.} x(t)$  as  $h \rightarrow 0$  if and only if the covariance function  $r(t)$  is continuous for  $t = 0$ . For sample function continuity, a sufficient condition in terms of the covariance function follows directly from Corollary 2.1.

**Theorem 2:2** *If  $r(t)$  is the covariance function of a stationary stochastic process  $x(t)$ , such that, as  $t \rightarrow 0$ ,*

$$r(t) = r(0) - O\left(\frac{|t|}{|\log |t||^q}\right), \quad (2.6)$$

*for some  $q > 3$ , then  $x(t)$  has<sup>3</sup> continuous sample functions.<sup>4</sup>*

### 2.2.2.3 Continuity of stationary Gaussian processes

For stationary Gaussian processes the conditions for sample function continuity can be considerably weakened, to require slightly more than just continuity of the covariance function. We state the sufficient condition both in terms of the covariance function and in terms of the spectrum, and to this end, we formulate an analytic lemma, the proof of which can be found in [9, Sect. 9.3].

**Lemma 2.1** *a) If, for some  $a > 0$ ,*

$$\int_0^{\infty} (\log(1 + \omega))^a dF(\omega) < \infty, \quad (2.7)$$

*then*

$$r(t) = r(0) - O\left(|\log |t||^{-b}\right), \quad \text{as } t \rightarrow 0, \quad (2.8)$$

*for any  $b \leq a$ .*

*b) If (2.8) holds for some  $b > 0$ , then (2.7) is satisfied for any  $a < b$ .*

**Theorem 2:3** *A stationary Gaussian process  $x(t)$  has, with probability one, continuous sample paths if, for some  $a > 3$ , any of the following conditions is satisfied:*

$$r(t) = r(0) - O\left(|\log |t||^{-a}\right), \quad \text{as } t \rightarrow 0, \quad (2.9)$$

$$\int_0^{\infty} (\log(1 + \omega))^a dF(\omega) < \infty. \quad (2.10)$$

<sup>3</sup>Or rather "Is equivalent to a process that has ..."

<sup>4</sup>The notation  $f(x) = g(x) + o(h(x))$  as  $x \rightarrow 0$  means that  $|(f(x) - g(x))/h(x)|$  is bounded by some finite constant  $C$  as  $x \rightarrow 0$ .

**Proof:** In a stationary Gaussian process,  $x(t+h) - x(t)$  has a normal distribution with mean zero and variance  $\sigma_h^2 = 2(r(0) - r(h))$ , where by assumption

$$\sigma_h \leq \frac{C}{|\log |h||^{a/2}}$$

for some constant  $C > 0$ . Writing  $\Phi(x) = \int_{-\infty}^x \phi(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$  for the standard normal distribution, we have

$$P(|x(t+h) - x(t)| > g(h)) = 2 \left\{ 1 - \Phi \left( \frac{g(h)}{\sigma_h} \right) \right\}$$

Now, take  $g(h) = |\log |h|/\log 2|^{-b}$ , where  $b$  is chosen so that  $1 < b < (a-1)/2$ , which is possible since  $a > 3$  by assumption. From the bound (1.13) of the normal distribution tail,  $1 - \Phi(x) \leq \phi(x)/x$ , we then get

$$\begin{aligned} P(|x(t+h) - x(t)| > g(h)) &\leq 2 \left\{ 1 - \Phi \left( \frac{g(h)|\log |h||^{a/2}}{C} \right) \right\} \\ &\leq \frac{2C}{g(h)|\log |h||^{a/2}} \phi \left( \frac{g(h)|\log |h||^{a/2}}{C} \right) \\ &= q(h), \text{ say.} \end{aligned}$$

The reader should complete the proof and show that

$$\sum g(2^{-n}) < \infty \quad \text{and} \quad \sum 2^n q(2^{-n}) < \infty,$$

and then apply Theorem 2:1 to see that (2.9) is sufficient.

Lemma 2.1 shows that also (2.10) is sufficient for sample function continuity for a Gaussian process.  $\square$

**Example 2:2** Any stationary process with

$$r(t) = r(0) - C|t|^\alpha + o(|t|^\alpha),$$

some  $C > 0$ , as  $t \rightarrow 0$ ,<sup>5</sup> has continuous sample paths if  $1 < \alpha \leq 2$ . If it furthermore is a Gaussian process it is continuous if  $0 < \alpha \leq 2$ .

**Remark 2:1** *The sufficient conditions for sample function continuity given in the theorems are satisfied for almost all covariance functions that are encountered in applied probability. But for Gaussian processes, even the weak condition (2.9) for  $a > 3$ , can be relaxed to require only that  $a > 1$ , which is very close to being necessary; see [9, Sect. 9.5].*

*Gaussian stationary processes which are not continuous behave with necessity very badly, and it can be shown that the sample functions are, with probability one, unbounded in any interval. This was shown by Belyaev [4] but is also a consequence of a theorem by Dobrushin [10], see also [9, Ch. 9.5].*

<sup>5</sup>At this stage you should convince yourself that  $\alpha > 2$  impossible?

### 2.2.3 Probability measures on $\mathbb{C}[0, 1]$

We can now complete the table at the end of Section 1.3.1 and define probabilities on  $\mathbb{C}[0, 1]$ , the space of continuous functions on the interval  $[0, 1]$ . A stochastic process  $\{x(t)\}_{0 \leq t \leq 1}$  on a probability space  $(\Omega, \mathcal{F})$  has its realizations in  $\mathbb{R}^{[0,1]}$ . If the finite-dimensional distributions of  $x$  satisfy any of the sufficient conditions for sample function continuity, either  $x(t)$  or an equivalent process  $y(t)$  has, with probability one, continuous sample functions, and hence has its realizations in  $\mathbb{C}[0, 1] \subset \mathbb{R}^{[0,1]}$ .

In the same way as the finite-dimensional distributions define a probability measure on  $(\mathbb{R}^{[0,1]}, \mathcal{B}_{[0,1]})$ , assigning probabilities to all Borel sets, we can now define a probability measure on  $\mathbb{C}[0, 1]$ . The question is, what is the  $\sigma$ -field of events which get probability? In fact, we can take the simplest choice,

$$\mathcal{B} = \mathcal{B}_{[0,1]} \cap \mathbb{C}[0, 1] = \{B \cap \mathbb{C}[0, 1]; B \in \mathcal{B}_{[0,1]}\},$$

i.e. take those parts of the Borel-sets which intersect  $\mathbb{C}[0, 1]$ .

**Theorem 2:4** *If  $\{F_{\mathbf{t}^n}\}_{n=1}^{\infty}$  is a family of finite-dimensional distributions that satisfy any of the sufficient conditions for sample functions continuity, then there exists a probability measure on  $(\mathbb{C}[0, 1], \mathcal{B})$  such that the co-ordinate process*

$$\{x(t)\}_{0 \leq t \leq 1}$$

*has the given finite-dimensional distributions.*

#### 2.2.3.1 Open sets in $\mathbb{C}[0, 1]$

The family  $\mathcal{B}$  can be described alternatively in terms of open sets in  $\mathbb{C}[0, 1]$ . Take a continuous function  $x(t) \in \mathbb{C}[0, 1]$ . By an  $\epsilon$ -surrounding of  $x$  we mean the set of functions which are in an  $\epsilon$ -band around  $x$ ,

$$\left\{ y \in \mathbb{C}[0, 1]; \max_{0 \leq t \leq 1} |y(t) - x(t)| < \epsilon \right\}.$$

A set of functions  $A \subseteq \mathbb{C}[0, 1]$  is called *open* if for every  $x \in A$ , there is an  $\epsilon$ -surrounding of  $x$  that is completely in  $A$ . The open sets in  $\mathbb{C}[0, 1]$  generate a  $\sigma$ -field, the smallest  $\sigma$ -field that contains all open sets, and that  $\sigma$ -field is exactly  $\mathcal{B}$ .

## 2.3 Derivatives, tangents, and other characteristics

### 2.3.1 Differentiability

#### 2.3.1.1 General conditions

When is a continuous stochastic process differentiable in the sense that its sample functions are continuously differentiable? The answer can be given

as conditions similar to those for sample function continuity, but now with bounds on the second order differences. By pasting together piecewise linear approximations by means of smooth arcs, one can prove the following theorem; see [9, Sect. 4.3].

**Theorem 2:5** *a) Suppose the stochastic process  $\{x(t); 0 \leq t \leq 1\}$  satisfies the conditions for sample function continuity in Theorem 2:1. If, furthermore, for all  $t - h$  and  $t + h$  in  $[0, 1]$ ,*

$$P(|x(t+h) - 2x(t) + x(t-h)| \geq g_1(h)) \leq q_1(h), \quad (2.11)$$

where  $g_1$  and  $q_1$  are two non-decreasing functions, such that

$$\sum_{n=1}^{\infty} 2^n g_1(2^{-n}) < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} 2^n q_1(2^{-n}) < \infty,$$

then there exists an equivalent process  $\{y(t); 0 \leq t \leq 1\}$  with continuously differentiable sample paths.

*b) The sufficient condition in (a) is satisfied if*

$$E(|x(t+h) - 2x(t) + x(t-h)|^p) \leq \frac{K|h|^{1+p}}{|\log|h||^{1+r}}, \quad (2.12)$$

for some constants  $p < r$  and  $K$ .

*c) Many processes satisfy a stronger inequality than (2.12), namely*

$$E(|x(t+h) - 2x(t) + x(t-h)|^p) \leq C|h|^{1+p+c}, \quad (2.13)$$

for some constants  $C$ , and  $c > 0$ ,  $p > 0$ . Then (2.12) is satisfied, and the process has, with probability one, continuously differentiable sample paths.

In Section 2.1 we mentioned a condition for quadratic mean (q.m.) differentiability of a stationary process. One may ask: What is the relation between the q.m.-derivative and the sample function derivative? They are both limits of the differential quotient  $(x(t+h) - x(t))/h$  as  $h \rightarrow 0$ . Now, it is easy to prove that if the limit exists in both quadratic mean and as a sample function limit with probability one, then the two limits are equal (also with probability one), and hence the two derivative processes are equivalent, and have the same finite-dimensional distributions. It follows that the covariance function of the derivative  $\{x'(t), t \in \mathbb{R}\}$  is  $r_{x'}(t) = -r_x''(t)$ .

2.3.1.2 Differentiable Gaussian processes

In order that a stationary Gaussian process has continuously differentiable sample functions it is necessary that its covariance function has a smooth higher order Taylor expansion; cf. condition (2.9).

As we shall see in the following theorem, demanding just slightly more than a finite second spectral moment guarantees that a Gaussian process has continuously differentiable sample paths. For a proof, the reader is referred to [9, Sect. 9.3].

**Theorem 2:6** *a) A stationary Gaussian process is continuously differentiable<sup>6</sup> if, for some  $a > 3$ , its covariance function has the expansion*

$$r(t) = r(0) - \frac{\omega_2 t^2}{2} + O\left\{\frac{t^2}{|\log |t||^a}\right\}, \tag{2.14}$$

where  $\omega_2 = -r''(0) < \infty$ .

*b) Condition (2.14) can be replaced by the condition that, for some  $a > 3$ ,*

$$\int_0^\infty \omega^2 (\log(1 + \omega))^a dF(\omega) < \infty. \tag{2.15}$$

As for sample function continuity, it can be shown that for continuously differentiable sample functions, it suffices that the constant  $a$  is greater than 1.

**Example 2:3** Condition (2.14) is easy to use for Gaussian processes. Most covariance functions used in practice have an expansion

$$r(t) = r(0) - \frac{\omega_2 t^2}{2} + O(|t|^a),$$

where  $a$  is an integer, either 3 or 4. Then the process is continuously differentiable. Processes with covariance function admitting an expansion  $r(t) = r(0) - C|t|^\alpha + o(|t|^\alpha)$  with  $\alpha < 2$  are not differentiable; they are not even differentiable in quadratic mean. An example is the Ornstein-Uhlenbeck process with  $r(t) = r(0)e^{-C|t|}$ .

As a final example of covariance conditions, we encourage the reader to prove a sufficient condition in terms of  $-r''(t)$ , the covariance function of the quadratic mean derivative.

**Theorem 2:7** *A stationary process  $x(t)$  is continuously differentiable if any of the following two conditions hold,*

---

<sup>6</sup>As usual, this means that there exists an equivalent process that, with probability one, has continuously differentiable sample functions.

- a)  $-r''(t) = -r''(0) - C|t|^\alpha + o(|t|^\alpha)$  with  $1 < \alpha \leq 2$ ,
- b) it is Gaussian and  $-r''(t) = -r''(0) - C|t|^\alpha + o(|t|^\alpha)$  with  $0 < \alpha \leq 2$ .

### 2.3.2 Jump discontinuities and Hölder conditions

What type of discontinuities are possible for stochastic processes which do not have continuous sample functions? For example, how do we know that the Poisson process has sample functions that increase only with jumps of size one? We would like to have a condition that guarantees that only simple discontinuities are possible, and such a condition exists, with restriction on the increment over two adjacent interval. Similarly, for a process which is not continuously differentiable, how far from differentiable are the sample functions.

#### 2.3.2.1 Jump discontinuities

The proof of the following theorem is indicated in [9, Sec. 4.4].

**Theorem 2:8** *If there are positive constants  $C, p, r$ , such that for all  $s, t$  with  $0 \leq t < s < t + h \leq 1$ ,*

$$E \{|x(t+h) - x(s)|^p \cdot |x(s) - x(t)|^p\} \leq C|h|^{1+r} \quad (2.16)$$

*then the process  $\{x(t); 0 \leq t \leq 1\}$  has, with probability one,<sup>7</sup> sample functions with at most jump discontinuities, i.e.*

$$\lim_{t \downarrow t_0} x(t) \quad \text{and} \quad \lim_{t \uparrow t_0} x(t)$$

*exist for every  $t_0 \in [0, 1]$ .*

**Example 2:4** The Poisson process with intensity  $\lambda$  has independent increments and hence

$$\begin{aligned} & E \left\{ |x(t+h) - x(s)|^2 \cdot |x(s) - x(t)|^2 \right\} \\ &= E(|x(t+h) - x(s)|^2) \cdot E(|x(s) - x(t)|^2) \\ &= (\lambda(t+h-s) + (\lambda(t+h-s))^2) \cdot (\lambda(s-t) + (\lambda(s-t))^2) \leq C\lambda^2 h^2. \end{aligned}$$

The conditions of the theorem are obviously satisfied.

---

<sup>7</sup>As usual, this means that there exists an equivalent process with this property.

### 2.3.2.2 Hölder continuity and the continuity modulus

How large are the increments in a non-differentiable process? In fact, moment bounds on the increments give precise estimates for the distribution of the *continuity modulus*  $\omega_x(h)$  of the sample functions. This is defined as

$$\omega_x(h) = \sup_{|s-t| \leq h} |x(s) - x(t)| \quad (2.17)$$

where the supremum is taken over  $0 \leq s, t \leq 1$ . (For a continuous process the supremum in (2.17) can be taken over the countable rationals, so it is a well defined random variable.)

Functions for which there exist constants  $A, a$  such that for all  $t, t+h \in [0, 1]$

$$|x(t+h) - x(t)| \leq A|h|^a,$$

are said to be *Hölder continuous* of order  $a$ , and to satisfy a *Lipschitz condition* of order  $a$ .

In this section we shall present stochastic estimates of the continuity modulus for a stochastic process, and also give a sufficient condition for a stochastic process to be Hölder continuous of order  $a < r/p \leq 1$ .

**Theorem 2:9** *If there are constants  $C, p \geq r > 0$  such that*

$$E\{|x(t+h) - x(t)|^p\} \leq C|h|^{1+r} \quad (2.18)$$

*then there exists a random variable  $A$  with  $P(A < \infty) = 1$ , such that the continuity modulus satisfies an inequality*

$$\omega_x(h) \leq A|h|^a, \quad \text{for all } h > 0,$$

*for all  $0 < a < r/p$ .*

**Proof:** First examine the increments over the dyadic numbers

$$t_n^{(k)} = k/2^n, \quad k = 0, 1, \dots, 2^n; \quad n = 1, 2, \dots$$

Take an  $a < r/p$  and write  $\delta = 2^{-a}$ . Then, by Markov's inequality,

$$\begin{aligned} P(|x(t_n^{(k+1)}) - x(t_n^{(k)})| > \delta^n) &\leq \frac{E(|x(t_n^{(k+1)}) - x(t_n^{(k)})|^p)}{\delta^{np}} \\ &\leq \frac{C(2^{-n})^{1+r}}{2^{-anp}} = \frac{C}{2^{n(1+r-ap)}}. \end{aligned}$$

From Boole's inequality, summing over  $n$ , we obtain in succession,

$$\begin{aligned} P\left(\max_{0 \leq k \leq 2^n - 1} |x(t_n^{(k+1)}) - x(t_n^{(k)})| > \delta^n\right) &\leq \frac{C2^n}{2^{n(1+r-ap)}} \\ &= \frac{C}{2^{n(r-ap)}}, \end{aligned}$$

$$\sum_{n=0}^{\infty} P\left(\max_{0 \leq k \leq 2^n - 1} |x(t_n^{(k+1)}) - x(t_n^{(k)})| > \delta^n\right) \leq \sum_{n=0}^{\infty} \frac{C}{2^{n(r-ap)}} < \infty,$$

since  $r - ap > 0$ . The Borel-Cantelli lemma gives that only finitely many events

$$A_n = \left\{ \max_{0 \leq k \leq 2^n - 1} |x(t_n^{(k+1)}) - x(t_n^{(k)})| > \delta^n \right\}$$

occur, which means that there exists a random index  $\nu$  such that for all  $n > \nu$ ,

$$|x(t_n^{(k+1)}) - x(t_n^{(k)})| \leq \delta^n \quad \text{for all } k = 0, 1, \dots, 2^n - 1. \quad (2.19)$$

Next we estimate the increment from a dyadic point  $t_n^{(k)}$  to an arbitrary point  $t$ . To that end, take any  $t \in [t_n^{(k)}, t_n^{(k+1)})$ , and consider its dyadic expansion, ( $\alpha_m = 0$  or  $1$ ),

$$t = t_n^{(k)} + \sum_{m=1}^{\infty} \frac{\alpha_m}{2^{n+m}}.$$

Summing all the inequalities (2.19), we obtain that the increment from  $t_n^{(k)}$  to  $t$  is bounded (for  $n > \nu$ ),

$$|x(t) - x(t_n^{(k)})| \leq \sum_{m=1}^{\infty} \delta^{n+m} = \frac{\delta^{n+1}}{1 - \delta}. \quad (2.20)$$

The final estimate relates  $t + h$  to the dyadic points. Let  $\nu < \infty$  be the random index just found to exist. Then, suppose  $h < 2^{-\nu}$  and find  $n, k$  such that  $2^{-n} \leq h < 2^{-n+1}$  and  $k/2^n \leq t < (k+1)/2^n$ . We see that  $n > \nu$  and

$$t_n^{(k)} \leq t < t_n^{(k+1)} < t + h \leq t_n^{(k+\ell)},$$

where  $\ell$  is either 2 or 3. As above, we obtain

$$|x(t+h) - x(t_n^{(k+1)})| \leq \frac{\delta^{n+1}}{1 - \delta} + \delta^n. \quad (2.21)$$

Summing the three estimates (2.19-2.21), we see that

$$|x(t+h) - x(t)| \leq \delta^n + \frac{\delta^{n+1}}{1 - \delta} + \frac{\delta^{n+1}}{1 - \delta} + \delta^n = \frac{2}{1 - \delta} (2^{-n})^a \leq \frac{2}{1 - \delta} h^a,$$

for  $2^{-n} \leq h < 2^{-\nu}$ . For  $h \geq 2^{-\nu}$  it is always true that

$$|x(t+h) - x(t)| \leq M \leq \frac{M}{2^{-\nu}} h^a$$

for some random  $M$ . If we take  $A = \max(M2^\nu, 2/(1 - \delta))$ , we complete the proof by combining the last two inequalities, to obtain  $|x(t+h) - x(t)| \leq Ah^a$ .  $\square$

**Example 2:5** For the Wiener process,

$$E(|x(t+h) - x(t)|^p) = C_p |h|^{p/2} = C_p |h|^{1+(p/2-1)},$$

so the Wiener process is Hölder continuous of order  $a < (p/2 - 1)/p = 1/2 - 1/p$  for every  $p > 0$ . This means that it is Hölder continuous of all orders  $a < 1/2$ .

In the next section we shall investigate the existence of tangents of a predetermined level. Then we shall need a small lemma on the size of the continuity modulus of a continuous stochastic process.

**Lemma 2.2** *Let  $x(t)$  be a stochastic process with continuous sample functions in  $0 \leq t \leq 1$ , and let  $\omega_x(h)$  be its (random) continuity modulus, defined by (2.17). Then, to every  $\epsilon > 0$  there is a (deterministic) function  $\omega_\epsilon(h)$  such that  $\omega_\epsilon(h) \downarrow 0$  as  $h \downarrow 0$ , and*

$$P(\omega_x(h) < \omega_\epsilon(h) \text{ for } 0 < h \leq 1) > 1 - \epsilon.$$

**Proof:** The sample continuity of  $x(t)$  says that the continuity modulus tends to 0 for  $h \rightarrow 0$ ,

$$\lim_{h \rightarrow 0} P(\omega_x(h) < c) = 1$$

for every fixed  $c > 0$ . Take a sequence  $c_1 > c_2 \dots > c_n \downarrow 0$ . For a given  $\epsilon > 0$  we can find a decreasing sequence  $h_n \downarrow 0$  such that

$$P(\omega_x(h_n) < c_n) > 1 - \epsilon/2^{n+1}.$$

Since  $\omega_x(h)$  is non-increasing as  $h$  decreases, then also

$$P(\omega_x(h) < c_n \text{ for } 0 < h \leq h_n) > 1 - \epsilon/2^{n+1},$$

for  $n = 1, 2, \dots$ . Summing the exceptions, we get that

$$P(\omega_x(h) < c_n \text{ for } 0 < h \leq h_n \text{ and } n = 1, 2, \dots) > 1 - \sum_{n=1}^{\infty} \epsilon/2^{n+1} = 1 - \epsilon/2. \tag{2.22}$$

Now we can define the deterministic function  $\omega_\epsilon(h)$  from the sequences  $c_n$  and  $h_n$ . Take

$$\omega_\epsilon(h) = \begin{cases} c_0 & \text{for } h_1 < h \leq 1 \\ c_n & \text{for } h_{n+1} < h \leq h_n, \quad n = 1, 2, \dots \end{cases}$$

If we take  $c_0$  large enough to make

$$P(\omega_x(h) < c_0 \text{ for } h_1 < h \leq 1) > 1 - \epsilon/2,$$

and combine with (2.22) we get the desired estimate. □

### 2.3.2.3 Tangencies

We start with a theorem due to E.V. Bulinskaya on the non-existence of tangents of a pre-specified level.

**Theorem 2:10** Suppose the density  $f_t(x)$  of  $x(t)$  is bounded for  $0 \leq t \leq 1$ ,

$$f_t(x) \leq c_0 < \infty,$$

and that  $x(t)$  has, with probability one, continuously differentiable sample paths. Then,

a) for any level  $u$ , the probability is zero that there exists a  $t \in [0, 1]$  such that simultaneously  $x(t) = u$  and  $x'(t) = 0$ , i.e. there exists no points where  $x(t)$  has a tangent on the level  $u$  in  $[0, 1]$ ,

b) there are only finitely many  $t \in [0, 1]$  for which  $x(t) = u$ .

**Proof:** a) By assumption,  $x(t)$  has continuously differentiable sample paths. We identify the location of those  $t$ -values for which  $x'(t) = 0$  and  $x(t)$  is close to  $u$ . For that sake, take an integer  $n$  and a constant  $h > 0$ , let  $H_\tau$  be the event

$$H_\tau = \{x'(\tau) = 0\} \cap \{|x(\tau) - u| \leq h\},$$

and define, for  $k = 1, 2, \dots, n$ ,

$$A_h = \{H_t \text{ occurs for at least one } t \in [0, 1]\},$$

$$A_h(k, n) = \{H_\tau \text{ occurs for at least one } \tau \in [\frac{k-1}{n}, \frac{k}{n}]\},$$

$$A_h = \cup_{k=1}^n A_h(k, n).$$

Now take a sample function that satisfies the conditions for  $A_h(k, n)$  and let  $\omega_{x'}$  be the continuity modulus of its derivative. For such a sample function,

$$x(k/n) = x(\tau) + (k/n - \tau)x'(\tau + \theta(k/n - \tau)),$$

for some  $\theta, 0 \leq \theta \leq 1$ , and hence, on  $A_h(k, n)$

$$|x(k/n) - u| \leq h + n^{-1}\omega_{x'}(n^{-1}). \quad (2.23)$$

We now use Lemma 2.2 to bound  $\omega_{x'}$ . If  $\omega(t) \downarrow 0$  as  $t \downarrow 0$ , let  $B_\omega$  denote the sample functions for which  $\omega_{x'}(t) \leq \omega(t)$  for all  $t$  in  $[0, 1]$ . By the lemma, given  $\epsilon > 0$ , there exists at least one function  $\omega_\epsilon(t) \downarrow 0$  such that  $P(B_{\omega_\epsilon}) > 1 - \epsilon/2$ . For outcomes satisfying (2.23) we use the bound  $\omega_\epsilon$ , and obtain

$$\begin{aligned} P(A_h) &\leq \sum_{k=1}^n P(A_h(k, n) \cap B_{\omega_\epsilon}) + (1 - P(B_{\omega_\epsilon})) \\ &\leq \sum_{k=1}^n P(|x(k/n) - u| \leq h + n^{-1}\omega_\epsilon(n^{-1})) + \epsilon/2 \\ &\leq 2nc_0(h + n^{-1}\omega_\epsilon(n^{-1})) + \epsilon/2, \end{aligned}$$

where  $c_0$  is the bounding constant for the density  $f_t(x)$ .

Since  $\omega_\epsilon(t) \rightarrow 0$  as  $t \rightarrow 0$ , we can select first an  $n_0$  and then an  $h_0$  to make  $P(A_{h_0}) \leq \epsilon$ . But if there exists a time point  $t$  for which  $x'(t) = 0$  and  $x(t) = u$  simultaneously, then certainly  $A_h$  has occurred for any  $h > 0$  and the event of interest has probability less than  $\epsilon$ , which was arbitrary. The probability of simultaneous occurrence is therefore 0 as stated.

b) To prove there are only a finite number of  $u$ -values in  $[0, 1]$ , assume, on the contrary, that there is an infinite sequence of points  $t_i \in [0, 1]$  with  $x(t_i) = u$ . There is then at least one limit point  $t_0 \in [0, 1]$  of  $\{t_i\}$  for which, by continuity,  $x(t_0) = u$ . Since the derivative of  $x(t)$  is assumed continuous, we must also have  $x'(t_0) = 0$ , and we have found a point where simultaneously  $x(t_0) = u$ ,  $x'(t_0) = 0$ . By (a), that event has probability 0.  $\square$

## 2.4 Quadratic mean properties a second time

Continuous or differentiable sample paths are what we expect to encounter in practice when we *observe* a stochastic process. To *prove* that a mathematical model for a random phenomenon has continuous or differentiable sample paths is a quite different matter. Much more simple is to base the stochastic analysis on correlation properties, which could be checked against data, at least in principle. Such second order properties are studied in quite some detail in elementary courses in stochastic processes, and we give here only some refinements and extra comments in addition to those in Section 2.1. We assume throughout in this section, as in most of the chapter, that the process  $x(t)$  has mean zero.

We first remind about the definition of convergence in quadratic mean of a sequence of random variables  $\{x_n\}$  with  $E(x_n^2) < \infty$  to a random variable  $x$ :

$$x_n \xrightarrow{q.m.} x \quad \text{if and only if} \quad E((x_n - x)^2) \rightarrow 0,$$

as  $n \rightarrow \infty$ ; see Appendix B. We shall use the Loève criterion (B.8) for quadratic mean convergence: the sequence  $x_n$  converges in quadratic mean if and only if

$$E(x_m \bar{x}_n) \text{ has a finite limit } c, \tag{2.24}$$

when  $m$  and  $n$  tend to infinity independently of each other.

### 2.4.1 Quadratic mean continuity

A stochastic process  $x(t)$  is continuous in quadratic mean (or  $L^2$ -continuous) at  $t$  if

$$x(t+h) \xrightarrow{q.m.} x(t)$$

when  $h \rightarrow 0$ , i.e. if  $E(|x(t+h) - x(t)|^2) \rightarrow 0$ . We formulate the condition for quadratic mean continuity in terms of the covariance function

$$r(s, t) = \text{Cov}(x(s), x(t)) = E(x(s) \cdot \overline{x(t)}),$$

for a, not necessarily stationary, process.

**Theorem 2:11** *A stochastic process  $x(t)$  with mean zero is continuous in quadratic mean at  $t_0$  if and only if the covariance function  $r(s, t)$  is continuous on the diagonal point  $s = t = t_0$ .*

**Proof:** If  $r(s, t)$  is continuous at  $s = t = t_0$ , then

$$\begin{aligned} E(|x(t_0 + h) - x(t_0)|^2) &= E(|x(t_0 + h)|^2) + E(|x(t_0)|^2) - 2E(x(t_0 + h) \cdot \overline{x(t_0)}) \\ &= r(t_0 + h, t_0 + h) - 2r(t_0 + h, t_0) + r(t_0, t_0) \rightarrow 0 \end{aligned}$$

as  $h \rightarrow 0$ , which shows the "if" part.

For the "only if" part, expand

$$\begin{aligned} &r(t_0 + h, t_0 + k) - r(t_0, t_0) \\ &= E((x(t_0 + h) - x(t_0)) \cdot \overline{(x(t_0 + k) - x(t_0))}) \\ &\quad + E((x(t_0 + h) - x(t_0)) \cdot \overline{x(t_0)}) \\ &\quad + E(x(t_0) \cdot \overline{(x(t_0 + k) - x(t_0))}) = e_1 + e_2 + e_3, \text{ say.} \end{aligned}$$

Here

$$\begin{aligned} e_1 &\leq \sqrt{E(|x(t_0 + h) - x(t_0)|^2) \cdot E(|x(t_0 + k) - x(t_0)|^2)} \rightarrow 0, \\ e_2 &\leq \sqrt{E(|x(t_0 + h) - x(t_0)|^2) \cdot E(|x(t_0)|^2)} \rightarrow 0, \\ e_3 &\leq \sqrt{E(|x(t_0)|^2) \cdot E(|x(t_0 + k) - x(t_0)|^2)} \rightarrow 0, \end{aligned}$$

so  $r(t_0 + h, t_0 + k) \rightarrow r(t_0, t_0)$  as  $h, k \rightarrow 0$ . □

### 2.4.2 Quadratic mean differentiability

A stochastic process  $x(t)$  is called differentiable in quadratic mean at  $t$  if there exists a random variable, naturally denoted  $x'(t)$ , such that

$$\frac{x(t+h) - x(t)}{h} \xrightarrow{q.m.} x'(t),$$

as  $h \rightarrow 0$ , i.e. if  $E((\frac{x(t+h) - x(t)}{h} - x'(t))^2) \rightarrow 0$ . If a process is differentiable both in quadratic mean and in sample function meaning, with probability one, then the two derivatives are equal with probability one. We shall now actually prove the condition for quadratic mean differentiability of a stationary process, which was stated in Section 2.1.

**Theorem 2:12** *A stationary process  $x(t)$  is quadratic mean differentiable if and only if its covariance function  $r(t)$  is twice continuously differentiable in a neighborhood of  $t = 0$ . The derivative process  $x'(t)$  has covariance function*

$$r_{x'}(t) = C(x'(s+t), x'(s)) = -r''(t).$$

**Proof:** For the "if" part we use the Loève criterion, and show that, if  $h, k \rightarrow 0$  independently of each other, then

$$E \left( \frac{x(t+h) - x(t)}{h} \cdot \frac{x(t+k) - x(t)}{k} \right) = \frac{1}{hk} (r(h-k) - r(h) - r(-k) + r(0)) \quad (2.25)$$

has a finite limit  $c$ . Define

$$\begin{aligned} f(h, k) &= r(h) - r(h-k), \\ f'_1(h, k) &= \frac{\partial}{\partial h} f(h, k) = r'(h) - r'(h-k), \\ f''_{12}(h, k) &= \frac{\partial^2}{\partial h \partial k} f(h, k) = r''(h-k). \end{aligned}$$

By applying the mean value theorem we see that there exist  $\theta_1, \theta_2 \in (0, 1)$  such that (2.25) is equal to

$$\begin{aligned} -\frac{f(h, k) - f(0, k)}{hk} &= -\frac{f'_1(\theta_1 h, k)}{k} \\ &= -\frac{f'_1(\theta_1 h, 0) + k f''_{12}(\theta_1 h, \theta_2 k)}{k} \\ &= -f''_{12}(\theta_1 h, \theta_2 k) = -r''(\theta_1 h - \theta_2 k). \end{aligned} \quad (2.26)$$

Since  $r''(t)$  by assumption is continuous, this tends to  $-r''(0)$  as  $h, k \rightarrow 0$ , which is the required limit in the Loève criterion.

To prove the "only if" we need a fact about Fourier integrals and the spectral representation  $r(t) = \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega) = \int_{-\infty}^{\infty} \cos \omega t dF(\omega)$ .

**Lemma 2.3** a)  $\lim_{t \rightarrow 0} \frac{2(r(0) - r(t))}{t^2} = \omega_2 = \int_{-\infty}^{\infty} \omega^2 dF(\omega) \leq \infty$ .

b) If  $\omega_2 < \infty$  then  $r''(0) = -\omega_2$  and  $r''(t)$  exists for all  $t$ .

c) If  $r''(0)$  exists, finite, then  $\omega_2 < \infty$  and then, by (b),  $r''(t)$  exists for all  $t$ .

**Proof of lemma:** If  $\omega_2 < \infty$ , (a) follows from

$$\frac{2(r(0) - r(t))}{t^2} = \int_{-\infty}^{\infty} \omega^2 \frac{1 - \cos \omega t}{\omega^2 t^2 / 2} dF(\omega)$$

by dominated convergence, since  $0 \leq \frac{1 - \cos \omega t}{\omega^2 t^2 / 2} \leq 1$ . If  $\omega_2 = \infty$ , the result follows from Fatou's lemma, since  $\lim_{t \rightarrow 0} \frac{1 - \cos \omega t}{\omega^2 t^2 / 2} = 1$ .

To prove (b), suppose  $\omega_2 < \infty$ . Then it is possible to differentiate twice under the integral sign in  $r(t) = \int_{-\infty}^{\infty} \cos \omega t dF(\omega)$  to obtain that  $r''(t)$  exists and

$$-r''(t) = \int_{-\infty}^{\infty} \omega^2 \cos(\omega t) dF(\omega),$$

and in particular  $-r''(0) = \omega_2$ .

For (c), suppose  $r(t)$  has a finite second derivative at the origin, and implicitly is differentiable near the origin. By the same argument as for (2.26), with  $h = k = t$ , we see that

$$\frac{2(r(0) - r(t))}{t^2} = -r''((\theta_1 - \theta_2)t) \rightarrow -r''(0)$$

as  $t \rightarrow 0$ . Then part (a) shows that  $\omega_2 < \infty$ , and the lemma is proved.  $\square$

**Proof of the "only if" part of Theorem 2:12:** If  $x(t)$  is quadratic mean differentiable,  $(x(t+h) - x(t))/h \xrightarrow{q.m.} x'(t)$  with  $E(|x'(t)|^2)$  finite, and (see further Appendix B.2),

$$E(|x'(t)|^2) = \lim_{h \rightarrow 0} E \left\{ \left| \frac{x(t+h) - x(t)}{h} \right|^2 \right\} = \lim_{h \rightarrow 0} \frac{2(r(0) - r(h))}{h^2} = \omega_2 < \infty.$$

Part (b) of the lemma shows that  $r''(t)$  exists.

To see that the covariance function of  $x'(t)$  is equal to  $-r''(t)$ , just take the limit of

$$\begin{aligned} E \left( \frac{x(s+t+h) - x(s+t)}{h} \cdot \frac{x(s+k) - x(s)}{k} \right) \\ = \frac{1}{hk} (r(t+h-k) - r(t+h) - r(t-k) + r(t)) \\ = -f''_{12}(t + \theta_1 h, \theta_2 k) = -r''(t + \theta_1 h - \theta_2 k) \rightarrow -r''(t), \end{aligned}$$

for some  $\theta_1, \theta_2$  as  $h, k \rightarrow 0$ .  $\square$

### 2.4.3 Higher order derivatives and their correlations

To obtain higher order derivatives one just has to examine the covariance function  $r_{x'}(t) = -r''_x(t)$  for the conditions in the existence theorem, etc.

The derivative of a (mean square) differentiable stationary process is also stationary and has covariance function  $r_{x'}(t) = -r''_x(t)$ , from Theorem 2:12. One can easily derive the cross-covariance relations between the derivative and the process (strictly, we need (B.7) from Appendix B).

**Theorem 2:13** *The cross-covariance between derivatives  $x^{(j)}(s)$  and  $x^{(k)}(t)$  of a stationary process  $\{x(t), t \in \mathbb{R}\}$  is*

$$Cov(x^{(j)}(s), x^{(k)}(t)) = (-1)^{(k)} r_x^{(j+k)}(s-t). \quad (2.27)$$

*In particular*

$$Cov(x(s), x'(t)) = -r'_x(s-t). \quad (2.28)$$

In Chapter 3 we will need the covariances between the process and its first two derivatives. The covariance matrix for  $x(t), x'(t), x''(t)$  is

$$\begin{pmatrix} \omega_0 & 0 & -\omega_2 \\ 0 & \omega_2 & 0 \\ -\omega_2 & 0 & \omega_4 \end{pmatrix},$$

where  $\omega_{2k} = (-1)^k r_x^{2k}(0) = \int \omega^{2k} dF(\omega)$  are spectral moments. Thus, the slope at a specified point is uncorrelated both with the process value at that point and with the curvature, while process value and curvature have negative correlation. We have, for example,  $V(x''(t) | x(0), x'(0)) = \omega_4 - \omega_2^2/\omega_0$ .

## 2.5 Summary of smoothness conditions

The following table summarizes some crude and simple sufficient conditions for different types of quadratic mean and sample function smoothness.

Condition on $r(t)$ as $t \rightarrow 0$	further condition	property
$r(t) = r(0) - o(1)$		q.m. continuous
$r(t) = r(0) - C t ^\alpha + o( t ^\alpha)$	$1 < \alpha \leq 2$	a.s. continuous
$r(t) = r(0) - C t ^\alpha + o( t ^\alpha)$	$0 < \alpha \leq 2$ Gaussian	a.s. continuous
$r(t) = r(0) - \omega_2 t^2/2 + o(t^2)$		q.m. differentiable
$-r''(t) = \omega_2 - C t ^\alpha + o( t ^\alpha)$	$1 < \alpha \leq 2$	a.s. differentiable
$-r''(t) = \omega_2 - C t ^\alpha + o( t ^\alpha)$	$0 < \alpha \leq 2$ aGaussian	a.s. differentiable

## 2.6 Stochastic integration

In this section we shall define the two simplest types of stochastic integrals, of the form

$$J_1 = \int_a^b g(t)x(t) dt,$$

$$J_2 = \int_a^b g(t) dx(t),$$

where  $g(t)$  is a deterministic function and  $x(t)$  a stochastic process with mean 0. The integrals can be defined either as quadratic mean limits of approximating Riemann or Riemann-Stieltjes sums, and depending on the type of convergence we require, the process  $x(t)$  has to satisfy suitable regularity conditions.

The two types of integrals are sufficient for our needs in these notes. A third type of stochastic integrals, needed for stochastic differential equations, are those of the form

$$J_3 = \int_a^b g(t, x(t)) dx(t),$$

in which  $g$  also is random and dependent on the integrator  $x(t)$ . These will not be dealt with here.

The integrals are defined as limits in quadratic mean of the approximating sums

$$J_1 = \lim_{n \rightarrow \infty} \sum_{k=1}^n g(t_k) x(t_k) (t_k - t_{k-1}),$$

$$J_2 = \lim_{n \rightarrow \infty} \sum_{k=1}^n g(t_k) (x(t_k) - x(t_{k-1})),$$

when  $a = t_0 < t_2 < \dots < t_n = b$ , and  $\max |t_k - t_{k-1}| \rightarrow 0$  as  $n \rightarrow \infty$ , provided the limits exist, and are independent of the subdivision  $\{t_k\}$ . To simplify the writing we have suppressed the double index in the sequences of  $t_k$ -values; there is one sequence for each  $n$ .

The limits exist as quadratic mean limits if the corresponding integrals of the covariance function  $r(s, t) = C(x(s), x(t)) = E(x(s)\overline{x(t)})$  are finite, as formulated in the following theorem, in which we assume  $E(x(t)) = 0$ . Since we shall mainly use the integrals with complex functions  $g(t)$  (in fact  $g(t) = e^{i\omega t}$ ), we formulate the theorem for complex random functions.

**Theorem 2:14** *a) If  $r(s, t)$  is continuous in  $[a, b] \times [a, b]$ , and  $g(t)$  is such that the Riemann integral*

$$Q_1 = \iint_{[a,b] \times [a,b]} g(s)\overline{g(t)}r(s, t) ds dt < \infty,$$

*then  $J_1 = \int_a^b g(t)x(t) dt$  exists as a quadratic mean limit, and  $E(J_1) = 0$  and  $E(|J_1|^2) = Q_1$ .*

*b) If  $r(s, t)$  has bounded variation<sup>8</sup> in  $[a, b] \times [a, b]$  and  $g(t)$  is such that the Riemann-Stieltjes integral*

$$Q_2 = \iint_{[a,b] \times [a,b]} g(s)\overline{g(t)} d_{s,t}r(s, t) < \infty,$$

*then  $J_2 = \int_a^b g(t) dx(t)$  exists as a quadratic mean limit, and  $E(J_2) = 0$  and  $E(|J_2|^2) = Q_2$ .*

**Proof:** The simple proof uses the Loève criterion (B.8) for quadratic mean convergence: take two sequences of partitions of  $[a, b]$  with points  $s_0, s_2, \dots, s_m$

<sup>8</sup>That  $f(t)$  is of bounded variation in  $[a, b]$  means that  $\sup \sum |f(t_k) - f(t_{k-1})|$  is bounded, with the sup taken over all possible partitions.

and  $t_0, t_1, \dots, t_n$ , respectively, and consider

$$E(S_m \overline{S_n}) = \sum_{k=1}^m \sum_{j=1}^n g(s_k) \overline{g(t_j)} r(s_k, t_j) (s_k - s_{k-1})(t_j - t_{j-1}).$$

If  $Q_1$  exists, then this converges to  $Q_1$  as the limits becomes infinitely fine. This proves (a). The reader should complete the proof for (b).  $\square$

**Example 2:6** Take  $x(t) = w(t)$ , as the Wiener process. Since  $r_w(s, t) = \sigma^2 \min(s, t)$ , we see that  $\int_a^b g(t)w(t) dt$  exists for all integrable  $g(t)$ .

**Example 2:7** If  $g(t)$  is a continuously differentiable function and  $x(t) = w(t)$ , the Wiener process, then

$$\int_a^b g(t) dw(t) = g(b)w(b) - g(a)w(a) - \int_a^b g'(t)w(t) dt.$$

To prove this, consider

$$\begin{aligned} S_2 &= \sum_{k=1}^m g(t_k)(w(t_k) - w(t_{k-1})) \\ &= g(t_m)w(t_m) - g(t_1)w(t_0) - \sum_{k=2}^m (g(t_k) - g(t_{k-1}))w(t_{k-1}). \end{aligned}$$

Since  $g(t)$  is continuously differentiable, there is a  $\rho_k$  such that  $g(t_k) - g(t_{k-1}) = (t_k - t_{k-1})(g'(t_k) + \rho_k)$ , and  $\rho_k \rightarrow 0$  uniformly in  $k = 1, \dots, m$  as  $m \rightarrow \infty$ ,  $\max_k |t_k - t_{k-1}| \rightarrow 0$ . Thus

$$\begin{aligned} S_2 &= g(t_m)w(t_m) - g(t_1)w(t_0) - \sum_{k=2}^m g(t_k)'w(t_{k-1})(t_k - t_{k-1}) \\ &\quad + \sum_{k=2}^m \rho_k w(t_{k-1})(t_k - t_{k-1}) \\ &\rightarrow g(b)w(b) - g(a)w(a) - \int_a^b g'(t)w(t) dt. \end{aligned}$$

The proofs of the following two theorems are left to the reader.

**Theorem 2:15** If  $x(s)$  and  $y(t)$  are stochastic processes with cross-covariance

$$r_{x,y}(s, t) = \text{Cov}(x(s), y(t)),$$

and if the conditions of Theorem 2:14 are satisfied, then

$$E \left( \int_a^b g(s)x(s) ds \cdot \overline{\int_c^d h(t)y(t) dt} \right) = \int_a^b \int_c^d g(s) \overline{h(t)} r_{x,y}(s, t) ds dt \quad (2.29)$$

$$E \left( \int_a^b g(s) dx(s) \cdot \overline{\int_c^d h(t) dy(t)} \right) = \int_a^b \int_c^d g(s) \overline{h(t)} d_{s,t} r_{x,y}(s, t). \quad (2.30)$$

**Theorem 2:16** For the Wiener-process with  $r_{x,x}(s,t) = \min(s,t)$  one has

$$d_{s,t}r_{x,x}(s,t) = ds$$

for  $s = t$  and 0 otherwise, which gives, for  $a < c < b < d$ ,

$$E \left( \int_a^b g(s) dx(s) \cdot \int_c^d h(t) dx(t) \right) = \int_c^b g(t)h(t) dt.$$

**Remark 2:2** A natural question is: are quadratic mean integrals and ordinary integrals equal? If a stochastic process has a continuous covariance function, and continuous sample paths, with probability one, and if  $g(t)$  is, for example, continuous, then  $\int_a^b g(t)x(t) dt$  exists both as a regular Riemann integral and as a quadratic mean integral. Both integrals are random variables and they are limits of the same approximating Riemann sum, the only difference being the mode of convergence – with probability one, and in quadratic mean, respectively. But then the limits are equivalent, i.e. equal with probability one.

## 2.7 An ergodic result

An ergodic theorem deals with convergence properties of *time or sample function averages* to *ensemble averages* i.e. to *statistical expectation*:

$$\frac{1}{T} \int_0^T f(x(t)) dt \rightarrow E(f(x(0))) \quad \text{as } T \rightarrow \infty,$$

for a function of a stationary stochastic process  $x(t)$ . Such theorems will be the theme of the entire Chapter 5, but we show already here a simple such result based only on covariance properties. The process  $x(t)$  need not even be stationary, but we assume  $E(x(t)) = 0$  and that the covariance function  $r(s,t) = \text{Cov}(x(s), x(t))$  exists.

**Theorem 2:17** a) If  $r(s,t)$  is continuous for all  $s,t$  and

$$\frac{1}{T^2} \int_0^T \int_0^T r(s,t) ds dt \rightarrow 0, \quad \text{as } T \rightarrow \infty, \quad (2.31)$$

then

$$\frac{1}{T} \int_0^T x(t) dt \xrightarrow{q.m.} 0.$$

b) If there exist constants  $K, \alpha, \beta$ , such that  $0 \leq 2\alpha < \beta < 1$ , and

$$|r(s,t)| \leq K \frac{s^\alpha + t^\alpha}{1 + |s-t|^\beta}, \quad \text{for } s, t \geq 0 \quad (2.32)$$

then

$$\frac{1}{T} \int_0^T x(t) dt \xrightarrow{a.s.} 0.$$

**Proof:** a) This is immediate from Theorem 2:15, since

$$\sigma_T^2 = E \left( \left( \frac{1}{T} \int_0^T x(t) dt \right)^2 \right) = \frac{1}{T^2} \int_0^T \int_0^T r(s, t) ds dt \rightarrow 0,$$

by assumption.

b) Before we prove the almost sure convergence, note that the condition  $2\alpha < \beta < 1$  puts a limit on how fast  $E(x(t)^2) = r(t, t)$  is allowed to increase as  $t \rightarrow \infty$ , and it limits the amount of dependence between  $x(s)$  and  $x(t)$  for large  $|s - t|$ . If the dependence is too strong, it may very well happen that  $\frac{1}{T} \int_0^T x(t) dt$  converges, but not to 0 but to a (random) constant different from 0.

We show here only part of the theorem, and refer the reader to [9, p. 95] for a completion. What we show is that there exists a subsequence of times,  $T_n \rightarrow \infty$ , such that  $\frac{1}{T_n} \int_0^{T_n} x(t) dt \xrightarrow{a.s.} 0$ .

First estimate  $\sigma_T^2$  from the proof of (a):

$$\begin{aligned} \sigma_T^2 &= E \left( \left( \frac{1}{T} \int_0^T x(t) dt \right)^2 \right) = \frac{1}{T^2} \int_0^T \int_0^T r(s, t) ds dt \\ &\leq \frac{K}{T^2} \int_0^T \int_0^T \frac{s^\alpha + t^\alpha}{1 + |s - t|^\beta} ds dt \\ &= \frac{K}{T^{\beta-\alpha}} \cdot \frac{1}{T^{2-\beta}} \int_0^T \int_0^T \frac{(s/T)^\alpha + (t/T)^\alpha}{1 + |s - t|^\beta} ds dt \\ &\leq \frac{K}{T^{\beta-\alpha}} \cdot \frac{2}{T^{1-\alpha}} \int_0^T \frac{1}{1 + u^\beta} du. \end{aligned}$$

Here  $(2/T^{1-\alpha}) \int_0^T (1 + u^\beta)^{-1} du$  tends to a constant as  $T \rightarrow \infty$ , which implies that  $\sigma_T^2 \leq K'/T^{\beta-\alpha}$  for some constant  $K'$ .

Take the constant  $\gamma$  such that  $\gamma(\beta - \alpha) > 1$ , which is possible from the properties of  $\alpha$  and  $\beta$ , and put  $T_n = n^\gamma$ , with

$$\sum_{n=1}^{\infty} \sigma_{T_n}^2 \leq \sum_{n=1}^{\infty} \frac{K'}{n^{\gamma(\beta-\alpha)}} < \infty. \quad (2.33)$$

That the sum (2.33) is finite implies, by the Borel-Cantelli lemma and the Chebysjev inequality, see (B.3) in Appendix B, that  $T_n^{-1} \int_0^{T_n} x(t) dt \xrightarrow{a.s.} 0$ , and so we have showed the convergence for a special sequence of times.

To complete the proof, we have to show that

$$\sup_{T_n \leq T \leq T_{n+1}} \left| \frac{1}{T} \int_0^T x(t) dt - \frac{1}{T_n} \int_0^{T_n} x(t) dt \right| \xrightarrow{a.s.} 0,$$

as  $n \rightarrow \infty$ ; see [9, p. 95]. □

For stationary processes, the theorem yields the following ergodic theorem about the observed average.

**Theorem 2:18** a) If  $x(t)$  is stationary and  $\frac{1}{T} \int_0^T r(t) dt \rightarrow 0$  as  $T \rightarrow \infty$  then  $\frac{1}{T} \int_0^T x(t) dt \xrightarrow{q.m.} 0$ .

b) If moreover there is a constant  $K > 0$  and a  $\beta > 0$ , such that  $|r(t)| \leq \frac{K}{|t|^\beta}$  as  $t \rightarrow \infty$ , then  $\frac{1}{T} \int_0^T x(t) dt \xrightarrow{a.s.} 0$ .

## Exercises

- 2:1. Prove the following useful inequality valid for any non-negative, integer-valued random variable  $N$ ,

$$E(N) - \frac{1}{2}E(N(N-1)) \leq P(N > 0) \leq E(N).$$

Generalize it to the following inequalities where

$$\alpha_i = E(N(N-1) \cdots (N-i+1))$$

is the  $i^{\text{th}}$  factorial moment:

$$\frac{1}{k!} \sum_{i=0}^{2n-1} (-1)^i \frac{1}{i!} \alpha_{(k+i)} \leq P(N = k) \leq \frac{1}{k!} \sum_{i=0}^{2n} (-1)^i \frac{1}{i!} \alpha_{(k+i)}.$$

- 2:2. Let  $\{x(t), t \in \mathbb{R}\}$  and  $\{y(t), t \in \mathbb{R}\}$  be equivalent processes which both have, with probability one, continuous sample paths. Prove that

$$P(x(t) = y(t), \text{ for all } t \in \mathbb{R}) = 1.$$

- 2:3. Find the values on the constants  $a$  and  $b$  that make a Gaussian process twice continuously differentiable if its covariance function is

$$r(t) = e^{-|t|}(1 + a|t| + bt^2).$$

- 2:4. Complete the proof of Theorem 2:3 and show that, in the notation of the proof,

$$\sum_n g(2^{-n}) < \infty, \quad \text{and} \quad \sum_n 2^n q(2^{-n}) < \infty.$$

- 2:5. Show that the sample paths of the Wiener process have infinite variation, a.s., by showing the stronger statement that if

$$Y_n = \sum_{k=0}^{2^n-1} \left| W\left(\frac{k+1}{2^n}\right) - W\left(\frac{k}{2^n}\right) \right|$$

then  $\sum_{n=1}^{\infty} P(Y_n < n) < \infty$ .

- 2:6. Show that a non-stationary process is continuous in quadratic mean at  $t = t_0$  only if its mean value function  $m(t) = E(x(t))$  is continuous at  $t_0$  and its covariance function  $r(s, t) = \text{Cov}(x(s), x(t))$  is continuous at  $s = t = t_0$ .

- 2:7. Convince yourself of the “trivial” fact that a sequence of normal variables  $\{x_n, n \in \mathbb{Z}\}$ , such that  $E(x_n)$  and  $V(x_n)$  have finite limits, then the sequence converges in distribution to a normal variable.

2:8. Give an example of a stationary process that violates the sufficient conditions in Theorem 2:10 and for which the sample functions can be tangents of the level  $u = 1$ .

2:9. Assume that sufficient conditions on  $r(s, t) = E(x(s)\overline{x(t)})$  are satisfied so that the integral

$$\int_0^T g(t)x(t) dt$$

exists for all  $T$ , both as a quadratic mean integral and as a sample function integral. Show that, if

$$\int_0^\infty |g(t)|\sqrt{r(t, t)} dt < \infty,$$

then the generalized integral  $\int_0^\infty g(t)x(t) dt$  exists as a limit as  $T \rightarrow \infty$ , both in quadratic mean and with probability one.

2:10. Let  $(x_n, y_n)$  have a bivariate Gaussian distribution with mean 0, variance 1, and correlation coefficient  $\rho_n$ .

a) Show that  $P(x_n < 0 < y_n) = \frac{1}{2\pi} \arccos \rho_n$ .

b) Calculate the conditional density functions for

$$(x_n + y_n) \mid x_n < 0 < y_n, \quad \text{and} \quad (y_n - x_n) \mid x_n < 0 < y_n.$$

c) Let  $z_n$  and  $u_n$  be distributed with the in (b) derived density functions and assume that  $\rho_n \rightarrow 1$  as  $n \rightarrow \infty$ . Take  $c_n = 1/\sqrt{2(1 - \rho_n)}$ , and show that the density functions for  $c_n z_n$  and  $c_n u_n$  converge to density functions  $f_1$  and  $f_2$ , respectively.

Hint:  $f_2(u) = u \exp(-u^2/2)$ ,  $u > 0$  is the Rayleigh density.

2:11. Let  $\{x(t), t \in \mathbb{R}\}$  be a stationary Gaussian process with mean 0, and with a covariance function that satisfies

$$-r''(t) = -r''(0) + o(|t|^a), \quad t \rightarrow 0,$$

for some  $a > 0$ . Define  $x_n = x(0)$ ,  $y_n = x(1/n)$ ,  $\rho_n = r(1/n)$  and use the previous exercise to derive the asymptotic distribution of

$$\frac{x(1/n) - x(0)}{1/n} \mid x(0) < 0 < x(1/n)$$

as  $n \rightarrow \infty$ . What conclusion do you draw about the derivative at a point with an upcrossing of the zero level? (Answer: it has a Rayleigh distribution, not a half normal distribution.)

2:12. Find an example of two dependent normal random variables  $U$  and  $V$  such that  $C(U, V) = 0$ ; obviously you cannot let  $(U, V)$  have a bivariate normal distribution.

2:13. Prove that Theorem 2:18 follows from Theorem 2:17.

# Chapter 3

## Crossings

### 3.1 Level crossings and Rice's formula

In applications one encounters level crossings, and in particular the distribution of the number of solutions to  $x(t) = u$  for a specified level  $u$ . This is a difficult question in general and very few explicit results can be derived. There exists however, a very famous formula, found by Marc Kac and Steve O. Rice [27], for the *expected number of upcrossings* of a level  $u$ . We shall here state and prove that formula. In Section 3.2 we shall use different forms of Rice's formula to investigate the conditional behavior of a stationary process when it is observed in the neighborhood of a crossing of a predetermined level. Quantities like the height and time extension of the excursion above the level will be analyzed by means of an explicit stochastic model, the *Slepian model*, used first by David Slepian 1962, [31], for a Gaussian process after zero crossings.

#### 3.1.1 Level crossings

In practice, level crossing counting is often used as a means to describe the variability and extremal behavior of a continuous stochastic process. For example, the maximum of the process in an interval is equal to the lowest level above which there exists no genuine level crossing, provided, of course, that the process starts below that level. Since it is often easier to find the statistical properties of the number of level crossings than to find the maximum distribution, crossing methods are of practical importance.

For sample functions of a continuous process  $\{x(t), t \in \mathbb{R}\}$  we say that  $x(t)$  has an *upcrossing* of the level  $u$  at  $t_0$  if, for some  $\epsilon > 0$ ,  $x(t) \leq u$  for all  $t \in (t_0 - \epsilon, t_0]$  and  $x(t) \geq u$  for all  $t \in [t_0, t_0 + \epsilon)$ . For any interval  $I = [a, b]$ , write  $N_I^+(x, u)$  for the number of upcrossings by  $x(t)$  in  $I$ ,

$$N_I^+ = N_I^+(x, u) = \text{the number of } u\text{-upcrossings by } x(t), t \in I.$$

For continuous processes which have only a finite number of  $u$ -values, there must be intervals to the left and to the right of any upcrossing point such that

$x(t)$  is *strictly less than*  $u$  immediately to the left and *strictly greater than*  $u$  immediately to the right of the upcrossing point. Also define

$$N_I = N_I(x, u) = \text{the number of } t \in I \text{ such that } x(t) = u.$$

By the *intensity of upcrossings* we mean any function  $\mu_t^+(u)$  such that

$$\int_{t \in I} \mu_t^+(u) dt = E(N_I^+(x, u)).$$

Similarly, we define the *intensity of crossings*, as  $\mu_t(u)$  if

$$\int_{t \in I} \mu_t(u) dt = E(N_I(x, u)).$$

For a stationary process,  $\mu_t^+(u) = \mu^+(u)$  and  $\mu_t(u) = \mu(u)$  are independent of  $t$ . In general, the intensity is the mean number of events per time unit, calculated at time  $t$ .

In reliability applications of stochastic processes one may want to calculate the distribution of the maximum of a continuous process  $x(t)$  in an interval  $I = [0, T]$ . The following approximation is then often useful, and also sufficiently accurate for short intervals,

$$\begin{aligned} P(\max_{0 \leq t \leq T} x(t) > u) &= P(\{x(0) \leq u\} \cap \{N_I^+(x, u) \geq 1\}) + P(x(0) > u) \\ &\leq P(N_I^+(x, u) \geq 1) + P(x(0) > u) \\ &\leq E(N_I^+(x, u)) + P(x(0) > u) = T \cdot \mu^+(u) + P(x(0) > u). \end{aligned}$$

The upcrossing intensity  $\mu^+(u)$  was found by Rice for Gaussian processes, results which were later given strict proofs through counting methods developed by Kac. The classical reference is [27]. We give first a general formulation and then specialize to Gaussian processes; see [22, 30].

### 3.1.2 Rice's formula for absolutely continuous processes

We present the simplest version and proof of Rice's formula, valid for processes  $\{x(t), t \in \mathbb{R}\}$  with absolutely continuous sample paths<sup>1</sup> and absolutely continuous distribution with density  $f_{x(t)}(u) = f_{x(0)}(u)$ , independent of  $t$ . For such a process, the derivative  $x'(t)$  exists almost everywhere, and the conditional expectations

$$E(x'(0)^+ | x(0) = u) \quad \text{and} \quad E(|x'(0)| | x(0) = u)$$

exist, (with  $x^+ = \max(0, x)$ ).

<sup>1</sup>A function  $x(t), t \in [a, b]$  is absolutely continuous if it is equal to the integral  $x(t) = \int_a^t y(s) ds$  of an integrable function  $y(s)$ . This is equivalent to the requirement that for every  $\epsilon > 0$  there is a  $\delta > 0$  such that for every collection  $(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)$  of disjoint intervals in  $[a, b]$  with  $\sum_1^n (b_k - a_k) < \delta$  one has  $\sum_1^n |x(b_k) - x(a_k)| < \epsilon$ . An absolutely continuous function is always continuous and its derivative exists almost everywhere,  $x'(t) = y(t)$ , almost everywhere.

**Theorem 3:1** (*Rice's formula*) For any stationary process  $\{x(t), t \in \mathbb{R}\}$  with density  $f_{x(0)}(u)$ , the crossings and up-crossings intensities are given by

$$\begin{aligned}\mu(u) &= E(N_{[0,1]}(x, u)) = \int_{-\infty}^{\infty} |z| f_{x(0), x'(0)}(u, z) dz \\ &= f_{x(0)}(u) E(|x'(0)| \mid x(0) = u),\end{aligned}\tag{3.1}$$

$$\begin{aligned}\mu^+(u) &= E(N_{[0,1]}^+(x, u)) = \int_0^{\infty} z f_{x(0), x'(0)}(u, z) dz \\ &= f_{x(0)}(u) E(x'(0)^+ \mid x(0) = u).\end{aligned}\tag{3.2}$$

These expressions hold for almost any  $u$ , whenever the involved densities exist.

Before we state the short proof we shall review some facts about functions of bounded variation, proved by Banach. To formulate the proof, write for any continuous function  $f(t)$ ,  $t \in [0, 1]$ , and interval  $I = [a, b] \subset [0, 1]$ ,

$$N_I(f, u) = \text{the number of } t \in I \text{ such that } f(t) = u.$$

Further, define the *total variation* of  $f(t), t \in I$  as  $\sup \sum |f(t_{k+1}) - f(t_k)|$ , where the supremum is taken over all subdivisions  $a \leq t_0 < t_1 < \dots < t_n \leq b$ .

**Lemma 3.1** (*Banach*) For any continuous function  $f(t), t \in I$ , the total variation is equal to

$$\int_{-\infty}^{\infty} N_I(f, u) du.$$

Further, if  $f(t)$  is absolutely continuous with derivative  $f'(t)$  then

$$\int_{-\infty}^{\infty} N_I(f, u) du = \int_I |f'(t)| dt.$$

Similarly, if  $A \subseteq \mathbb{R}$  is any Borel measurable set, and  $\mathbf{1}_A$  is its indicator function, then

$$\int_{-\infty}^{\infty} \mathbf{1}_A(u) N_I(f, u) du = \int_I \mathbf{1}_A(f(t)) |f'(t)| dt\tag{3.3}$$

**Proof of Rice's formula** We prove (3.1) by applying Banach's theorem on the stationary process  $\{x(t), t \in \mathbb{R}\}$  with absolutely continuous, and hence a.s. differentiable, sample paths. If  $x(t)$  has a.s. absolutely continuous sample functions, then (3.3) holds for almost every realization, i.e.

$$\int_{-\infty}^{\infty} \mathbf{1}_A(u) N_I(x, u) du = \int_I \mathbf{1}_A(x(t)) |x'(t)| dt$$

Taking expectations and using Fubini's theorem to change the order of integration and expectation, we get

$$\begin{aligned} |I| \int_{u \in A} \mu(u) du &= \int_{-\infty}^{\infty} \mathbf{1}_A(u) E(N_I(x, u)) du \\ &= E \left( \int_I \mathbf{1}_A(x(t)) |x'(t)| dt \right) = |I| E(\mathbf{1}_A(x(0)) |x'(0)|) \\ &= |I| \int_{u \in A} f_{x(0)}(u) E(|x'(0)| | x(0) = u) du; \end{aligned}$$

here we also used that  $\{x(t), t \in \mathbb{R}\}$  is stationary.

Since  $A$  is an arbitrary measurable set, we get the desired result,

$$\mu(u) = f_{x(0)}(u) E(|x'(0)| | x(0) = u)$$

for almost all  $u$ . The proof of (3.2) is similar.  $\square$

### 3.1.3 Alternative proof of Rice's formula

The elegant and general proof of Rice's formula just given does not give an intuitive argument for the presence of the factor  $z$  in the integral (3.2). The following more pedestrian proof is closer to an explanation.

**Theorem 3:2** *For a stationary process  $\{x(t), t \in \mathbb{R}\}$  with almost surely continuous sample paths, suppose  $x(0)$  and  $\zeta_n = 2^n(x(1/2^n) - x(0))$  have a joint density  $g_n(u, z)$  which is continuous in  $u$  for all  $z$  and all sufficiently large  $n$ . Also suppose  $g_n(u, z) \rightarrow p(u, z)$  uniformly in  $u$  for fixed  $z$  as  $n \rightarrow \infty$  and that  $g_n(u, z) \leq h(z)$  with  $\int_0^\infty zh(z) dz < \infty$ .<sup>2</sup> Then*

$$\mu^+(u) = E(N_{[0,1]}(x, u)) = \int_0^\infty zp(u, z) dz. \quad (3.4)$$

**Proof:** We first devise a counting technique for the upcrossings by dividing the interval  $[0, 1]$  into dyadic subintervals,  $[(k-1)/2^n, k/2^n]$ ,  $k = 1, \dots, 2^n$ , and checking the values at the endpoints. Let  $N_n$  denote the number of points  $k/2^n$  such that  $x((k-1)/2^n) < u < x(k/2^n)$ . Since  $x(t)$  has continuous sample paths (a.s.), there is at least one  $u$ -upcrossing in every interval such that  $x((k-1)/2^n) < u < x(k/2^n)$ , and hence

$$N_n \leq N_{[0,1]}(x, u).$$

Furthermore, since  $x(t)$  has a continuous distribution, we may assume that  $x(k/2^n) \neq u$  for all  $n$  and  $k = 1, \dots, 2^n$ . When  $n$  increases to  $n+1$  the

<sup>2</sup>It is of course tempting to think of  $p(u, z)$  as the joint density of  $x(0), x'(0)$ , but no argument for this is involved in the proof.

number of subintervals doubles, each interval contributing one upcrossing to  $N_n$  will contribute at least one upcrossing to  $N_{n+1}$  in at least one of the two new subintervals. Hence  $N_n$  is increasing and it is easy to see that, regardless of if  $N_{[0,1]}(x, u) = \infty$  or  $N_{[0,1]}(x, u) < \infty$ ,  $N_n \uparrow N_{[0,1]}(x, u)$  as  $n \rightarrow \infty$ . Monotone convergence implies that  $\lim_{n \rightarrow \infty} E(N_n) = E(N_{[0,1]}(x, u))$ .

Now, define

$$J_n(u) = 2^n P(x(0) < u < x(1/2^n)),$$

so, by stationarity,

$$E(N_{[0,1]}(x, u)) = \lim_{n \rightarrow \infty} E(N_n) = \lim_{n \rightarrow \infty} J_n(u).$$

By writing the event  $\{x(0) < u < x(1/2^n)\}$  as

$$\{x(0) < u < x(0) + \zeta_n/2^n\} = \{x(0) < u\} \cap \{\zeta_n > 2^n(u - x(0))\},$$

we have

$$\begin{aligned} J_n(u) &= 2^n P(x(0) < u, \zeta_n > 2^n(u - x(0))) \\ &= 2^n \int_{x=-\infty}^u \int_{y=2^n(u-x)}^{\infty} g_n(x, y) dy dx. \end{aligned}$$

By a change of variables,  $x = u - zv/2^n$ ,  $y = z$ , ( $v = 2^n(u - x)/y$ ), this is equal to

$$\int_{z=0}^{\infty} z \int_{v=0}^1 g_n(u - zv/2^n, z) dv dz,$$

where  $g_n(u - zv/2^n, z)$  tends pointwise to  $p(u, z)$  as  $n \rightarrow \infty$ , by the assumptions of the theorem. Since  $g_n$  is dominated by the integrable  $h(z)$  it follows that the double integral tends to  $\int_0^{\infty} zp(u, z) dz$  as  $n \rightarrow \infty$ .  $\square$

**Remark 3:1** *The proof of Rice's formula as illustrated in Figure 3.1 shows the relation to the Kac and Slepian horizontal window conditioning: one counts the number of times the process passes through a small horizontal window; we will dwell upon this concept in Section 3.2.1, page 65.*

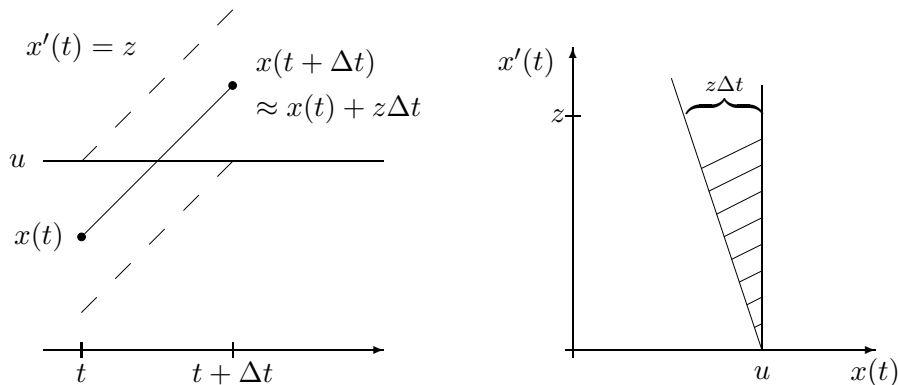
**Remark 3:2** *Rice's formula can be extended to non-stationary processes, in which case the crossings intensity is time dependent. Then the density function  $f_{x(t), x'(t)}(u, z)$  in the integral depends on  $t$ :*

$$E(N_{[a,b]}(x, u)) = \int_a^b \int_{-\infty}^{\infty} |z| f_{x(t), x'(t)}(u, z) dz dt.$$

The integral

$$\mu_t(u) = \int_{-\infty}^{\infty} |z| f_{x(t), x'(t)}(u, z) dz$$

is the local crossings intensity at time  $t$ .



**Figure 3.1:** Upcrossing occurs between  $t$  and  $t + \Delta t$  when  $x(t)$  is between  $u - x'(t)\Delta t$  and  $u$ . The probability is the integral of the joint density  $f_{x(t),x'(t)}(u, z)$  over the dashed area.

### 3.1.4 Rice's formula for differentiable Gaussian processes

For a Gaussian stationary process, Rice's formula becomes particularly simple. We know from elementary courses that for a stationary differentiable Gaussian process  $\{x(t), t \in \mathbb{R}\}$ , the process value,  $x(t)$ , and the derivative,  $x'(t)$  at the same time point, are independent<sup>3</sup> and Gaussian, with mean  $E(x(t)) = m$ ,  $E(x'(t)) = 0$ , respectively, and variances given by the spectral moments,  $V(x(t)) = r(0) = \omega_0$ ,  $V(x'(t)) = -r''(0) = \omega_2$ . Then

$$f_{x(0),x'(0)}(u, z) = \frac{1}{2\pi\sqrt{\omega_0\omega_2}} e^{-(u-m)^2/2\omega_0} e^{-z^2/2\omega_2}. \quad (3.5)$$

Simple integration of (3.1) and (3.2) gives that for Gaussian stationary processes,

$$\begin{aligned} \mu(u) &= E(N_{[0,1]}(x, u)) = \frac{1}{\pi} \sqrt{\frac{\omega_2}{\omega_0}} e^{-(u-m)^2/2\omega_0}, \\ \mu^+(u) &= E(N_{[0,1]}^+(x, u)) = \frac{1}{2\pi} \sqrt{\frac{\omega_2}{\omega_0}} e^{-(u-m)^2/2\omega_0}, \end{aligned}$$

which are the original forms of Rice's formula. These formulas hold regardless of whether  $\omega_2$  is finite or not, so  $\omega_2 = \infty$  if and only if the expected number of crossings in any interval is infinite. This does not mean, however, that there necessarily are infinitely many crossings, but if there is crossings, then there may be infinitely many in the neighborhood.

<sup>3</sup>If you have not seen this before, prove it by showing that  $x(t)$  and  $x'(t) = \lim_{h \rightarrow 0} (x(t+h) - x(t))/h$  are uncorrelated.

**Remark 3:3** *The expected number of mean-level upcrossings per time unit in a stationary Gaussian process is*

$$\mu^+(m) = \frac{1}{2\pi} \sqrt{\omega_2/\omega_0} = \frac{1}{2\pi} \sqrt{\frac{\int \omega^2 f(\omega) d\omega}{\int f(\omega) d\omega}},$$

and it is called the (root) mean square frequency of the process. The inverse is equal to the long run average time distance between successive mean level upcrossings,  $1/\mu^+(m) = 2\pi \sqrt{\omega_0/\omega_2}$ , also called the mean period.

A local extreme, minimum or maximum, for a differentiable stochastic process  $\{x(t), t \in \mathbb{R}\}$  corresponds to, respectively, an upcrossing and a downcrossing of the zero level by the process derivative  $\{x'(t), t \in \mathbb{R}\}$ . Rice's formula applied to  $x'(t)$  therefore gives the expected number of local extremes. For a Gaussian process the formulas involve the fourth spectral moment  $\omega_4 = V(x''(t)) = \int \omega^4 f(\omega) d\omega$ . The general and Gaussian expressions are, respectively,

$$\begin{aligned} \mu_{min} &= \int_0^\infty z f_{x',x''}(0, z) dz = \frac{1}{2\pi} \sqrt{\frac{\omega_4}{\omega_2}}, \\ \mu_{max} &= \int_{-\infty}^0 |z| f_{x',x''}(0, z) dz = \frac{1}{2\pi} \sqrt{\frac{\omega_4}{\omega_2}}. \end{aligned}$$

If we combine this with Remark 3:3 we get the average number of local maxima per mean level upcrossings,

$$1/\alpha = \frac{\frac{1}{2\pi} \sqrt{\omega_4/\omega_2}}{\frac{1}{2\pi} \sqrt{\omega_2/\omega_0}} = \sqrt{\frac{\omega_0 \omega_4}{\omega_2^2}}.$$

The parameter  $\alpha$  is always bounded by  $0 < \alpha < 1$ , and it is used as an irregularity measure: an  $\alpha$  near 1 indicates a very regular process with approximately one local maximum and minimum between mean level upcrossings. If  $\alpha$  is near zero one can expect many local extremes between the upcrossings.

Seen in relation to the spectrum, the parameter  $\alpha$  can be seen as a measure of *spectral width*. A spectrum with  $\alpha$  near 1 is *narrow banded*, i.e. the spectral density is concentrated to a small frequency band around a dominating center frequency. A narrow banded process has very regular sample functions, with slowly varying random amplitude; see Section 4.4.5 that deals with the envelope process.

## 3.2 Prediction from a random crossing time

In the prediction theory briefly presented in Section 1.5.2 our concern was to predict, as accurately as possible, the unknown future value  $x(t_0 + \tau)$  of a process  $\{x(t), t \in \mathbb{R}\}$ , from what we know by observations available at time

$t_0$ . An implicit assumption has been that there is no stochastic dependence between the choice of  $t_0$  and  $x(t_0 + \tau)$ .

For example, given that we have a complete record of all old values, the best predictor, in the sense of smallest mean square error, is the conditional expectation  $\hat{x}(t_0 + \tau) = E(x(t_0 + \tau) \mid x(s), s \leq t_0)$ . If the process is Gaussian the predictor is linear in the observed value; for example, when we know only the value of  $x(t_0)$ , the optimal solution that has smallest mean square error, taken over all outcomes of  $x(t_0)$ , is

$$\hat{x}(t_0 + \tau) = E(x(t_0 + \tau) \mid x(t_0)) = \frac{r(\tau)}{r(0)} x(t_0). \quad (3.6)$$

### 3.2.1 Prediction from upcrossings

There are situations where the time point from which we want to predict the future process is not a deterministic time point but a random time, determined by the process itself. An example is an *alert predictor* of the water level in a flood protection system: when the water level reaches a certain warning level special actions are taken, such as special surveillance, more detailed prediction, etc. Another example is a surveillance system in a medical intense care unit.

We assume throughout this section that the process  $\{x(t), t \in \mathbb{R}\}$  has continuous sample functions. Prediction from an upcrossing time point  $t_0$  with specified level  $u$ , shall be based on the conditional distributions, given that an upcrossing of the level  $u$  has occurred at  $t_0$ . So, we need to find those conditional distributions,<sup>4</sup> in particular

$$\hat{x}^u(t_0 + \tau) = E(x(t_0 + \tau) \mid x(t_0) = u, \text{ upcrossing}). \quad (3.7)$$

#### 3.2.1.1 A word about conditioning

A conditional expectation was defined in an elementary way in Section 1.2.4 as  $\varphi(v) = E(x \mid y = v) = \int_u u f_{x,y}(u, v) / f_y(v) du$ . The main property of the conditional expectation is relation (1.5),  $E(x) = \int_y \varphi(y) f(y) dy$  and its more refined version,

$$E(x \mid y \in A) = \frac{\int_{y \in A} \varphi(y) f(y) dy}{\int_{y \in A} f(y) dy},$$

for every Borel set  $A$ . For example, when the density  $f(y)$  is continuous, with  $A = [u - \epsilon, u]$ ,

$$\lim_{\epsilon \rightarrow 0} E(x \mid u - \epsilon \leq y \leq u) = E(x \mid y = u) = \varphi(u). \quad (3.8)$$

The meaning of this is that  $E(x \mid y = u)$  is the expected value of  $x$  for those outcomes where the value of  $y$  is close to  $u$ ; for more details, see [5, Ch. 4].

---

<sup>4</sup>In Markov process theory one has introduced the *strong Markov property* to handle conditioning from a time point that depends on the process.

If  $\{x(t), t \in \mathbb{R}\}$  is a stationary process, take  $y = x(t_0)$  and  $x = x(t_0 + \tau)$ . Then

$$\begin{aligned} P(x(t_0 + \tau) \leq v \mid x(t_0) = u) \\ = \lim_{\epsilon \rightarrow 0} P(X(t_0 + \tau) \leq v \mid u - \epsilon \leq x(t_0) \leq u), \end{aligned} \quad (3.9)$$

$$\begin{aligned} E(x(t_0 + \tau) \mid x(t_0) = u) \\ = \lim_{\epsilon \rightarrow 0} E(X(t_0 + \tau) \mid u - \epsilon \leq x(t_0) \leq u), \end{aligned} \quad (3.10)$$

calculated at time  $t_0$ , and (3.10) gives the best predictor of the future value  $x(t_0 + \tau)$  in the sense that it minimizes the squared error taken as an average over all the possible outcomes of  $x(t_0)$ . By "average" we then mean expected value as well as an empirical average over many realizations observed at the fixed predetermined time  $t_0$ , chosen independently of the process.

### 3.2.1.2 Empirical limits and horizontal window conditioning

We now consider prediction from the times of upcrossings of a fixed level  $u$ . This differs from the previous type of conditioning in that the last observed value of the process is known, and the time points are variable. The interpretation of "average future value" is then not clear at this moment and has to be made precise. Obviously, what we should aim at is a prediction method that works well on the average, in the long run, for all the  $u$ -level upcrossings we observe in the process. Call these upcrossings time points  $t_k > 0$ . To this end, define the following distribution.

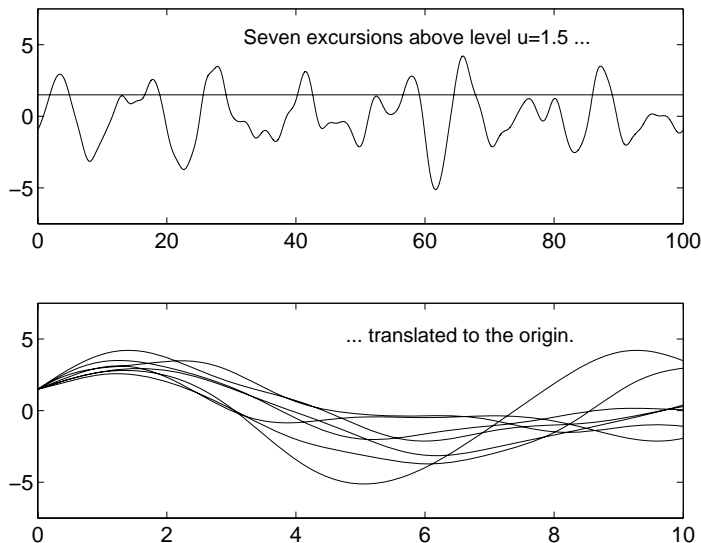
**Definition 3:1** For a stationary process  $\{x(t), t \in \mathbb{R}\}$ , the (long run, ergodic) conditional distribution of  $x(t_0 + \cdot)$  after  $u$ -upcrossing at  $t_0$  is defined as

$$\begin{aligned} P^u(A) &= "P(x(t_0 + \cdot) \in A \mid x(t_0) = u, \text{ upcrossing})" \\ &= \lim_{T \rightarrow \infty} \frac{\#\{t_k; 0 < t_k < T \text{ and } x(t_k + \cdot) \in A\}}{\#\{t_k; 0 < t_k < T\}}. \end{aligned} \quad (3.11)$$

Thus, in  $P^u(A)$  is counted all those  $u$ -upcrossings  $t_k$  for which the process, taken with  $t_k$  as new origin, satisfies the condition given by  $A$ .

The definition makes sense only if the limit exists, but, as we shall prove in Chapter 5, the limit exists for every stationary process  $\{x(t), t \in \mathbb{R}\}$ , but it may be random. If the process is *ergodic* the limit is non-random and it defines a proper distribution on  $\mathbb{C}$  for a (non-stationary) stochastic process.

The empirical long run distribution is related to Kac and Slepian's *horizontal window conditioning*, [19]. For a stationary process  $\{x(t), t \in \mathbb{R}\}$ , the (horizontal window) conditional distribution of  $x(t_0 + \cdot)$  after  $u$ -upcrossing at



**Figure 3.2:** Excursions above  $u = 1.5$  contribute to the distribution  $P^{1.5}(\cdot)$ .

$t_0$  is

$$\begin{aligned} P^{hw}(A) &= "P(x(t_0 + \cdot) \in A \mid x(t_0) = u, \text{ upcrossing in h.w.sense})" \quad (3.12) \\ &= \lim_{\epsilon \rightarrow 0} P(x(t_0 + \cdot) \in A \mid x(s) = u, \text{ upcrossing, some } s \in [t_0 - \epsilon, t_0]). \end{aligned}$$

It is easy to show that the two distributions just defined are identical,  $P^u(A) = P^{hw}(A)$ , for every (ergodic) stationary process such that there are only a finite number of  $u$ -upcrossings in any finite interval; see [9].

With point process terminology one can look at the upcrossings as a sequence of points in a stationary point process and the shape of the process around the upcrossing as a mark, attached to the point. The conditional distribution of the shape is then treated as a *Palm distribution* in the marked point process; see [9].

The term, horizontal window condition, is natural, since the process has to pass a horizontal window at level  $u$  somewhere near  $t_0$ . In analogy, the condition in (3.9) is called *vertical window condition*, since the process has to pass a vertical window near  $v$  exactly at time  $t_0$ .

The distribution  $P^u$  can be found via its finite-dimensional distribution functions. Take  $\mathbf{s} = (s_1, \dots, s_n)$ ,  $\mathbf{v} = (v_1, \dots, v_n)$ , write

$$x(\mathbf{s}) \leq \mathbf{v} \quad \text{for } x(s_j) \leq v_j, j = 1, \dots, n,$$

and define

$$N_{[0,T]}(x, u; \mathbf{s}, \mathbf{v}) = \#\{t_k; 0 \leq t_k \leq T, \text{ and } x(t_k + \mathbf{s}) \leq \mathbf{v}\},$$

as the number of  $u$ -upcrossings in  $[0, T]$  which are such that the process, at each of the times  $s_j$  after the upcrossing is less than  $v_j$ . Thus  $\frac{N_{[0,T]}(x, u; \mathbf{s}, \mathbf{v})}{N_{[0,T]}(x, u)}$ ,  $\mathbf{v} \in$

$\mathbb{R}^n$  is the empirical distribution function for the process at times  $s_j$  after  $u$ -upcrossings.

**Theorem 3:3** *If  $\{x(t), t \in \mathbb{R}\}$  is ergodic, with  $A = \{y \in \mathbb{C}; y(\mathbf{s}) \leq \mathbf{v}\}$ ,*

$$P^u(A) = \frac{E(N_{[0,1]}(x, u; \mathbf{s}, \mathbf{v}))}{E(N_{[0,1]}(x, u))} \quad (3.13)$$

$$= \frac{\int_{z=0}^{\infty} z f_{x(0), x'(0)}(u, z) P(x(\mathbf{s}) \leq \mathbf{v} \mid x(0) = u, x'(0) = z) dz}{\int_{z=0}^{\infty} z f_{x(0), x'(0)}(u, z) dz}. \quad (3.14)$$

**Proof:** We need a result from Chapter 5, namely that for an ergodic process (with probability one)

$$\begin{aligned} \frac{N_{[0,T]}(x, u)}{T} &\rightarrow E(N_{[0,1]}(x, u)) \\ \frac{N_{[0,T]}(x, u; \mathbf{s}, \mathbf{v})}{T} &\rightarrow E(N_{[0,1]}(x, u; \mathbf{s}, \mathbf{v})) \end{aligned}$$

as  $T \rightarrow \infty$ . This gives (3.13).

The proof of (3.14) is analogous to that of Theorem 3:2 and can be found in [22, Ch. 10] under the (unnecessarily strict) condition that  $\{x(t), t \in \mathbb{R}\}$  has continuously differentiable sample paths.  $\square$

Noting that

$$p_u(z) = \frac{z f_{x(0), x'(0)}(u, z)}{\int_{\zeta=0}^{\infty} \zeta f_{x(0), x'(0)}(u, \zeta) d\zeta}, \quad z \geq 0, \quad (3.15)$$

is a density function in  $z$  we can introduce a random variable  $\zeta$  with this density. Then (3.14) can be formulated as

$$P^u(A) = E_{\zeta}(P(x(\mathbf{s}) \leq \mathbf{v} \mid x(0) = u, x'(0) = \zeta)). \quad (3.16)$$

The interpretation is that the distribution of the process values at times  $\mathbf{s}$  after  $u$ -upcrossings is a mixture over the random slope  $\zeta$  at the upcrossing point, of the ordinary conditional distributions of  $x(\mathbf{s})$  given that  $x(0) = u$  and  $x'(0) = \zeta$ , when  $\zeta$  has density  $p_u(z)$ .

### 3.2.2 The Slepian model

Theorem 3:3 presents the conditional distribution of a stationary process in the neighborhood of upcrossings of the fixed level  $u$ . The definition in the form of integrals in (3.14) is not very informative as it is. However, for Gaussian processes one can construct an explicit and simple process that has exactly this distribution, and lends itself to easy simulation, numerical calculations, and asymptotic expansions.

**Definition 3:2** Let  $\{x(t), t \in \mathbb{R}\}$  be a stationary and ergodic stochastic process and fix a level  $u$ , such that  $x(t)$  has a finite number of  $u$ -upcrossings in any finite interval. A Slepian model process for  $\{x(t), t \in \mathbb{R}\}$  after  $u$ -upcrossings is any stochastic process  $\{\xi_u(t), t \in \mathbb{R}\}$  that has its distributions given by  $P^u$  in (3.14). In particular, its finite-dimensional distributions are given by

$$P(\xi_u(\mathbf{s}) \leq \mathbf{v}) = \int_0^\infty p_u(z) P(x(\mathbf{s}) \leq \mathbf{v} \mid x(0) = u, x'(0) = z) dz.$$

The Slepian model is a stochastic model for individual excursions after a level upcrossing, and its distribution is equal to the distribution of the sample functions illustrated in the lower diagram in Figure 3.2. More complex Slepian models can be formulated for other crossing problems, for example the process behavior after a local maximum or minimum, since these are defined as downcrossing or upcrossing of the zero level by the derivative process  $\{x'(t), t \in \mathbb{R}\}$ .

Every Slepian model has two elements, which depend on the type of crossing problem: the long run distribution of the gradient at the instances of the crossings, and the conditional distribution of the process given its value and the value of the gradient.

Typical problems that can be analyzed by a Slepian process are

- **Prediction after crossing:** What is the best predictor of the process a time  $\tau$  after one has observed a level  $u$  upcrossing?
- **Excursion shape:** How high above level  $u$  will an excursion extend, and how long does it take before the process returns below the level?
- **Crest shape:** What is the shape of the process near its local maxima?

We can immediately solve the first problem: the best predictor  $\hat{x}^u(t_0 + \tau)$  after  $u$ -upcrossings is the expectation of the Slepian model:

$$\hat{x}^u(t_0 + \tau) = E(\xi_u(\tau)), \quad (3.17)$$

in the sense that the average of  $(x(t_k + \tau) - a)^2$ , when  $t_k$  runs over all  $u$ -upcrossings, takes its minimum value when  $a = E(\xi_u(\tau))$ .

### 3.2.2.1 An explicit Slepian model for crossings in a Gaussian process

The conditional distribution of  $\{x(t), t \in \mathbb{R}\}$  after  $u$ -upcrossing is particularly simple in the Gaussian case, and the Slepian model can be expressed in a very explicit way.

We have to find the density  $p_u(z)$  and the conditional distributions of  $x(\mathbf{s})$  in (3.16). For Gaussian processes, formula (3.5) holds, and with mean  $\mu = 0$ , it says

$$f_{x(0), x'(0)}(u, z) = \frac{1}{2\pi\sqrt{\omega_0\omega_2}} e^{-u^2/2\omega_0} e^{-z^2/2\omega_2}.$$

Canceling  $\frac{1}{\sqrt{2\pi\omega_0}} e^{-u^2/2\omega_0}$  in (3.15) we get

$$p_u(z) = \frac{z}{\omega_2} e^{-z^2/\omega_2}, \quad z \geq 0, \quad (3.18)$$

i.e. the slope at upcrossings has a Rayleigh distribution with parameter  $\omega_2$ , regardless of the level  $u$ .<sup>5</sup>

Next, we need the conditional distribution of  $x(\mathbf{s})$  given  $x(0) = u$ ,  $x'(0) = z$  and average over  $\zeta = z$  with density  $p_u(z)$ . Since a conditional distribution in a multivariate normal distribution is still a normal distribution, we need only to find the conditional mean  $E(x(s) \mid x(0) = u, x'(0) = z)$  and the conditional covariances  $\text{Cov}(x(s_1), x(s_2) \mid x(0) = u, x'(0) = z)$ , and these were given in Section 1.5.1, equations (1.9, 1.11).

Take  $\boldsymbol{\xi} = (x(s_1), x(s_2))$ ,  $\boldsymbol{\eta} = (x(0), x'(0))$ , and calculate the joint covariance matrix of  $(\boldsymbol{\xi}, \boldsymbol{\eta})$  from the covariance function  $r(t)$  for  $\{x(t), t \in \mathbb{R}\}$ . By Theorem 2:13 it is

$$\boldsymbol{\Sigma} = \begin{pmatrix} r(0) & r(s_2 - s_1) & r(s_1) & -r'(s_1) \\ r(s_1 - s_2) & r(0) & r(s_2) & -r'(s_2) \\ r(s_1) & r(s_2) & r(0) & 0 \\ -r'(s_1) & -r'(s_2) & 0 & -r''(0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{pmatrix}.$$

Use the notation  $\omega_0 = r(0)$ ,  $\omega_2 = -r''(0)$  and remember that  $m_{\boldsymbol{\xi}} = m_{\boldsymbol{\eta}} = \mathbf{0}$ . Then we get the conditional expectation and covariance matrix given  $\boldsymbol{\eta} = \mathbf{y} = (u, z)$  as

$$E(\boldsymbol{\xi} \mid \boldsymbol{\eta} = \mathbf{y}) = \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} \mathbf{y}' = \begin{pmatrix} ur(s_1)/\omega_0 - zr'(s_1)/\omega_2 \\ ur(s_2)/\omega_0 - zr'(s_2)/\omega_2 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}|\boldsymbol{\eta}} = \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} - \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\xi}} = \begin{pmatrix} r_{\kappa}(s_1, s_1) & r_{\kappa}(s_1, s_2) \\ r_{\kappa}(s_2, s_1) & r_{\kappa}(s_2, s_2) \end{pmatrix},$$

say. Here

$$r_{\kappa}(s_1, s_2) = r(s_2 - s_1) - r(s_1)r(s_2)/\omega_0 - r'(s_1)r'(s_2)/\omega_2 \quad (3.19)$$

is the covariance function for a non-stationary process.

Note the structure of  $r_{\kappa}(s_1, s_2)$ : The first term is the unconditional covariance function of the process and the two other terms are the changes in covariance that is obtained by the knowledge of the uncorrelated  $x(0)$  and  $x'(0)$ . When  $s_1$  or  $s_2$  tend to infinity, the reduction terms go to 0 and the influence of the conditioning vanishes.

We have now found the explicit structure of the Slepian model in the Gaussian case; we formulate it as a theorem.

<sup>5</sup>This is the solution to Exercise 2:11.

**Theorem 3:4** a) The Slepian model for a Gaussian process  $\{x(t), t \in \mathbb{R}\}$  after  $u$ -upcrossings has the form

$$\xi_u(t) = \frac{ur(t)}{\omega_0} - \frac{\zeta r'(t)}{\omega_2} + \kappa(t), \quad (3.20)$$

where  $\zeta$  has the Rayleigh density  $p_\zeta(z) = (z/\omega_2)e^{-z^2/(2\omega_2)}$ , and  $\{\kappa(t), t \in \mathbb{R}\}$  is a non-stationary Gaussian process, independent of  $\zeta$ , with mean zero and covariance function  $r_\kappa(s_1, s_2)$  given by (3.19).

b) In particular, the best prediction of  $x(t_k + \tau)$  taken over all  $u$ -upcrossings  $t_k$ , is obtained by taking  $E(\zeta) = \sqrt{\pi\omega_2/2}$  and  $\widehat{\kappa}(\tau) = 0$  in (3.20) to get

$$\widehat{x}^u(t_0 + \tau) = \frac{ur(\tau)}{\omega_0} - \frac{E(\zeta)r'(\tau)}{\omega_2} = \frac{u}{\omega_0} r(\tau) - \sqrt{\frac{\pi}{2\omega_2}} r'(\tau). \quad (3.21)$$

We have now found the correct way of taking the apparent positive slope at a  $u$ -upcrossing into account in predicting the near future. Note that the simple formula (3.6),

$$\widehat{x}(t_0 + \tau) = E(x(t_0 + \tau) | x(t_0) = u) = \frac{u}{\omega_0} r(\tau),$$

in the Gaussian case, lacks the slope term. Of course, the slope at an upcrossing is always positive, but it is perhaps intuitively obvious that the observed slopes at the upcrossings are “more positive” than that. The slope = derivative of a stationary Gaussian process is normal with mean zero, but we do not expect slopes at fixed level upcrossings to have a half-normal distribution with a mode (= most likely values) at 0. The  $r'$ -term in (3.21) tells us how we shall take this “sample bias” into account.

The prediction of the slope is the expectation of the Rayleigh variable  $\zeta$ . If the slope at the  $u$ -upcrossing is observed and used in the prediction, then the difference between the two approaches disappears; see [23].

**Example 3:1** To illustrate the efficiency of the Slepian model, we shall analyse the shape of an excursion above a very high level  $u$  in a Gaussian process, and then expand the Slepian model  $\xi_u(t)$  in a Taylor series as  $u \rightarrow \infty$ . It will turn out that the length and height of the excursion will both be of the order  $u^{-1}$ , so we normalize the scales of  $\xi_u(t)$  by that factor. Using

$$r(t/u) = \omega_0 - \omega_2 \frac{t^2}{2u^2} (1 + o(1)), \quad r'(t/u) = -\omega_2 \frac{t}{u} (1 + o(1)),$$

as  $t/u \rightarrow 0$ , and that  $\kappa(t/u) = o(t/u)$ , and omitting all  $o$ -terms, we get

$$u \{\xi_u(t/u) - u\} = u \left\{ u \left( \frac{r(t/u)}{\omega_0} - 1 \right) - \zeta \frac{r'(t/u)}{\omega_2} + \kappa(t/u) \right\} \approx \zeta t - \frac{\omega_2 t^2}{2\omega_0}.$$

Thus, the excursion above a high level  $u$  takes the form of a parabola with height  $\frac{\zeta^2 \omega_0}{2u\omega_2}$  and length  $\frac{2\omega_0 \zeta}{u\omega_2}$ . It is easy to check that the normalized height of the excursion above  $u$  has an exponential distribution.

**Remark 3:4** *One should be aware that a Slepian model as it is described here represents the “marginal distribution” of the individual excursions above the defined level  $u$ . Of course one would like to use it also to analyse the dependence there may exist between successive excursions in the original process  $x(t)$ , and this is in fact possible. For example, suppose we want to find how often it happens that two successive excursions both exceed a critical limit  $T_0$  in length. Then, writing  $\tau_1 = \inf\{\tau > 0; \xi_u(\tau) = u, \text{ upcrossing}\}$  for the first  $u$ -upcrossing in  $\xi_u(t)$  strictly on the positive side, one can calculate*

$$P(\xi_u(s) > u, \text{ for } 0 < s < T_0, \text{ and } \xi_u(\tau_1 + s) > u, \text{ for } 0 < s < T_0).$$

### 3.2.2.2 A Slepian model around local maxima in a Gaussian process

A local maximum for  $x(t)$  is a zero-downcrossing point for the derivative  $x'(t)$ , and the second derivative  $x''(t)$  at these local maxima has a negative Rayleigh distribution. A Slepian model for the derivative after local maxima, therefore has the same structure as the level crossing model, with  $r_x(t)$  replaced by  $r_{x'}(t) = -r_x''(t)$ . The time difference between maximum and minimum can therefore be calculated as previously.

If we want the distribution of the *height difference* between the maximum and the following minimum, we need a more elaborate Slepian model, since now also the height of the maximum is random, not only the curvature. The reader is encouraged to prove the following theorem, copying Theorem 3:3 with analogous notation, now with  $t'_k$  for the times of local maxima

**Theorem 3:5** *If  $\{x(t), t \in \mathbb{R}\}$  is twice differentiable and ergodic the long run empirical distribution of  $x(t'_k + \mathbf{s})$  around local maxima is equal to*

$$P_1^{\max}(A) = \frac{\int_{z=-\infty}^0 |z| f_{x'(0), x''(0)}(0, z) P(x(\mathbf{s}) \leq \mathbf{v} \mid 0, z) dz}{\int_{-\infty}^0 |z| f_{x'(0), x''(0)}(0, z) dz},$$

where  $A = \{y \in \mathbb{C}; y(\mathbf{s}) \leq \mathbf{v}\}$  and

$$P(x(\mathbf{s}) \leq \mathbf{v} \mid 0, z) = P(x(\mathbf{s}) \leq \mathbf{v} \mid x'(0) = 0, x''(0) = z).$$

*An explicit representation of the model process for a Gaussian process is the*

$$\xi_1^{\max}(t) = \zeta_1^{\max} \frac{r''(t)}{\omega_4} + \Delta_1(t), \quad (3.22)$$

where  $\zeta_1^{\max}$  has a negative Rayleigh distribution with density

$$p_1^{\max}(z) = \frac{|z|}{\omega_4} e^{-z^2/(2\omega_4)}, \quad z < 0, \quad (3.23)$$

and the non-stationary Gaussian process  $\Delta_1(t)$  is independent of  $\zeta_1^{max}$ , has mean 0 and the covariance function,

$$r_{\Delta_1}(s_1, s_2) = r(s_1 - s_2) - \frac{r'(s_1)r'(s_2)}{\omega_2} - \frac{r''(s_1)r''(s_2)}{\omega_4}. \quad (3.24)$$

Since  $\Delta_1(0)$  is normal with mean 0 and variance  $r_{\Delta_1}(0, 0) = \omega_0 - \omega_2^2/\omega_4$ , we see in particular that the distribution of the height of a local maximum is the same as the distribution of

$$\begin{aligned} \xi^{max}(0) &= -\zeta_1^{max} \frac{\omega_2}{\omega_4} + \Delta^1(0) \\ &= \sqrt{\omega_0} \left\{ \sqrt{1 - \epsilon^2} \cdot \text{Rayleigh} + \epsilon \cdot \text{Normal} \right\}, \end{aligned} \quad (3.25)$$

with standard Rayleigh and normal variables, illustrating the relevance of the spectral width parameter  $\alpha = \sqrt{1 - \epsilon^2} = \sqrt{\omega_2^2/(\omega_0\omega_4)}$ .

Theorem 3:5 is the basis for numerical calculations of wave characteristic distributions like height and time difference between local maxima and minima, as they can be made by the routines in the MATLAB package WAFO, [34]. The model (3.22) contains an explicit function  $r''(t)/\omega_4$  with a simple random factor, representing the Rayleigh distributed curvature at the maximum, plus a continuous parameter Gaussian process. The numerical procedures work by successively replacing the continuous parameter process by explicit functions multiplied by random factors. We illustrate now the first steps in this procedure.

The model (3.22) contains the random curvature, and it is the simplest form of the Slepian model after maximum. There is nothing that prevents us to include also the random height of the local maximum in the model. We have seen in (3.25) how the height and the curvature depend on each other, so we can build an alternative Slepian model after maximum that explicitly includes both the height of the maximum and the curvature.

To formulate the extended model we define three functions,  $A(t), B(t), C(t)$ , by

$$\begin{aligned} E(x(t) \mid x(0) = u, x'(0) = y, x''(0) = z) \\ &= uA(t) + yB(t) + zC(t) \\ &= u \frac{\omega_4 r(t) + \omega_0 \omega_2 r''(t)}{\omega_0 \omega_4 - \omega_2^2} - y \frac{r'(t)}{\omega_2} + z \frac{\omega_2 r(t) + \omega_0 r''(t)}{\omega_0 \omega_4 - \omega_2^2}. \end{aligned}$$

The conditional covariance between  $x(s_1)$  and  $x(s_2)$  are found from the same theorem, and the explicit expression is given in the following Theorem 3:6.

As we have seen in Section 2.4.3 the derivative  $x'(0)$  is uncorrelated with both  $x(0)$  and  $x''(0)$ , but  $x(0)$  and  $x''(0)$  are correlated. To formulate the effect of observing a local maximum we will first introduce the crest height,

$x(0)$ , and then find the conditional properties of  $x''(0)$  given  $x(0)$  and  $x'(0)$ . We use Theorem 2:13 and define the function

$$b(t) = \frac{\text{Cov}(x(t), x''(t) \mid x(0), x'(0))}{\sqrt{V(x''(0) \mid x(0), x'(0))}} = \frac{r''(t) + (\omega_2/\omega_0)r(t)}{\sqrt{\omega_4 - \omega_2^2/\omega_0}}.$$

**Theorem 3:6** *If  $\{x(t), t \in \mathbb{R}\}$  is twice differentiable and ergodic the long run empirical distribution of  $x(t'_k + \mathbf{s})$  around local maxima is equal to*

$$P_2^{\max}(A) = \frac{\int_{u=-\infty}^{\infty} \int_{z=-\infty}^0 |z| f_{x(0), x'(0), x''(0)}(u, 0, z) P(x(\mathbf{s}) \leq \mathbf{v} \mid u, 0, z) dz du}{\int_{u=-\infty}^{\infty} \int_{z=-\infty}^0 |z| f_{x(0), x'(0), x''(0)}(u, 0, z) dz du},$$

where  $A = \{y \in \mathbb{C}; y(\mathbf{s}) \leq \mathbf{v}\}$  and

$$P(x(\mathbf{s}) \leq \mathbf{v} \mid u, 0, z) = P(x(\mathbf{s}) \leq \mathbf{v} \mid x(0) = u, x'(0) = 0, x''(0) = z).$$

An explicit representation of the model process for a Gaussian process is the

$$\xi_2^{\max}(t) = \eta_2^{\max} A(t) + \zeta_2^{\max} C(t) + \Delta_2(t), \quad (3.26)$$

where the random  $(\eta_2^{\max}, \zeta_2^{\max})$  has the two-dimensional density (with normalizing constant  $c$ ),

$$p_2^{\max}(u, z) = c|z| \exp \left\{ -\frac{\omega_0 z^2 + 2\omega_2 u z + \omega_4 u^2}{2(\omega_0 \omega_4 - \omega_2^2)} \right\}, \quad -\infty < u < \infty, z < 0.$$

The process  $\Delta_2(t)$  with mean zero is non-stationary Gaussian, independent of  $(\eta_2^{\max}, \zeta_2^{\max})$ , and has covariance function

$$r_{\Delta_2}(s, t) = r(s-t) - \frac{r(s)r(t)}{\omega_0} - \frac{r'(s)r'(t)}{\omega_2} - b(s)b(t).$$

### 3.2.3 Excursions and related distributions

What is the shape, height and extension, of an arbitrary excursion above a fixed level  $u$  for a stationary process? An example of this type of problem is when  $x(t)$  is an electrical potential which may not become too high, and should always stay below a certain level. The integral  $\int_{t_k}^{t_k+T_k} (x(t) - u) dt$  between an excursion and the critical level represents the amount of extra electrical charges that are transmitted between the upcrossing at  $t_k$  and the next downcrossing at  $t_k + T_k$ .

One of the advantages of a Slepian model is that it lends itself to efficient numerical calculations of important quantities related to crossings and maxima. The structure of the model is such that the random variables that represent slope, height, and curvature in the crossings and crest models are easy to handle

numerically. The only problems are the residual processes which require infinite dimensional probabilities to be calculated. To overcome this in a numerical algorithm one can use a successive conditioning technique that first introduces the value of the normal residual at a single point, say,  $\kappa(s_1)$ , and include that as a separate term in the model. The residual process will be correspondingly reduced and the procedure repeated.

For numerical calculations of interesting crossings probabilities one can truncate the conditioning procedure when sufficient accuracy is attained. This approximation technique is called *regression approximation* in crossing theory.

### 3.2.3.1 Length of excursions

The Slepian process  $\xi_u(t)$  has a  $u$ -upcrossing at  $t = 0$  and we denote by  $T$  the time of its first downcrossing of the same level, so  $T$  is the length of an excursion. Since  $T > t$  if and only if  $\xi_u(s)$  stays above the level  $u$  in the entire interval  $0 < s < t$ , we can express the probability  $P(T > t)$  by means of the indicator function

$$I_z(\kappa, t) = \begin{cases} 1, & \text{if } \frac{ur(s)}{\omega_0} - \frac{zr'(s)}{\omega_2} + \kappa(s) > u, \text{ for all } s \in (0, t), \\ 0, & \text{otherwise.} \end{cases}$$

The result is,

$$P(T > t) = \int_{z=0}^{\infty} p_u(z) \cdot E(I_z(\kappa, t)) dz, \quad (3.27)$$

where  $p_u(z)$  is the Rayleigh density for the derivative at  $u$ -upcrossing, and the expectation

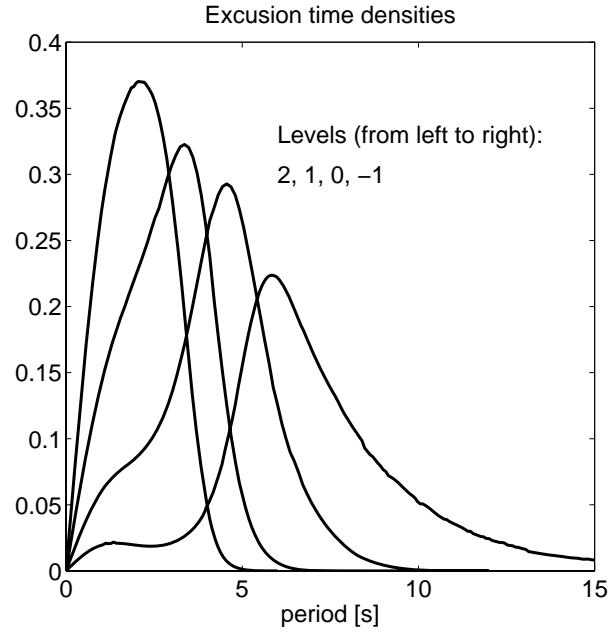
$$E(I_z(\kappa, t)) = P\left(\inf_{0 < s < t} \left\{ \frac{ur(s)}{\omega_0} - \frac{zr'(s)}{\omega_2} + \kappa(s) \right\} > u\right)$$

is an infinite dimensional normal probability. That probability has to be calculated numerically by special software. By means of routines in the MATLAB package WAFO, [34], which is available on the departments homepage, one can calculate the distribution with very high accuracy. Figure 3.3 shows the excursion length densities for a realistic water wave process with a common JONSWAP North Sea spectrum.

The probability density function for the excursion time  $T$  is of course minus the derivative of  $P(T > t)$ . It can also be expressed by means of *Durbin's formula*,

$$f_T(t) = f_{\xi_u(t)}(u) E\left(I\{\xi_u(s) > u, 0 < s < t\} \cdot (-\xi'_u(t)^-) \mid \xi_u(t) = u\right),$$

where  $\xi'_u(t)^- = \min(0, \xi'_u(t))$  is the negative part of the derivative. The expectation can be calculated by mean of algorithms from WAFO, by means of the regression technique with successive conditioning on the residual process.



**Figure 3.3:** Probability densities for excursions above  $u = -1, 0, 1, 2$  for process with North Sea wave spectrum JONSWAP.

### 3.2.3.2 Wave shape

The distribution of wave characteristics such as drop in height and time difference between a local maximum and the next local minimum can be derived from the Slepian models in Theorem 3:5 or Theorem 3:6.

First consider the model (3.22),

$$\xi_1^{max}(t) = \zeta_1^{max} \frac{r''(t)}{\omega_4} + \Delta_1(t),$$

which completely describes the stochastic properties of the shape around maximum. The simplest, zero order, approximation is to delete the residual process  $\Delta_1(t)$  completely, only keeping the curvature dependent term  $\zeta_1^{max} \frac{r''(t)}{\omega_4}$ . By replacing  $\zeta_1^{max}$  by its average  $-\sqrt{\omega_4\pi}/2$ . From this we can, for example, get the average shape, as

$$\widehat{\xi}^{max}(t) = -\sqrt{\omega_0} \alpha \frac{r''(t)}{\omega_2}.$$

The zero order approximation is usually too crude to be of any use. A better approximation is obtained from the model (3.26), which also includes the (random) height at the maximum point,

$$\xi_2^{max}(t) = \eta_2^{max} A(t) + \zeta_2^{max} C(t) + \Delta_2(t),$$

and define the random variable  $T$  as the time of the first local minimum of  $\xi_2^{max}(t)$ ,  $t > 0$ . The height drop is then  $H = \xi_2^{max}(0) - \xi_2^{max}(T)$  and we ask for the joint distribution of  $T$  and  $H$ .

Using the fact that  $A(0) = 1, C(0) = 0$  and  $\xi_2^{max}(T)' = 0$ , we get the following relations that need to be satisfied,

$$\begin{aligned} \eta_2^{max} A'(T) + \zeta_2^{max} C'(T) + \Delta_2'(T) &= 0, \\ \eta_2^{max} + \Delta_2(0) - (\eta_2^{max} A(T) + \zeta_2^{max} C(T) + \Delta_2(T)) &= H. \end{aligned}$$

We now describe the regression approximation of order 1, which is obtained by deleting all of the residual process terms. The relations will then be

$$\eta_2^{max} A'(T^r) + \zeta_2^{max} C'(T^r) = 0, \quad (3.28)$$

$$\eta_2^{max} - (\eta_2^{max} A(T^r) + \zeta_2^{max} C(T^r)) = H^r, \quad (3.29)$$

where we write  $T^r, H^r$  for the approximative time and height variables.

To write the solution in a form that can be generalized to more complicated problems, define

$$G(t) = \begin{pmatrix} 1 - A(t) & C(t) \\ A'(t) & C'(t) \end{pmatrix}$$

and write the equations (3.28) and (3.29) as ( $T$  for matrix transpose),

$$G(T^r)(\eta_2^{max} \zeta_2^{max})^T = (H^r T^r)^T.$$

If  $\det G(T^r) \neq 0$  we get from  $(\eta_2^{max} \zeta_2^{max})^T = G(T^r)^{-1}(H^r 0)^T$  that the variables with known distribution ( $\eta_2^{max}$  and  $\zeta_2^{max}$ ) are simple functions of the variables with unknown distribution,

$$\eta_2^{max} = H^r p(T^r) q(T^r), \quad \zeta_2^{max} = H^r q(T^r),$$

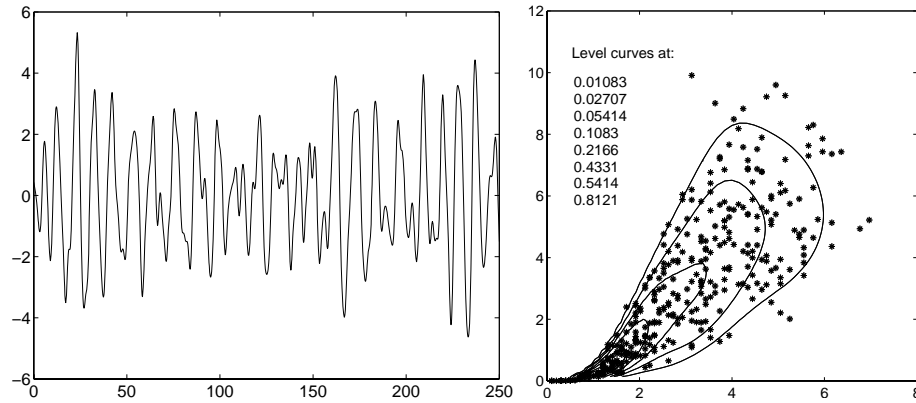
where

$$p(t) = \frac{-C'(t)}{A'(t)}, \quad q(t) = \frac{-A'(t)}{(1 - A(t))C'(t) - A'(t)C(t)}.$$

We want the density at the point  $T^r = t, H^r = h$ ; let  $\xi(t, h), \zeta(t, h)$  be the corresponding solution and define the indicator function  $I(t, h)$  to be 1 if the approximating process  $\xi(t, h)A(s) + \zeta(t, h)(s)$  is strictly decreasing for  $0 < s < t$ .

The Jacobian for the transformation is  $J(t, h) = hp'(t)q(t)^2$ , and therefore the density of  $T^r, H^r$  is

$$\begin{aligned} f_{T^r, H^r}(t, h) &= f_{\xi_{max}, \zeta_{max}}(hp(t)q(t), hq(t)) \cdot |J(t, h)| I(t, h) \\ &= \text{const } I(t, h) h^2 |q^2(t)|^3 p'(t) \\ &\quad \times \exp \left\{ -\frac{1}{2\varepsilon^2} h^2 q(t)^2 (T_m/\pi)^4 \left( ((\pi/T_m)^2 p(t) + 1)^2 + \frac{\varepsilon^2}{1 - \varepsilon^2} \right) \right\}. \end{aligned}$$



**Figure 3.4:** *Probability density for  $T, H$  for process with North Sea wave spectrum JONSWAP together with 343 observed cycles.*

This form of the  $T, H$  distribution is common in the technical literature, where  $T_m = \pi\sqrt{\omega_2/\omega_4}$  is called the *mean half wave period*. Note that dependence on the spectrum is only through the spectral width parameter  $\epsilon = \sqrt{1 - \frac{\omega_2^2}{\omega_0\omega_4}} = \sqrt{1 - \alpha^2}$ .

This first order approximation of the  $T, H$ -density is not very accurate, but it illustrates the basic principle of the regression approximation. The WAFO toolbox, [34], contains algorithms for very accurate higher order approximations. Figure 3.4 shows the result for a process with a common North Sea JONSWAP spectrum.

**Exercises**

- 3:1. Prove that  $\kappa(0) = \kappa'(0) = 0$  in the Slepian model after upcrossing.
- 3:2. Formulate conditions on the covariance function  $r_x(t)$  that guarantee that the residual process  $\kappa(t)$  has differentiable sample paths.
- 3:3. Complete the proof of Theorem 3:5.

## Chapter 4

# Spectral- and other representations

This chapter deals with the spectral representation of weakly stationary processes – stationary in the sense that the mean is constant and the covariance  $Cov(x(s), x(t))$  only depends on the time difference  $t - s$ . For real-valued Gaussian processes, the mean and covariance function determines all finite-dimensional distributions, and hence the entire process distribution. However, the spectral representation requires complex-valued processes, and then one needs to specify also the correlation structure between the real and the imaginary part of the process. We therefore start with a summary of the basic properties of complex-valued processes, in general, and in the Gaussian case. We remind the reader of the classical memoirs by S.O. Rice, [27], which can be recommended to anyone with the slightest historical interest. That work also contains many old references.

### 4.1 Complex processes and their covariance functions

#### 4.1.1 Stationary processes

A complex-valued process

$$x(t) = y(t) + iz(t)$$

is *strictly stationary* if all  $2n$ -dimensional distributions of

$$y(t_1 + \tau), z(t_1 + \tau), \dots, y(t_n + \tau), z(t_n + \tau)$$

are independent of  $\tau$ . It is called *weakly stationary* or *second order stationary* if  $E(x(t)) = m$  is constant, and

$$E(x(s) \cdot \overline{x(t)}) = r(s - t) + |m|^2$$

only depends on the time difference  $s - t$ . The *covariance function*

$$r(\tau) = E\left((x(s + \tau) - m)\overline{(x(s) - m)}\right)$$

is *Hermitian*, i.e.

$$r(-\tau) = \overline{r(\tau)}.$$

For real-valued processes, the covariance function  $r(\tau)$  determines all covariances between  $x(t_1), \dots, x(t_n)$ ,

$$\begin{aligned} \Sigma(t_1, \dots, t_n) &= \begin{pmatrix} r(0) & r(t_1 - t_2) & \dots & r(t_1 - t_n) \\ \vdots & \vdots & \ddots & \vdots \\ r(t_n - t_1) & r(t_n - t_2) & \dots & r(0) \end{pmatrix} \\ &= \begin{pmatrix} V(x(t_1)) & C(x(t_1), x(t_2)) & \dots & C(x(t_1), x(t_n)) \\ \vdots & \vdots & \ddots & \vdots \\ C(x(t_n), x(t_1)) & C(x(t_n), x(t_2)) & \dots & V(x(t_n)) \end{pmatrix}. \end{aligned} \quad (4.1)$$

#### 4.1.2 Non-negative definite functions

It is a unique characteristic property of a covariance function that is it *non-negative definite* in the following sense: Let  $t_1, \dots, t_n$  be any finite set of time points, and take arbitrary complex numbers  $a_1, \dots, a_n$ . Then, for simplicity assuming  $E(x(t)) = 0$ ,

$$\sum_{j,k} a_j \overline{a_k} r(t_j - t_k) = E\left(\sum_{j,k} a_j x(t_j) \overline{a_k x(t_k)}\right) \quad (4.2)$$

$$= E\left|\sum_{j=1}^n a_j x(t_j)\right|^2 \geq 0. \quad (4.3)$$

**Theorem 4:1** *Every non-negative definite, possibly complex, function  $r(\tau)$  is the covariance function for a strictly stationary Gaussian process. Thus, the class of covariance functions is equal to the class of non-negative definite functions.*

**Proof:** We have to show that if  $r(\tau)$  is non-negative definite then there are finite-dimensional distributions for a process  $x(t)$  with  $E(x(t)) = 0$ , such that

$$r(\tau) = E(x(s + \tau)\overline{x(s)}).$$

By Kolmogorov's existence theorem, see Appendix A, we only have to show that for every selection of time points  $t_1, \dots, t_n$ , there is an  $n$ -dimensional

distribution with mean 0 and covariances given by  $\Sigma(t_1, \dots, t_n)$  as defined by (4.1), and such that the obtained family of distributions forms a consistent family, i.e. for example  $F_{X,Y}(x, y) = F_{Y,X}(y, x)$  and  $F_{X,Y}(x, \infty) = F_X(x)$ .

If  $r(t)$  is a real function and  $\mathbf{u} = (u_1, \dots, u_n)$  a real vector, consider the non-negative quadratic form

$$Q(\mathbf{u}) = \sum_{j,k} u_j u_k r(t_j - t_k).$$

Then we recognize  $\exp(-Q(\mathbf{u})/2)$  as the characteristic function for an  $n$ -variate normal distribution with covariance matrix  $\Sigma(t_1, \dots, t_n)$ , and we have found a distribution with the specified properties.

If  $r(t)$  is complex, with  $r(-t) = \overline{r(t)}$ , there are real functions  $p(t) = p(-t)$ ,  $q(t) = -q(-t)$  such that

$$r(t) = p(t) + iq(t).$$

Take  $a_j = u_j - iv_j$  and consider the non-negative quadratic form in real variables  $(\mathbf{u}, \mathbf{v}) = (u_1, \dots, u_n, v_1, \dots, v_n)$ ,

$$\begin{aligned} Q(\mathbf{u}, \mathbf{v}) &= \sum_{j,k} a_j \overline{a_k} r(t_j - t_k) \\ &= \sum_{j,k} (u_j - iv_j)(u_k + iv_k)(p(t_j - t_k) + iq(t_j - t_k)) \\ &= \sum_{j,k} \{p(t_j - t_k)(u_j u_k + v_j v_k) - q(t_j - t_k)(u_j v_k - u_k v_j)\}; \end{aligned}$$

note that the imaginary part vanishes since  $Q$  is assumed to be non-negative, and hence real. Similarly, as in the real case,

$$\exp(-Q(\mathbf{u}, \mathbf{v})/2) = E \left( \exp(i \sum_j (u_j y_j + v_j z_j)) \right)$$

is the characteristic function of a  $2n$ -dimensional normal variable

$$(y_1, \dots, y_n, z_1, \dots, z_n)$$

with the specified properties:

$$\begin{aligned} E(y_j y_k) &= E(z_j z_k) = p(t_j - t_k) \\ E(y_j z_k) &= -E(y_k z_j) = -q(t_j - t_k). \end{aligned}$$

With  $x_j = (y_j + iz_j)/\sqrt{2}$ , we have  $E(x_j) = 0$  and

$$E(x_j \overline{x_k}) = p(t_j - t_k) + iq(t_j - t_k) = r(t_j - t_k),$$

as required. Since the Gaussian distribution of the  $y$ - and  $z$ -variables is determined by the covariances, and these depend only on the time difference, the process is strictly stationary.  $\square$

### 4.1.3 Strict and weak stationarity

Since the first two moments determines a real normal distribution, it is clear that each weakly (covariance) stationary normal process is strictly stationary. For complex processes matters are not that easy, and one has to impose two stationarity conditions in order to guarantee strict stationarity.

**Theorem 4:2** *A complex normal process  $x(t) = y(t) + iz(t)$  with mean zero is strictly stationary if and only if the two functions*

$$\begin{aligned} r(s, t) &= E(x(s)\overline{x(t)}) \\ q(s, t) &= E(x(s)x(t)), \end{aligned}$$

*only depend on  $t - s$ .*

**Proof:** To prove the "if" part, express  $r(s, t)$  and  $q(s, t)$  in terms of  $y$  and  $z$ ,

$$\begin{aligned} r(s, t) &= E(y(s)y(t) + z(s)z(t)) + iE(z(s)y(t) - y(s)z(t)), \\ q(s, t) &= E(y(s)y(t) - z(s)z(t)) + iE(z(s)y(t) + y(s)z(t)). \end{aligned}$$

Since these only depend on  $t - s$ , the same is true for the sums and differences of their real and imaginary parts, i.e. for  $E(y(s)y(t))$ ,  $E(z(s)z(t))$ ,  $E(z(s)y(t))$ ,  $E(y(s)z(t))$ . Therefore, the  $2n$ -dimensional distribution of

$$y(t_1), \dots, y(t_n), z(t_1), \dots, z(t_n)$$

only depends on time differences, and  $x(t)$  is strictly stationary. The converse is trivial.  $\square$

**Example 4:1** If  $x(t)$  is a real and stationary normal process, and  $\mu$  is a constant, then

$$x^*(t) = e^{i\mu t}x(t)$$

is a weakly, but not strictly, stationary complex normal process,

$$\begin{aligned} E(x^*(s)\overline{x^*(t)}) &= e^{i\mu(s-t)}E(x(s)x(t)), \\ E(x^*(s)x^*(t)) &= e^{i\mu(s+t)}E(x(s)x(t)). \end{aligned}$$

## 4.2 Bochner's theorem and the spectral distribution

### 4.2.1 The spectral distribution

We have seen that covariance functions for stationary processes are characterized by the property of being non-negative definite. From elementary courses we also know that covariance functions are Fourier-transforms of their *spectral distributions*. We shall now formulate this and prove this statement.

**Theorem 4:3** (Bochner's theorem) *A continuous function  $r(t)$  is non-negative definite, and hence a covariance function, if and only if there exists a non-decreasing, right continuous, and bounded real function  $F(\omega)$  such that*

$$r(t) = \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega).$$

The function  $F(\omega)$  is the spectral distribution function of the process, and it has all the properties of a statistical distribution function except that  $F(+\infty) - F(-\infty) = r(0)$  need not be equal to one. The function  $F(\omega)$  is defined only up to an additive constant, and one usually takes  $F(-\infty) = 0$ .

**Proof:** The "if" part is clear, since if  $r(t) = \int \exp(i\omega t) dF(\omega)$ , then

$$\begin{aligned} \sum_{j,k} z_j \bar{z}_k r(t_j - t_k) &= \sum_{j,k} z_j \bar{z}_k \int e^{i\omega t_j} \cdot e^{-i\omega t_k} dF(\omega) \\ &= \int \sum_{j,k} z_j e^{i\omega t_j} \overline{z_k e^{i\omega t_k}} dF(\omega) \\ &= \int \left| \sum_j z_j e^{i\omega t_j} \right|^2 dF(\omega) \geq 0. \end{aligned}$$

For the "only if" part we shall use some properties of characteristic functions, which are proved elsewhere in the probability course. We shall show that, given  $r(t)$ , there exists a proper distribution function  $F_\infty(\omega) = F(\omega)/F(\infty)$  such that

$$\begin{aligned} F_\infty(\infty) - F_\infty(-\infty) &= 1, \\ \int e^{i\omega t} dF_\infty(\omega) &= \frac{r(t)}{r(0)}. \end{aligned}$$

To this end, take a real  $A > 0$ , and define

$$\begin{aligned} g(\omega, A) &= \frac{1}{2\pi A} \int_0^A \int_0^A r(t-u) e^{-i\omega(t-u)} dt du \\ &= \frac{1}{2\pi A} \lim \sum_{j,k} r(t_j - t_k) e^{-i\omega t_j} \overline{e^{-i\omega t_k}} \Delta t_j \Delta t_k \\ &= \frac{1}{2\pi A} \lim \sum_{j,k} r(t_j - t_k) \Delta t_j e^{-i\omega t_j} \cdot \overline{\Delta t_k e^{-i\omega t_k}} \geq 0, \end{aligned}$$

since  $r(t)$  is non-negative definite. (Here, of course, the  $t_j$  define a subdivision of the interval  $[0, A]$ .) Going to the limit,  $g$  will give the density of the desired spectral distribution.

Before we proceed, we express  $g(\omega, A)$  as

$$\begin{aligned} g(\omega, A) &= \frac{1}{2\pi A} \int_0^A \int_0^A r(t-u)e^{-i\omega(t-u)} dt du \\ &= \frac{1}{2\pi} \int_{-A}^A \left(1 - \frac{|t|}{A}\right) r(t)e^{-i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mu(t/A)r(t)e^{-i\omega t} dt, \end{aligned}$$

where

$$\mu(t) = \begin{cases} 1 - |t| & \text{for } |t| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The proof will now proceed in three steps:

**Step 1)** Prove that  $g(\omega, A) \geq 0$  is integrable, and

$$\int_{\omega} g(\omega, A) d\omega = r(0), \quad (4.4)$$

so  $g(\cdot, A)/r(0)$  is a regular statistical density function.

**Step 2)** Show that

$$\left(1 - \frac{|t|}{A}\right) \frac{r(t)}{r(0)} = \int_{-\infty}^{\infty} \frac{g(\omega, A)}{r(0)} e^{it\omega} d\omega, \quad (4.5)$$

so the function  $\left(1 - \frac{|t|}{A}\right) r(t)/r(0)$  for  $|t| \leq A$  is the characteristic function for the density  $g(\omega, A)/r(0)$ .

**Step 3)** Take limits as  $A \rightarrow \infty$ ,

$$\lim_{A \rightarrow \infty} \left(1 - \frac{|t|}{A}\right) r(t) = r(t). \quad (4.6)$$

Since the limit of a convergent sequence of characteristic functions is also a characteristic function, provided it is continuous, we have shown that there exists a statistical distribution such that  $r(t)/r(0)$  is its characteristic function.

We now show step (1) and (2). Multiply  $g(\omega, A)$  by  $\mu(\omega/2M)$ , integrate, and change the order of integration (since  $\mu(\omega/2M)\mu(t/A)r(t)e^{-i\omega t}$  is bounded and has support in  $[-2M, 2M] \times [-A, A]$  Fubini's theorem permits this):

$$\begin{aligned} \int_{-\infty}^{\infty} \mu(\omega/2M)g(\omega, A) d\omega &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mu(\omega/2M) \int_{-\infty}^{\infty} \mu(t/A)r(t)e^{-i\omega t} dt d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mu(t/A)r(t) \int_{-\infty}^{\infty} \mu(\omega/2M)e^{-i\omega t} d\omega dt \end{aligned} \quad (4.7)$$

Here,

$$\begin{aligned} \int_{-\infty}^{\infty} \mu(\omega/2M)e^{-i\omega t} d\omega &= \int_{-2M}^{2M} \left(1 - \frac{|\omega|}{2M}\right) e^{-i\omega t} d\omega \\ &= \int_{-2M}^{2M} \left(1 - \frac{|\omega|}{2M}\right) \cos \omega t d\omega = 2M \left(\frac{\sin Mt}{Mt}\right)^2, \end{aligned}$$

so (4.7) is equal to

$$\begin{aligned} \frac{M}{\pi} \int_{-\infty}^{\infty} \mu(t/A)r(t) \left(\frac{\sin Mt}{Mt}\right)^2 dt &= \frac{1}{\pi} \int_{-\infty}^{\infty} \mu(s/MA)r(s/M) \left(\frac{\sin s}{s}\right)^2 ds \\ &\leq \frac{1}{\pi} r(0) \int_{-\infty}^{\infty} \left(\frac{\sin s}{s}\right)^2 ds = r(0). \end{aligned}$$

Now,  $\mu(\omega/2M)g(\omega, A) \uparrow g(\omega, A)$  as  $M \uparrow \infty$ , so

$$\int_{-\infty}^{\infty} g(\omega, A) d\omega = \lim_{M \rightarrow \infty} \int_{-\infty}^{\infty} \mu(\omega/2M)g(\omega, A) d\omega \leq r(0).$$

We have now shown that  $g(\omega, A)$  and  $\mu(t/A)r(t)$  are both absolutely integrable over the whole real line. Since they form a Fourier transform pair, i.e.

$$g(\omega, A) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mu(t/A)r(t)e^{-i\omega t} dt,$$

we can use the Fourier inversion theorem, which states that

$$\mu(t/A)r(t) = \int_{-\infty}^{\infty} g(\omega, A)e^{i\omega t} d\omega,$$

which is step (2) in the proof.

By taking  $t = 0$  we also get step (1), and  $f_A(\omega) = g(\omega, A)/r(0)$  is a probability density function for some distribution with characteristic function

$$\phi_A(t) = \int_{-\infty}^{\infty} \frac{g(\omega, A)}{r(0)} e^{i\omega t} d\omega = \frac{\mu(t/A)}{r(0)} r(t).$$

For step (3) we need one of the basic lemmas in probability theory, the convergence properties of characteristic functions: if  $F_A(x)$  is a family of distribution functions with characteristic functions  $\phi_A(t)$ , and  $\phi_A(t)$  converges to a continuous function  $\phi(t)$ , as  $A \rightarrow \infty$ , then there exists a distribution function  $F(x)$  with characteristic function  $\phi(t)$  and  $F_A(x) \rightarrow F(x)$ , for all  $x$  where  $F(x)$  is continuous.

Here the characteristic functions  $\phi_A(t) = \frac{\mu(t/A)}{r(0)} r(t)$  converge to  $\phi(t) = r(t)/r(0)$ , and since we have assumed  $r(t)$  to be continuous, we know from the

basic lemma that  $F_A(x) = \int_{-\infty}^x f_A(\omega) d\omega$  converges to a distribution function  $F_\infty(x)$  as  $A \rightarrow \infty$ , with characteristic function  $\phi(t)$ :

$$\frac{r(t)}{r(0)} = \int_{-\infty}^{\infty} e^{i\omega t} dF_\infty(\omega).$$

We get the desired spectral representation with  $F(\omega) = r(0)F_\infty(\omega)$ .  $\square$

## 4.2.2 Properties of the spectral distribution

### 4.2.2.1 The inversion theorem

The covariance function  $r(t)$  and the spectral density  $f(\omega)$  form a Fourier transform pair. In general, the spectral distribution is uniquely determined by the covariance function but the precise relationship is somewhat complicated if the spectrum is not absolutely continuous. To formulate a general *inversion theorem* we need to identify those  $\omega$  for which the spectral distribution function is not continuous.<sup>1</sup> Write  $\Delta F_\omega = F(\omega) - F(\omega - 0) \geq 0$  for the jump (possibly 0) at  $\omega$ , and define  $\tilde{F}(\omega)$  as the average between the left and right limits of  $F(\omega)$ ,

$$\tilde{F}(\omega) = \frac{F(\omega) + F(\omega - 0)}{2} = F(\omega) - \frac{1}{2}\Delta F_\omega. \quad (4.8)$$

For a proof of the following theorem we refer to [12].

**Theorem 4:4** a) If  $\omega_1 < \omega_2$ , then we have

$$\tilde{F}(\omega_2) - \tilde{F}(\omega_1) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{-it} r(t) dt. \quad (4.9)$$

b) If the covariance function  $r(t)$ ,  $t \in \mathbb{R}$  is absolutely integrable, i.e.

$$\int_{-\infty}^{\infty} |r(t)| dt < \infty,$$

then the spectrum is absolutely continuous and the Fourier inversion formula holds,

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} r(t) dt. \quad (4.10)$$

**Remark 4:1** The inversion formula (4.9) defines the spectral distribution for all continuous covariance functions. One can also use (4.10) to calculate the spectral density in case  $r(t)$  is absolutely integrable, but if it not, one may use 4.9) and take  $\tilde{f}(\omega) = \lim_{h \rightarrow 0} (\tilde{F}(\omega + h) - \tilde{F}(\omega))/h$ . This is always possible, but one has to be careful in case  $f(\omega)$  is not continuous. Even when the limit  $\tilde{f}(\omega)$  exists it need not be equal to  $f(\omega)$  as the following example shows. The limit, which always exists, is called the Cauchy principal value.

<sup>1</sup>Note that  $F(\omega)$  can have only a denumerable number of discontinuity points.

**Example 4:2** We use (4.9) to find the spectral density of low frequency white noise, with covariance function  $r(t) = \frac{\sin t}{t}$ . We get

$$\begin{aligned} \frac{\tilde{F}(\omega + h) - \tilde{F}(\omega - h)}{2h} &= \frac{1}{2\pi} \frac{1}{2h} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-i(\omega+h)t} - e^{-i(\omega-h)t}}{-it} \frac{\sin t}{t} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \frac{\sin ht}{ht} \frac{\sin t}{t} dt \\ &= \begin{cases} \frac{1}{2}, & \text{for } |\omega| < 1 - h, \\ \frac{1}{4}(1 + (1 - |\omega|)/h), & \text{for } 1 - h < |\omega| < 1 + h, \\ 0, & \text{for } |\omega| > 1 + h. \end{cases} \end{aligned}$$

The limit as  $h \rightarrow 0$  is  $1/2$ ,  $1/4$ , and  $0$ , respectively, which gives as spectral density

$$f(\omega) = \begin{cases} 1/2, & \text{for } |\omega| < 1, \\ 1/4, & \text{for } |\omega| = 1, \\ 0, & \text{for } |\omega| > 1. \end{cases}$$

Note that the Fourier inversion formula (4.10) gives  $1/4$  for  $\omega = 1$  as the Cauchy principal value,

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T e^{-i\omega t} r(t) dt = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{\sin 2t}{2t} dt = 1/4.$$

**4.2.2.2 The one-sided real form**

A real stationary process has a symmetric covariance function and a symmetric spectrum. In practical applications one often uses only the positive side of the spectrum. The one-sided spectral distribution will be denoted by  $G(\omega)$ , with

$$G(\omega) = \begin{cases} 0 & \text{for } \omega < 0, \\ F(\omega) - F(-\omega - 0) = 2F(\omega) - r(0) & \text{for } \omega \geq 0. \end{cases} \tag{4.11}$$

Then  $G(0-) = 0$  and  $G(\infty) = r(0)$ . If  $F(\omega)$  is discontinuous at  $\omega = 0$  then  $G(\omega)$  will have a jump  $F(0) - F(-0) = G(0+)$  at  $\omega = 0$ . For discontinuity points  $\omega > 0$  the jump of  $G(\omega)$  will be twice that of  $F(\omega)$ .

The covariance function can be expressed as

$$r(t) = \int_{0-}^{\infty} \cos \omega t dG(\omega) = \int_{-\infty}^{\infty} \cos \omega t dF(\omega),$$

and the inversion formula (4.9) gives that for any continuity points

$$G(\omega) = F(\omega) - F(-\omega) = \frac{2}{\pi} \int_0^{\infty} \frac{\sin \omega t}{t} r(t) dt.$$

Note that  $\int_{-\infty}^{\infty} \sin \omega t dF(\omega) = 0$ , since  $F$  is symmetric.

### 4.2.3 Spectrum for stationary sequences

If a stationary process  $\{x(t), t \in \mathbb{R}\}$ , with spectral distribution  $F(\omega)$  and covariance function  $r(t) = \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega)$ , is observed only at integer time points  $t \in \mathbb{Z}$  one obtains a stationary sequence  $\{x_n, n \in \mathbb{Z}\}$  for which the covariance function is the same as that of  $x(t)$  restricted to integer  $n = t$ . In the spectral formula the factor  $e^{i\omega t} = e^{i(\omega+2k\pi)t}$  for all integer  $t$  and  $k$ , and the spectrum may be restricted to the interval  $(-\pi, \pi]$ :

$$\begin{aligned} r(t) &= \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega) = \sum_{k=-\infty}^{\infty} \int_{2k\pi-\pi+0}^{2k\pi+\pi} e^{i\omega t} dF(\omega) \\ &= \int_{-\pi+0}^{\pi} e^{i\omega t} \sum_{k=-\infty}^{\infty} dF(\omega + 2k\pi). \end{aligned}$$

This means that all frequencies  $\omega + 2k\pi$  for  $k \neq 0$  are lumped together with the frequency  $\omega$  and cannot be individually distinguished. This is the *aliasing* or *folding* effect of sampling a continuous time process.

For a stationary sequence  $\{x_n, n \in \mathbb{Z}\}$ , the covariance function  $r(t)$  is defined only for  $t \in \mathbb{Z}$ . Instead of Bochner's theorem we have the following theorem, in the literature called *Herglotz' lemma*.

**Theorem 4:5** (*Herglotz' lemma*) *A function  $r(t), t \in \mathbb{Z}$ , defined on the integers, is non-negative definite, and hence a covariance function for a stationary sequence, if and only if there exists a non-decreasing, right-continuous, and bounded real function  $F(\omega)$  on  $(-\pi, \pi]$ , such that*

$$r(t) = \int_{-\pi+0}^{\pi} e^{i\omega t} dF(\omega). \quad (4.12)$$

Note that the spectrum is defined over the half-open interval to keep the right-continuity of  $F(\omega)$ . It is possible to move half the spectral mass in  $\pi$  to  $-\pi$  without changing the representation (4.12).

The inversion theorem states that if  $\sum_{t=-\infty}^{\infty} |r(t)| < \infty$  then the spectrum is absolutely continuous with spectral density given by

$$f(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-i\omega t} r(t),$$

while in general, for  $-\pi < \omega_1 < \omega_2 \leq \pi$ ,

$$\tilde{F}(\omega_2) - \tilde{F}(\omega_1) = \frac{1}{2\pi} r(0)(\omega_2 - \omega_1) + \lim_{T \rightarrow \infty} \frac{1}{2\pi} \sum_{\substack{t=-T, \\ t \neq 0}}^T r(t) \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{-it},$$

where as before  $\tilde{F}(\omega)$  is defined as the average of left and right hand side limits of  $F(\omega)$ .

## 4.3 Spectral representation of a stationary process

### 4.3.1 The spectral process

In elementary courses one could have encountered processes of the form

$$x(t) = \sum_k A_k \cos(\omega_k t + \phi_k), \quad (4.13)$$

where  $\omega_k > 0$  are fixed frequencies, while  $A_k$  are random amplitudes, and  $\phi_k$  random phases, uniformly distributed in  $(0, 2\pi)$  and independent of the  $A_k$ . The uniformly distributed phases make the process stationary, and its spectrum is discrete, concentrated at  $\{\omega_k\}$ . The covariance function and one-sided spectral distribution function are, respectively,

$$\begin{aligned} r(t) &= \sum_k E(A_k^2/2) \cos \omega_k t, \\ G(\omega) &= \sum_{k; \omega_k \leq \omega} E(A_k^2/2), \quad \omega > 0. \end{aligned}$$

The process (4.13) can also be defined as the real part of a complex process

$$x(t) = \Re \sum_k A_k e^{i\phi_k} e^{i\omega_k t},$$

and it is in fact a special example of the general spectral representation of a stationary process, which in takes the form of an integral

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega),$$

where  $\{Z(\omega); \omega \in \mathbb{R}\}$  is a complex spectral process with  $E(Z(\omega)) = 0$  and *orthogonal increments*, i.e.

$$E((Z(\omega_4) - Z(\omega_3)) \cdot \overline{(Z(\omega_2) - Z(\omega_1))}) = 0,$$

for  $\omega_1 < \omega_2 < \omega_3 < \omega_4$ . The variance of its increments is equal to the increments of the spectral distribution, i.e. for  $\omega_1 < \omega_2$ ,

$$E(|Z(\omega_2) - Z(\omega_1)|^2) = F(\omega_2) - F(\omega_1).$$

One can summarize the relations between  $Z(\omega)$  and  $F(\omega)$  as

$$E(dZ(\omega) \cdot \overline{dZ(\mu)}) = \begin{cases} dF(\omega) & \text{if } \omega = \mu, \\ 0 & \text{if } \omega \neq \mu. \end{cases} \quad (4.14)$$

It follows that  $Z(\omega)$  is continuous in quadratic mean if and only if the spectral distribution function  $F$  is continuous. If  $F$  has a jump at a point  $\omega_0$ ,

$$F(\omega_0+) - F(\omega_0-) = \sigma_0^2,$$

then  $\lim_{\epsilon \rightarrow 0} (Z(\omega_0 + \epsilon) - Z(\omega_0 - \epsilon))$  exists and has variance  $\sigma_0^2$ .

Now, let us start with a spectral process  $\{Z(\omega); \omega \in \mathbb{R}\}$ , a complex process with  $E(Z(\omega)) = 0$  and with orthogonal increments, and define the function  $F(\omega)$  by

$$F(\omega) = \begin{cases} E(|Z(\omega) - Z(0)|^2) & \text{for } \omega \geq 0, \\ -E(|Z(\omega) - Z(0)|^2) & \text{for } \omega < 0. \end{cases}$$

Since only the increments of  $Z(\omega)$  are used in the theory, we can fix its value at any point, and we take  $Z(0) = 0$ . Following the definition of a stochastic integral in Section 2.6, we can define a stochastic process

$$x(t) = \int e^{i\omega t} dZ(\omega) = \lim \sum e^{i\omega_k t} (Z(\omega_{k+1}) - Z(\omega_k)),$$

where the limit is in quadratic mean. It is then easy to prove that  $E(x(t)) = 0$  and that its covariance function is given by the Fourier-Stieltjes transform of  $F(\omega)$ : use Theorem 2:14, and (4.14), to get

$$\begin{aligned} E \left( \int e^{i\omega s} dZ(\omega) \cdot \overline{\int e^{i\mu t} dZ(\mu)} \right) &= \int \int e^{i(\omega s - \mu t)} E \left( dZ(\omega) \cdot \overline{dZ(\mu)} \right) \\ &= \int e^{i\omega(s-t)} dF(\omega). \end{aligned}$$

### 4.3.2 The spectral theorem

We shall now prove one of the central results in the theory, namely that every  $L^2$ -continuous weakly stationary process  $\{x(t), t \in \mathbb{R}\}$  has a spectral representation,  $x(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$ , where  $Z(\omega) \in \mathcal{H}(x)$ , i.e.  $Z(\omega)$  is an element in the Hilbert space which is spanned by limits of linear combinations of  $x(t)$ -values. In fact, one can define  $Z(\omega)$  explicitly for a continuity point  $\omega$  of  $F$ ,

$$Z(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-i\omega t} - 1}{-it} x(t) dt, \quad (4.15)$$

and prove that it has all the required properties. This is the technique used in Yaglom's classical book, [38]; see Exercise 5. We shall present a functional analytic proof, as in [9], and find a relation between  $\mathcal{H}(x) = \mathcal{S}(x(t); t \in \mathbb{R})$  and  $\mathcal{H}(F) = L_2(F) =$  the set of all functions  $g(\omega)$  with  $\int |g(\omega)|^2 dF(\omega) < \infty$ . We start by the definition of an *isometry*.

**Definition 4:1** *A linear one-to-one mapping  $f$  between two Hilbert spaces  $X$  and  $Y$  is called an isometry if it conserves the inner product  $(u, v)_X = (f(u), f(v))_Y$ . In particular  $\|u - v\|_X = \|f(u) - f(v)\|_Y$ , so distances are also preserved.*

**Theorem 4:6** *If  $\{x(t), t \in \mathbb{R}\}$  is a zero mean continuous stationary process with spectral distribution  $F(\omega)$  there exists a complex-valued spectral process  $\{Z(\omega), \omega \in \mathbb{R}\}$  with orthogonal increments, such that*

$$E\left(|Z(\omega_2) - Z(\omega_1)|^2\right) = F(\omega_2) - F(\omega_1),$$

for  $\omega_1 < \omega_2$  and

$$x(t) = \int e^{i\omega t} dZ(\omega).$$

**Proof:** We shall build an isometry between the Hilbert space of random variables  $\mathcal{H}(x) = \mathcal{S}(x(t); t \in \mathbb{R})$  and the function Hilbert space  $\mathcal{H}(F) = L_2(F)$ , with scalar products defined as

$$(u, v)_{\mathcal{H}(x)} = E(u\bar{v}),$$

$$(g, h)_{\mathcal{H}(F)} = \int g(\omega)\overline{h(\omega)} dF(\omega).$$

First consider the norms in the two spaces,

$$\|y\|_{\mathcal{H}(x)}^2 = E(|y|^2),$$

$$\|g\|_{\mathcal{H}(F)}^2 = \int |g(\omega)|^2 dF(\omega),$$

and note that

$$\|x(t)\|_{\mathcal{H}(x)}^2 = E(|x(t)|^2) = r(0),$$

$$\|e^{i\cdot t}\|_{\mathcal{H}(F)}^2 = \int |e^{i\omega t}|^2 dF(\omega) = r(0).$$

This just means that  $x(t)$  has the same length as an element of  $\mathcal{H}(x)$  as has  $e^{i\cdot t}$  as an element of  $\mathcal{H}(F)$ , i.e.  $\|x(t)\|_{\mathcal{H}(x)} = \|e^{i\cdot t}\|_{\mathcal{H}(F)}$ . Furthermore, scalar products are preserved,

$$(x(s), x(t))_{\mathcal{H}(x)} = E(x(s)\overline{x(t)}) = \int e^{i\omega s}\overline{e^{i\omega t}} dF(\omega) = (e^{i\cdot s}, e^{i\cdot t})_{\mathcal{H}(F)}.$$

This is the start of our isometry:  $x(t)$  and  $e^{i\cdot t}$  are the corresponding elements of the two spaces. Instead of looking for random variables  $Z(\omega_0)$  in  $\mathcal{H}(x)$  we shall look for functions  $g_{\omega_0}(\cdot)$  in  $\mathcal{H}(F)$  with the same properties.

**Step 1:** Extend the correspondence to finite linear combinations of  $x(t)$  and  $e^{i\omega t}$  by letting

$$y = \alpha_1 x(t_1) + \dots + \alpha_n x(t_n) \tag{4.16}$$

$$g(\omega) = \alpha_1 e^{i\omega t_1} + \dots + \alpha_n e^{i\omega t_n} \tag{4.17}$$

be corresponding elements. Check by yourself that scalar product is preserved, i.e.

$$(y_1, y_2)_{\mathcal{H}(x)} = (g_1, g_2)_{\mathcal{H}(F)}.$$

**Step 2:** Distances are preserved, i.e.  $\|y_1 - y_2\|_{\mathcal{H}(x)} = \|g_1 - g_2\|_{\mathcal{H}(F)}$ , so  $y_1 = y_2$  if and only if  $g_1 = g_2$  where equality means equal with probability one, and almost everywhere, respectively.

**Step 3:** Convergence in the two spaces means the same. If  $y_1, y_2, \dots$  converges towards  $y$  in  $\mathcal{H}(x)$ , and  $g_1, g_2, \dots$  are the corresponding elements in  $\mathcal{H}(F)$ , then

$$\|y_n - y_m\|_{\mathcal{H}(x)} \rightarrow 0 \quad \text{implies} \quad \|g_n - g_m\|_{\mathcal{H}(F)} \rightarrow 0,$$

and since  $\mathcal{H}(F)$  is complete there exists a limit element  $g \in \mathcal{H}(F)$  such that  $\|y\|_{\mathcal{H}(x)} = \|g\|_{\mathcal{H}(F)}$ . The reverse implication also holds. Thus we have extended the correspondence between the two spaces to all limits of sums of  $x(t)$ -variables and  $e^{i\omega t}$ -functions.<sup>2</sup>

**Step 4:** The correspondence can then be extended to all of  $\mathcal{H}(F)$  and  $\mathcal{H}(x)$ . The set  $\mathcal{H}(x)$  consists by definition of limits of linear combinations of  $x(t_k)$ , and every function in  $L_2(F)$  can be approximated by a polynomial in  $e^{i\omega t_k}$  for different  $t_k$ -s. This is the famous *Stone-Weierstrass theorem*. We have then found the isometry between  $\mathcal{H}(x)$  and  $\mathcal{H}(F)$ : if  $u$  and  $v$  are elements in  $\mathcal{H}(x)$  and  $f(u)$  and  $f(v)$  the corresponding elements in  $\mathcal{H}(F)$ , then  $\|u - v\|_{\mathcal{H}(x)} = \|f(u) - f(v)\|_{\mathcal{H}(F)}$ .

**Step 5:** The following function  $g_{\omega_0}$  in  $\mathcal{H}(F)$  corresponds to  $Z(\omega_0)$ ,

$$g_{\omega_0}(\omega) = \begin{cases} 1 & \text{for } \omega \leq \omega_0, \\ 0 & \text{for } \omega > \omega_0. \end{cases}$$

Obviously,  $\|g_{\omega_0}\|_{\mathcal{H}(F)}^2 = \int |g_{\omega_0}(\omega)|^2 dF(\omega) = \int_{-\infty}^{\omega_0} dF(\omega)$ , and, with  $\omega_1 < \omega_2$ ,

$$\|g_{\omega_2} - g_{\omega_1}\|_{\mathcal{H}(F)}^2 = F(\omega_2) - F(\omega_1).$$

**Step 6:** Let  $Z(\omega)$  be the elements in  $\mathcal{H}(x)$  that correspond to  $g_{\omega}(\cdot)$  in  $\mathcal{H}(F)$ . It is easy to see that  $Z(\omega)$  is a process with orthogonal increments and incremental variance given by  $F(\omega)$ :

$$\begin{aligned} E \left( (Z(\omega_4) - Z(\omega_3)) \cdot \overline{(Z(\omega_2) - Z(\omega_1))} \right) \\ = \int (g_{\omega_4}(\omega) - g_{\omega_3}(\omega))(g_{\omega_2}(\omega) - g_{\omega_1}(\omega)) dF(\omega) = 0, \end{aligned}$$

$$E(|Z(\omega_2) - Z(\omega_1)|^2) = F(\omega_2) - F(\omega_1),$$

---

<sup>2</sup>Remember that in  $\mathcal{H}(F)$  all  $e^{i\omega t}$  are functions of  $\omega$ , and that we have one function for every  $t$ . Similarly, in  $\mathcal{H}(x)$ ,  $x(t) = x(t, \omega)$  is a function of  $\omega$ .

for  $\omega_1 < \omega_2 < \omega_3 < \omega_4$ .

**Step 7:** It remains to prove that  $Z(\omega)$  is the spectral process to  $x(t)$ , i.e. that

$$\begin{aligned} x(t) &= \int e^{i\omega t} dZ(\omega) \\ &= \lim \sum e^{i\omega_k t} (Z(\omega_{k+1}) - Z(\omega_k)) = \lim S^{(n)}(t), \end{aligned}$$

for an increasingly dense subdivision  $\{\omega_k\}$  with  $\omega_k < \omega_{k+1}$ . But we have that  $x(t) \in \mathcal{H}(x)$  and  $e^{i\cdot t} \in \mathcal{H}(F)$  are corresponding elements. Further,

$$e^{i\omega t} = \lim \sum e^{i\omega_k t} (g_{\omega_{k+1}}(\omega) - g_{\omega_k}(\omega)) = \lim g_t^{(n)}(\omega),$$

where the difference of  $g$ -functions within parentheses is equal to 1 for  $\omega_k < \omega \leq \omega_{k+1}$  and 0 otherwise. The limits are in  $\mathcal{H}(F)$ , i.e.  $g(\cdot) = e^{i\cdot t} = \lim g_t^{(n)}(\cdot)$ . Since  $S^{(n)}(t)$  corresponds to  $g_t^{(n)}(\cdot)$  and limits are preserved under the isometry, we have that  $x(t) = \lim S^{(n)}(t)$ , as was to be shown.  $\square$

**Corollary 4.1** *Every  $y \in \mathcal{H}(x)$  can be written*

$$y = \int g(\omega) dZ(\omega),$$

for some function  $g(\omega) \in \mathcal{H}(F)$ .

**Proof:** Every  $y \in \mathcal{H}(x)$  is the limit of a sequence of linear combinations,

$$y = \lim_n \sum_k \alpha_k^{(n)} x(t_k^{(n)}) = \lim_n \int_\omega \sum_k \alpha_k^{(n)} e^{i\omega t_k^{(n)}} dZ(\omega),$$

and  $g^{(n)}(\cdot) = \sum \alpha_k^{(n)} e^{i\omega_k t_k^{(n)}}$  converges in  $\mathcal{H}(F)$  to some function  $g(\cdot)$ , and then

$$\int g^{(n)}(\omega) dZ(\omega) \rightarrow \int g(\omega) dZ(\omega)$$

in  $\mathcal{H}(x)$  which was to be shown.  $\square$

### 4.3.3 More on the spectral representation

#### 4.3.3.1 Discrete spectrum

If the spectral distribution function  $F(\omega)$  is piecewise constant with jumps of height  $\Delta F_k$  at  $\omega_k$ , then  $Z(\omega)$  is also piecewise constant with jumps of random size  $\Delta Z_k$  at  $\omega_k$ , and  $E(|\Delta Z_k|^2) = \Delta F_k$ , so  $x(t) = \sum \Delta Z_k e^{i\omega_k t}$ . Note that the covariance function then has the corresponding form,  $r(t) = \sum \Delta F_k e^{i\omega_k t}$ .

In the general spectral representation, the complex  $Z(\omega)$  defines a random amplitude and phase for the different components  $e^{i\omega t}$ . This fact is perhaps

difficult to appreciate in the integral form, but is easily understood for processes with discrete spectrum. Take the polar form,  $\Delta Z_k = |\Delta Z_k| e^{i \arg \Delta Z_k} = \rho_k e^{i \theta_k}$ . Then,

$$x(t) = \sum \rho_k e^{i(\omega_k t + \theta_k)} = \sum \rho_k \cos(\omega_k t + \theta_k) + i \sum \rho_k \sin(\omega_k t + \theta_k).$$

For a real process, the imaginary part vanishes, and we have the form, well known from elementary courses – see also later in this section –

$$x(t) = \sum \rho_k \cos(\omega_k t + \phi_k). \quad (4.18)$$

If the phases  $\phi_k$  are independent and uniformly distributed between 0 and  $2\pi$ , then  $x(t)$  is also strictly stationary.

For discrete spectrum we also have the following important ways of recovering the discrete components of  $F(\omega)$  and of  $Z(\omega)$ ; the proof of the properties are part of the Fourier theory.

**Theorem 4:7** *If  $F(\omega)$  is a step function, with jumps of size  $\Delta F_k$  at  $\omega_k$ , then*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T r(t) e^{-i\omega_k t} dt = \Delta F_k, \quad (4.19)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T r(t)^2 dt = \sum_k (\Delta F_k)^2, \quad (4.20)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) e^{-i\omega_k t} dt = \Delta Z_k. \quad (4.21)$$

#### 4.3.3.2 Continuous spectrum

If the spectrum is absolutely continuous, with  $F(\omega) = \int_{-\infty}^{\omega} f(x) dx$ , then one can normalize the increments of  $Z(\omega)$  by dividing by  $\sqrt{f(\omega)}$ , at least for  $f(\omega) > 0$ , and use the spectral representation in the form

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} \sqrt{f(\omega)} d\tilde{Z}(\omega), \quad (4.22)$$

with  $\tilde{Z}(\omega) = \int_{\{x \leq \omega; f(x) > 0\}} \frac{dZ(x)}{\sqrt{f(x)}}$ , and  $E \left( \left| d\tilde{Z}(\omega) \right|^2 \right) = \frac{dF(\omega)}{f(\omega)} = d\omega$ . Even if  $\tilde{Z}(\omega)$  is not a true spectral process – it may for example have infinite incremental variance – it is useful as a model for *white noise*. We will meet this “constant spectral density” formulation several times in later sections.

#### 4.3.3.3 One-sided spectral representation of a real process

For a real processes  $x(t)$ , the complex spectral representation has to produce a real integral. This of course requires  $Z(\omega)$  to have certain symmetry properties,

which we shall now investigate. Write  $\Delta Z_0$  for a possible  $Z$ -jump at  $\omega = 0$ . Then

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega) \\ &= \Delta Z_0 + \int_{0+}^{\infty} e^{i\omega t} dZ(\omega) + \int_{0+}^{\infty} e^{-i\omega t} dZ(-\omega) \\ &= \Delta Z_0 + \int_{0+}^{\infty} \cos \omega t \cdot (dZ(\omega) + dZ(-\omega)) \\ &\quad + i \int_{0+}^{\infty} \sin \omega t \cdot (dZ(\omega) - dZ(-\omega)). \end{aligned}$$

For this to be real for all  $t$  it is necessary that  $\Delta Z_0$  is real, and also that  $dZ(\omega) + dZ(-\omega)$  is real, and  $dZ(\omega) - dZ(-\omega)$  is purely imaginary, which implies  $dZ(-\omega) = \overline{dZ(\omega)}$ , i.e.  $\arg Z(-\omega) = -\arg Z(\omega)$  and  $|Z(-\omega)| = |Z(\omega)|$ . (These properties also imply that  $x(t)$  is real.)

Now, introduce two real processes  $\{u(\lambda), 0 \leq \lambda < \infty\}$  and  $\{v(\lambda), 0 \leq \lambda < \infty\}$ , with mean zero, and with  $u(0-) = v(0-) = 0$ ,  $du(0) = \Delta Z_0$ ,  $v(0+) = 0$ , and such that, for  $\omega > 0$ ,

$$\begin{aligned} du(\omega) &= dZ(\omega) + dZ(-\omega) = 2 \Re dZ(\omega) \\ dv(\omega) &= i(dZ(\omega) - dZ(-\omega)) = -2 \Im dZ(\omega). \end{aligned}$$

The real spectral representation of  $x(t)$  will then take the form

$$\begin{aligned} x(t) &= \int_0^{\infty} \cos \omega t du(\omega) + \int_0^{\infty} \sin \omega t dv(\omega) \\ &= \int_{0+}^{\infty} \cos \omega t du(\omega) + \int_0^{\infty} \sin \omega t dv(\omega) + du(0). \end{aligned} \quad (4.23)$$

It is easily checked that with the one-sided spectral distribution function  $G(\omega)$ , defined by (4.11),

$$E(du(\omega) \cdot dv(\mu)) = 0, \quad \text{for all } \omega \text{ and } \mu, \quad (4.24)$$

$$E(du(\omega)^2) = \begin{cases} 2dF(\omega) = dG(\omega), & \omega > 0, \\ dF(0) = dG(0), & \omega = 0, \end{cases} \quad (4.25)$$

$$E(dv(\omega)^2) = \begin{cases} 2dF(\omega) = dG(\omega), & \omega > 0, \\ dF(0) = dG(0), & \omega = 0. \end{cases} \quad (4.26)$$

In almost all applications, when a spectral density for a time process  $x(t)$  is presented, it is the one-sided density  $g(\omega) = 2f(\omega) = dG(\omega)/d\omega$  that is given.

#### 4.3.3.4 Why negative frequencies?

One may ask why at all use negative frequencies  $\omega < 0$  in the spectral representation of a real process. Since the two complex functions  $dZ(\omega) e^{i\omega t}$  and  $dZ(-\omega) e^{-i\omega t} = \overline{dZ(\omega) e^{i\omega t}}$ , which build up the spectral representation, circle the origin in the counter clockwise and clockwise directions their contribution to the total  $x(t)$ -process is real, and there seems to be no point in using the complex formulation.

One reason for the complex approach is, besides from some mathematical convenience, that negative frequencies are necessary when we want to build models for simultaneous *time and space* processes, for example a random water wave which moves with time  $t$  along a straight line with coordinate  $s$ . As described in Section 1.6.3 on page 22, a random wave model can be built from elementary harmonics  $A_\omega \cos(\omega t - \kappa s + \phi_\omega)$  where  $\omega$  is the frequency in radians per time unit and  $\kappa$  is the wave number in radians per length unit. If  $\omega$  and  $\kappa$  have the same sign the elementary wave moves to the right with increasing  $t$  and if they have opposite sign it moves to the left. In stochastic wave models for infinite water depth the dispersion relation states that  $\kappa = \omega^2/g > 0$ , with both positive and negative  $\omega$  possible. The (average) “energy” attached to the elementary wave  $A_\omega \cos(\omega t - \kappa s + \phi_\omega)$  is  $A_\omega^2/2$  or, in the random case  $E(A_\omega^2)/2$ .

If one observes the wave only at a single point  $s = s_0$  it is not possible to determine in which direction the wave is moving, and one can divide the elementary energy in an arbitrary way between  $\omega$  and  $-\omega$ . When we deal with the spectral density for the time process we have chosen to divide it equally between positive and negative frequencies.

If we have more than one observation point, perhaps a whole space interval of observations, we can determine wave direction and see how the energy should be divided between positive and negative  $\omega$ . This can be done by splitting the process in two independent components, one  $\{x_+(t), t \in \mathbb{R}\}$  with only positive frequencies, moving to the right, and one  $\{x_-(t), t \in \mathbb{R}\}$  with only negative frequencies, moving to the left. The spectra on the positive and negative side need not be equal.

For a wave model with one time parameter and a two-dimensional space parameter  $(s_1, s_2)$  the wave direction is taken care of by a two-dimensional wave number and the spectrum defined by one part that defines the energy distribution over frequencies, and one directional spreading part that determines the energy for different wave directions.

#### 4.3.3.5 Gaussian processes

For Gaussian processes  $\{x(t), t \in \mathbb{R}\}$ , the spectral process is complex Gaussian, with independent real and imaginary parts. Since  $Z(\omega)$  is an element in the space  $\mathcal{H}(x)$  of limits of linear combinations of  $x$ -variables, this is immediate from the characterization of Gaussian processes as those processes for which all linear combinations have a Gaussian distribution. Also the real spectral

processes  $\{u(\lambda), 0 \leq \lambda < \infty\}$  and  $\{v(\lambda), 0 \leq \lambda < \infty\}$  are Gaussian, and since they have uncorrelated increments, they are Gaussian processes with independent increments.

The sample paths of  $u(\omega)$  and  $v(\omega)$  can be continuous, or they could contain jump discontinuities, which then are normal random variables. In the continuous case, when there is a spectral density  $f(\omega)$ , they are almost like Wiener processes, and they can be transformed into Wiener processes by normalizing the incremental variance. In analogy with  $\tilde{Z}(\omega)$  in (4.22), define  $w_1(\omega)$  and  $w_2(\omega)$  by

$$w_1(\omega) = \int_{\{x \leq \omega; f(x) > 0\}} \frac{du(x)}{\sqrt{2f(x)}}, \quad w_2(\omega) = \int_{\{x \leq \omega; f(x) > 0\}} \frac{dv(x)}{\sqrt{2f(x)}}, \quad (4.27)$$

to get, Theorem 2:16,

$$x(t) = \int_0^\infty \sqrt{2f(\omega)} \cos \omega t dw_1(\omega) + \int_0^\infty \sqrt{2f(\omega)} \sin \omega t dw_2(\omega). \quad (4.28)$$

Note that if  $f(\omega) > 0$  then  $E(dw_1(\omega)^2) = E(dw_2(\omega)^2) = d\omega$ .

The representation (4.28) is particularly useful for simulation of stationary Gaussian processes, as described in detail in Appendix D. Then the continuous spectrum is discretized to frequencies  $\omega_k = k\Delta$ ,  $k \in \mathbb{N}$ , and the integrals (4.28) replaced by sums. Since the increments in the Wiener processes are independent normal variables, the approximative expressions become

$$x(t) = \sum U_k \sqrt{2\Delta F(\omega_k)} \cos \omega_k t + \sum V_k \sqrt{2\Delta F(\omega_k)} \sin \omega_k t, \quad (4.29)$$

$$= \sum A_k \sqrt{2\Delta F(\omega_k)} \cos(\omega_k t + \phi_k), \quad (4.30)$$

where  $U_k$  and  $V_k$  are independent standard normal variables, and

$$A_k = \sqrt{U_k^2 + V_k^2}, \quad \phi_k = -\arg(U_k + iV_k).$$

Historically, the representation (4.29) was used explicitly already by Lord Rayleigh in connection with heat radiation and by Einstein (1910) and others to introduce Gaussian randomness. The form (4.30) appears to have come later, at least according to S.O. Rice, who cites work written by W.R. Bennett in the 1930's.

#### 4.3.3.6 Gaussian white noise

The differentials  $dw_1(\omega)$  and  $dw_2(\omega)$  in (4.28) are examples of *Gaussian white noise*. White noise in general is a common notion in stochastic process theory when one needs a process in continuous time where all process values are virtually independent, regardless of how close they are in time. Complete independence would require  $r_x(t) = 0$  for all  $t$  except  $t = 0$ , i.e. the covariance

function is not continuous and Bochner's theorem, Theorem 4:3, is of no use to find a corresponding spectrum. Fourier's inversion formula (4.10) hints that the spectrum should be independent of  $\omega$  but  $f(\omega) > 0$  is not a spectral density. On the other hand, the  $\delta$ -distribution,  $\delta(\omega)$ , also called the *Dirac delta function*, forms a Fourier transform pair together with the constant function  $f(\omega) = 1/2\pi$ . It is in fact possible to formulate a theory for "distribution valued" stationary processes and covariance functions, but that theory is little used in practical work and we do not go into any details on this; for a brief introduction, see [38, Appendix I].

Instead we will use the two Wiener processes defined by (4.27) to illustrate the common way to go around the problem with constant spectral density. We used them as spectral processes in (4.28) without any difficulty; we only noted that  $E(dw_1(\omega)^2) = E(dw_2(\omega)^2) = d\omega$ .

In the theory of stochastic differential equations, one often uses the notation  $w'(t)$  or  $dw(t)$  with the understanding that it is shorthand for a stochastic integral of the form  $\int_{t_0}^t g(u) dw(u)$ , for  $\int g(t)^2 dt < \infty$ . We will illustrate this with the previously mentioned Langevin equation, (1.15), and deal with these more in detail in Section 4.4.4.

**Example 4:3** (The Ornstein-Uhlenbeck process) The Ornstein-Uhlenbeck process is a Gaussian stationary process with covariance function  $r(t) = \sigma^2 e^{-\alpha|t|}$ , and spectral density

$$f(\omega) = \frac{\sigma^2}{\pi} \cdot \frac{\alpha}{\alpha^2 + \omega^2}.$$

We saw in Chapter 2 that a Gaussian process with this covariance function and spectrum is continuous but not differentiable.

The process can be realized as a stochastic integral

$$x(t) = \sqrt{2\alpha\sigma^2} \int_{-\infty}^t e^{-\alpha(t-\tau)} dw(\tau). \quad (4.31)$$

As we will see in Section 4.4.4 the Ornstein-Uhlenbeck process is the solution of the linear stochastic differential equation

$$\alpha x(t) + x'(t) = \sqrt{2\alpha}\sigma w'(t) \quad (4.32)$$

with Gaussian white noise  $w'(t)$ . We met this equation in Section 1.6.1 under the name Langevin's equation.

For large  $\alpha$  ( $\alpha \rightarrow \infty$ ), the covariance function falls off very rapidly around  $t = 0$  and the correlation between  $x(s)$  and  $x(t)$  becomes negligible when  $s \neq t$ . In the integral (4.31) each  $x(t)$  depends asymptotically only on the increment  $dw(t)$  and are hence approximately independent. With increasing  $\alpha$ , the spectral density becomes increasingly flatter at the same time as  $f(\omega) \rightarrow 0$ . In order to keep the variance of the process constant, not going to 0 or  $\infty$ , we

can let  $\sigma^2 \rightarrow \infty$  in such a way that  $\sigma^2/\alpha \rightarrow C > 0$ . Therefore, the Ornstein-Uhlenbeck process with large  $\alpha$  and  $\sigma^2/\alpha = C$  can be used as an approximation to Gaussian white noise.

As a stationary process the Ornstein-Uhlenbeck has a spectral representation of the type (4.28), and one may ask what connection there is between the two integral representations.

To see the analogy, take  $w_1(\omega)$  and  $w_2(\omega)$  from (4.28) and define  $w_C(\omega)$  for  $-\infty < \omega < \infty$ , as

$$w_C(\omega) = \begin{cases} w_1(\omega) + iw_2(\omega), & \omega > 0, \\ w_1(-\omega) - iw_2(-\omega), & \omega < 0, \end{cases}$$

to get

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} \sqrt{f(\omega)} dw_C(\omega). \quad (4.33)$$

We then do some formal calculation with white noise:  $w'(t)$  is the formal derivative of the Wiener process, and it is a stationary process with *constant spectral density* equal to  $1/2\pi$  over the whole real line, i.e. by (4.33),

$$w'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} dw_C(\omega),$$

for some complex Wiener process  $w_C(\omega)$ . Inserting this in (4.31), we obtain

$$\begin{aligned} x(t) &= \sqrt{2\alpha\sigma^2} \int_{-\infty}^t e^{-\alpha(t-\tau)} w'(\tau) d\tau \\ &= \frac{\sqrt{2\alpha\sigma^2}}{\sqrt{2\pi}} \int_{\tau=-\infty}^t e^{-\alpha(t-\tau)} \left\{ \int_{\omega=-\infty}^{\infty} e^{i\omega\tau} dw_C(\omega) \right\} d\tau \\ &= \frac{\sqrt{2\alpha\sigma^2}}{\sqrt{2\pi}} \int_{\omega=-\infty}^{\infty} \left\{ \int_{\tau=-\infty}^t e^{-(\alpha+i\omega)(t-\tau)} d\tau \right\} e^{i\omega t} dw_C(\omega) \\ &= \frac{\sqrt{2\alpha\sigma^2}}{\sqrt{2\pi}} \int_{\omega=-\infty}^{\infty} \frac{1}{\alpha+i\omega} e^{i\omega t} dw_C(\omega) \\ &= \frac{\sqrt{2\alpha\sigma^2}}{\sqrt{2\pi}} \int_{\omega=-\infty}^{\infty} \frac{1}{\sqrt{\alpha^2+\omega^2}} e^{i(-\arg(\alpha+i\omega)+\omega t)} dw_C(\omega) \\ &= \int_{\omega=-\infty}^{\infty} e^{i(\omega t+\gamma(\omega))} \sqrt{f(\omega)} dw_C(\omega), \end{aligned}$$

with  $\gamma(-\omega) = -\gamma(\omega)$ . The same Wiener process  $w_C(\omega)$  which works in the spectral representation of the white noise in (4.31) can be used as spectral process in (4.33) after correction of the phase.

### 4.3.4 Spectral representation of stationary sequences

A stationary sequence  $\{x(t), t \in \mathbb{Z}\}$  can be thought of as a stationary process which is observed only at integer times  $t$ . The spectral representation can then be restricted to  $\omega$ -values only in  $(-\pi, \pi]$  as for the spectral distribution. In the formula

$$x(t) = \int_{-\pi+}^{\pi} e^{i\omega t} dZ(\omega), \quad (4.34)$$

there is now an explicit expression for the spectral process,

$$Z(\omega) = \frac{1}{2\pi} \left\{ \omega x(0) - \sum_{k \neq 0} \frac{e^{-i\omega k}}{ik} x(k) \right\}$$

## 4.4 Linear filters

### 4.4.1 Projection and the linear prediction problem

One of the most useful instruments in the theory of stochastic processes is the linear prediction device, by which we "predict" or approximate a random variable  $x$  by a linear combination of a set of observed random variables, or by a limit of such linear combinations. The general formulation in the theory of stochastic processes is the *linear filtering problem* in which one seeks a linear filter  $h(u)$  such that the linearly filtered process

$$\hat{y}(t) = \int_{u=-\infty}^{\infty} h(u)x(t-u) du$$

approximates some interesting random quantity  $y(t)$ , dependent on  $x(s)$ ,  $s \in \mathbb{R}$ . If the *impulse response function*  $h(u)$  is zero for  $u < 0$  we talk about *linear prediction*, otherwise we call it *linear reconstruction*. The impulse response may contain  $\delta$ -functions  $\delta_{\tau_k}$ , which act as time delays; for example  $y(t) = \int \delta_{\tau_0}(u)x(t-u) du = x(t - \tau_0)$ .

The projection theorem in Hilbert spaces states that if  $\mathcal{M}$  is a closed linear subspace of a Hilbert space  $\mathcal{H}$ , and  $x$  is a point in  $\mathcal{H}$  not in  $\mathcal{M}$ , then there is a unique element  $y$  in  $\mathcal{M}$  closest to  $x$ , and then  $z = x - y$  is orthogonal to  $\mathcal{M}$ ; see Appendix C.

Formulated in statistical terms, if  $x$  is a random variable and  $y_1, \dots, y_n$  is a finite set of random variables, then there is a unique linear combination  $\hat{x} = c_1 y_1 + \dots + c_n y_n$  that is closest to  $x$  in the  $\|\cdot\|$ -norm, i.e. such that

$$\left\| x - \sum c_j y_j \right\|^2 = E(|x - \sum c_j y_j|^2)$$

is minimal. This linear combination is characterized by the requirement that the residual  $x - \sum c_j y_j$  is orthogonal, i.e. uncorrelated with all the  $y_j$ -variables. This is the least squares solution to the common *linear regression problem*.

Expressed in terms of covariances, the coefficients in the optimal predictor  $\hat{x} = c_1 y_1 + \dots + c_n y_n$  satisfy the linear equation system

$$\text{Cov}(x, y_j) = c_1 \text{Cov}(y_1, y_j) + \dots + c_n \text{Cov}(y_n, y_j), \quad j = 1, \dots, n, \quad (4.35)$$

which follows from  $\text{Cov}(x - \sum_k c_k y_k, y_j) = 0$ .

Note that the projection theorem says that the random variable  $y = \hat{x}$  is unique, in the sense that if  $\tilde{y}$  is another random variable that minimizes the prediction error, i.e.  $E(|x - y|^2) = E(|x - \tilde{y}|^2)$  then  $E(|y - \tilde{y}|^2) = 0$  and  $P(y = \tilde{y}) = 1$ . This does not mean that the coefficients in the linear combination  $\sum c_j y_j$  are unique; if the variables  $y_1, \dots, y_n$  are linearly dependent then many combinations produce the same best predictor.

**Example 4:4** For the MA(1)-process,  $x(t) = e(t) + b_1 e(t - 1)$ , we found in Example C:1 that

$$e(t) = \begin{cases} \sum_{k=0}^{\infty} (-b_1)^k x(t - k), & \text{if } |b_1| < 1, \\ \lim_{n \rightarrow \infty} \sum_{k=0}^n (1 - \frac{k}{n}) x(t - k), & \text{for } b_1 = -1. \end{cases}$$

Thus,  $x(t + 1) = e(t + 1) + b_1 e(t)$  has been written as the sum of one variable  $e(t + 1) \perp \mathcal{H}(x, t)$  and one variable  $b_1 e(t) \in \mathcal{H}(x, t)$ . The projection theorem implies that the best linear prediction of  $x(t + 1)$  based on  $x(s), s \leq t$ , is

$$\hat{x}_t(t + 1) = b_1 e(t) = \begin{cases} b_1 \sum_{k=0}^{\infty} (-b_1)^k x(t - k), & \text{if } |b_1| < 1, \\ \lim_{n \rightarrow \infty} - \sum_{k=0}^n (1 - \frac{k}{n}) x(t - k), & \text{for } b_1 = -1. \end{cases}$$

Note that  $\mathcal{H}(e, t) = \mathcal{H}(x, t)$ .

**Example 4:5** We can extend the previous example to have  $\mathcal{H}(x, t) \subset \mathcal{H}(e, t)$  with strict inclusion. Take a series of variables  $e^*(t)$  and a random variable  $U$  with  $E(U) = 0$  and  $V(U) < \infty$ , everything uncorrelated, and set

$$e(t) = U + e^*(t).$$

Then  $x(t) = e(t) - e(t - 1) = e^*(t) - e^*(t - 1)$ , and  $\mathcal{H}(e, t) = \mathcal{H}(U) \oplus \mathcal{H}(e^*, t)$  with  $\mathcal{H}(U)$  and  $\mathcal{H}(e^*, t)$  orthogonal, and  $\mathcal{H}(e, t) \supseteq \mathcal{H}(e^*, t) = \mathcal{H}(x, t)$ .

## 4.4.2 Linear filters and the spectral representation

### 4.4.2.1 Frequency response

In the previous section we formulated the prediction solution as a linear filter with an impulse response functions. Here we take slightly more abstract approach and use a frequency formulation. A *linear time-invariant filter* is a

transformation  $\mathcal{S}$  that takes a stationary process  $x(t) = \int e^{i\omega t} dZ(\omega)$  into a new stationary process  $y(t)$  so that,

$$y(t) = \int_{-\infty}^{\infty} g(\omega) e^{i\omega t} dZ(\omega), \quad (4.36)$$

where  $g(\omega)$  is the *transfer function* (also called *frequency function*). It has to satisfy  $\int |g(\omega)|^2 dF(\omega) < \infty$ . That the filter is linear and time-invariant means that, for any (complex) constants  $a_1, a_2$  and time delay  $\tau$ ,

$$\begin{aligned} \mathcal{S}(a_1 x_1 + a_2 x_2) &= a_1 \mathcal{S}(x_1) + a_2 \mathcal{S}(x_2), \\ \mathcal{S}(x(\cdot + \tau)) &= \mathcal{S}(x)(\cdot + \tau). \end{aligned}$$

As an alternative to the impulse function approach in Section 4.4.1 we may take (4.36) as the *definition* of a linear time-invariant filter.

The process  $y(t)$  is also stationary, and it has covariance function given by

$$E\left(y(s+t)\overline{y(s)}\right) = \int_{-\infty}^{\infty} |g(\omega)|^2 e^{i\omega t} dF(\omega).$$

In particular, if  $x(t)$  has spectral density  $f_x(\omega)$  then the spectral density of  $y(t)$  is

$$f_y(\omega) = |g(\omega)|^2 f_x(\omega). \quad (4.37)$$

Many of the interesting processes we have studied in previous sections, were obtained as linear combinations of  $x(t)$ -variables, or, more commonly, as limits of linear combinations. To formulate the spectral forms of these operations, we need the following property, cf. Step 3, in the proof of Theorem 4.6.

**Lemma 4.1** *If  $g_n \rightarrow g$  in  $\mathcal{H}(F)$ , i.e.  $\int |g_n(\omega) - g(\omega)|^2 dF(\omega) \rightarrow 0$ , then*

$$\int g_n(\omega) e^{i\omega t} dZ(\omega) \rightarrow \int g(\omega) e^{i\omega t} dZ(\omega)$$

in  $\mathcal{H}(x)$ .

**Proof:** Use the isometry,

$$\begin{aligned} & \left\| \int g_n(\omega) e^{i\omega t} dZ(\omega) - \int g(\omega) e^{i\omega t} dZ(\omega) \right\|_{\mathcal{H}(x)}^2 \\ &= \int |g_n(\omega) e^{i\omega t} - g(\omega) e^{i\omega t}|^2 dF(\omega) = \|g_n - g\|_{\mathcal{H}(F)}^2. \end{aligned}$$

□

**Example 4:6** For example, the linear operation "derivation" of a stationary process is the limit of

$$\frac{x(t+h) - x(t)}{h} = \int_{-\infty}^{\infty} \frac{e^{ih\omega} - 1}{h} e^{i\omega t} dZ(\omega).$$

If  $x(t)$  satisfies the condition  $\int \omega^2 dF(\omega) < \infty$  for quadratic mean differentiability,  $(e^{i\omega h} - 1)/h \rightarrow i\omega$  in  $\mathcal{H}(F)$  as  $h \rightarrow 0$ , and hence

$$x'(t) = \int i\omega e^{i\omega t} dZ(\omega) = \int \omega e^{i(\omega t + \pi/2)} dZ(\omega).$$

The frequency function for derivation is therefore  $g(\omega) = i\omega$ , and the spectral density of the derivative is  $f_{x'}(\omega) = \omega^2 f_x(\omega)$ .

In general, writing  $y(t) = \int |g(\omega)| e^{i(\omega t + \arg g(\omega))} dZ(\omega)$ , we see how the filter amplifies the amplitude of  $dZ(\omega)$  by a factor  $|g(\omega)|$  and adds  $\arg g(\omega)$  to the phase. For the derivative, the phase increases by  $\pi/2$ , while the amplitude increases by a frequency dependent factor  $\omega$ .

#### 4.4.2.2 A practical rule

The spectral formulation of linear filters gives us an easy-to-use tool to find covariance and cross-covariance functions between stationary processes. If the process  $\{x(t), t \in \mathbb{R}\}$  is stationary with spectral distribution function  $F_x(\omega)$ , and  $\{u(t), t \in \mathbb{R}\}$  and  $\{v(t), t \in \mathbb{R}\}$  are generated from  $\{x(t), t \in \mathbb{R}\}$  by linear filters,

$$\begin{aligned} x(t) &= \int e^{i\omega t} dZ(\omega), \\ u(t) &= \int g(\omega) e^{i\omega t} dZ(\omega), \\ v(t) &= \int h(\omega) e^{i\omega t} dZ(\omega), \end{aligned}$$

then, by (4.14),

$$\begin{aligned} Cov(x(s), u(t)) &= \int_{\omega} \int_{\mu} e^{i\omega s} e^{-i\mu t} \overline{g(\mu)} E(dZ(\omega) \cdot \overline{dZ(\mu)}) \\ &= \int e^{i(s-t)\omega} \overline{g(\omega)} dF_x(\omega), \end{aligned}$$

and similarly,

$$Cov(u(s), v(t)) = \int g(\omega) \overline{h(\omega)} e^{i(s-t)\omega} dF_x(\omega). \quad (4.38)$$

### 4.4.2.3 Impulse response and frequency response

Suppose a linear filter is defined by its *impulse response function*  $h(t)$ , as in Section 4.4.1,

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u) du = \int_{-\infty}^{\infty} h(t-u)x(u) du.$$

Inserting the spectral representation of  $x(t)$  and changing the order of integration, we obtain a filter in frequency response form,

$$y(t) = \int_{\omega=-\infty}^{\infty} \left\{ \int_{u=-\infty}^{\infty} e^{i\omega u} h(t-u) du \right\} dZ(\omega) = \int_{-\infty}^{\infty} g(\omega) e^{i\omega t} dZ(\omega),$$

with

$$g(\omega) = \int_{u=-\infty}^{\infty} e^{-i\omega u} h(u) du, \quad (4.39)$$

if  $\int |h(u)| du < \infty$ .

Conversely, if  $h(u)$  is absolutely integrable,  $\int |h(u)| du < \infty$ , then  $g(\omega)$ , defined by (4.39), is bounded and hence  $\int |g(\omega)|^2 dF(\omega) < \infty$ . Therefore

$$y(t) = \int g(\omega) e^{i\omega t} dZ(\omega)$$

defines a linear filter with frequency function  $g(\omega)$  as in (4.36). Inserting the expression for  $g(\omega)$  and changing the order of integration we get the impulse response form,

$$\begin{aligned} y(t) &= \int e^{i\omega t} \left\{ \int e^{-i\omega u} h(u) du \right\} dZ(\omega) \\ &= \int h(u) \left\{ \int e^{i\omega(t-u)} dZ(\omega) \right\} du = \int h(u)x(t-u) du. \end{aligned}$$

The impulse response and frequency response function form a Fourier transform pair, and

$$h(u) = \frac{1}{2\pi} \int_{\omega=-\infty}^{\infty} e^{i\omega u} g(\omega) d\omega. \quad (4.40)$$

If  $h(u) = 0$  for  $u < 0$  the filter is called *causal* or *physically realizable*, indicating that then  $y(t) = \int_{u=0}^{\infty} h(u)x(t-u) du$  depends only on  $x(s)$  for  $s \leq t$ , i.e. the output from the filter at time  $t$  depends on the past and not on the future.

### 4.4.2.4 Linear processes

A stationary sequence  $x_t$  or a stationary process  $x(t)$  is called *linear* if it is the output of a linear time invariant filter acting on a sequence of orthogonal

random variables, i.e.

$$x_t = \sum_{k=-\infty}^{\infty} h_{t-k} y_k, \quad (4.41)$$

$$x(t) = \int_{u=-\infty}^{\infty} h(t-u) dY(u), \quad (4.42)$$

where  $y_k$  are uncorrelated with mean 0, and  $E(|y_k|^2) = 1$ , and  $\{Y(t), t \in \mathbb{R}\}$  is a stationary process with orthogonal increments,  $E(dY(u) \cdot dY(v))$  is equal to 0 for  $u \neq v$  and equal to  $du$  for  $u = v$ . The term *infinite moving average* is also used for processes of this type.

**Theorem 4:8** a) A stationary sequence  $\{x_t; t \in \mathbb{Z}\}$  is an infinite moving average

$$x_t = \sum_{k=-\infty}^{\infty} h_{t-k} y_k,$$

with orthonormal  $y_k$  and  $\sum_k |h_k|^2 < \infty$ , if and only if its spectrum is absolutely continuous,  $F(\omega) = \int_{-\pi}^{\omega} f(x) dx$ .

b) A stationary process  $\{x(t), t \in \mathbb{R}\}$  is an infinite continuous moving average  $x(t) = \int_{u=-\infty}^{\infty} h(t-u) dY(u)$ , with an orthogonal increment process  $Y(\omega)$ , and  $\int_u |h(u)|^2 du < \infty$ , if and only if its spectrum is absolutely continuous,  $F(\omega) = \int_{-\infty}^{\omega} f(x) dx$ .

**Proof:** We show part a); part b) is quite similar. For the "only if" part, use that  $y_k = \int_{-\pi}^{\pi} e^{i\omega k} dZ(\omega)$ , where  $E(|dZ(\omega)|^2) = \frac{d\omega}{2\pi}$ . Then

$$\begin{aligned} x_t &= \sum_k h_{t-k} \int_{-\pi}^{\pi} e^{i\omega k} dZ(\omega) = \int_{-\pi}^{\pi} e^{i\omega t} \left\{ \sum_k h_{t-k} e^{-i\omega(t-k)} \right\} dZ(\omega) \\ &= \int_{-\pi}^{\pi} e^{i\omega t} g(\omega) dZ(\omega), \end{aligned}$$

with  $g(\omega) = \sum_k h_k e^{-i\omega k}$ . Thus, the spectral distribution of  $x_k$  has

$$dF(\omega) = E(|g(\omega) dZ(\omega)|^2) = |g(\omega)|^2 \frac{d\omega}{2\pi},$$

with spectral density  $f(\omega) = \frac{1}{2\pi} |g(\omega)|^2$ .

For the "if" part,  $F(\omega) = \int_{-\infty}^{\omega} f(x) dx$ , write  $f(\omega) = \frac{1}{2\pi} |g(\omega)|^2$ , and expand  $|g(\omega)|$  in a Fourier series,

$$|g(\omega)| = \sum_k c_k e^{i\omega k}.$$

From the normalized spectral representation (4.22),

$$x_t = \int_{-\pi}^{\pi} e^{i\omega t} \sqrt{f(\omega)} d\tilde{Z}(\omega), \quad \text{with} \quad E(|d\tilde{Z}(\omega)|^2) = d\omega,$$

we then get

$$\begin{aligned} x_t &= \int_{-\pi}^{\pi} e^{i\omega t} \left\{ \frac{1}{\sqrt{2\pi}} \sum_k c_k e^{i\omega k} \right\} d\tilde{Z}(\omega) \\ &= \sum_k \frac{c_k}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\omega(t+k)} d\tilde{Z}(\omega) = \sum_k c_k e_{t+k} = \sum_k h_{t-k} e_k, \end{aligned}$$

with  $e_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\omega k} d\tilde{Z}(\omega)$ , and  $h_k = c_{k+t}$ . Since the  $\tilde{Z}(\omega)$  has constant incremental variance, the  $e_k$ -variables are uncorrelated and normalized as required.  $\square$

### 4.4.3 Linear filters and differential equations

Linear filters expressed in terms of differential equations are common in the engineering sciences. The linear oscillator, also called the harmonic oscillator, is the basic element in mechanical systems which exhibit resonant periodic movements. Its counterpart in electronic systems is the resonance circuit. We shall describe both of these as examples of a general technique, common in the theory of ordinary differential equations.

To illustrate the general ideas we start with the exponential smoothing filter, also called the RC-filter, with a term borrowed from electrical engineering.

#### 4.4.3.1 The RC-filter and exponential smoothing

Consider the electrical circuit in Figure 4.1 with potential difference  $x(t)$  on the left hand side and potential difference  $y(t)$  on the right hand side. The circuit consists of a resistance  $R$  and a capacitance  $C$ . Regarding  $x(t)$  as the driving process and  $y(t)$  as the resulting process we will see that this device acts as a *smoother* that reduces rapid high frequency variations in  $x(t)$ . The relation between the input  $x(t)$  and the output  $y(t)$  is

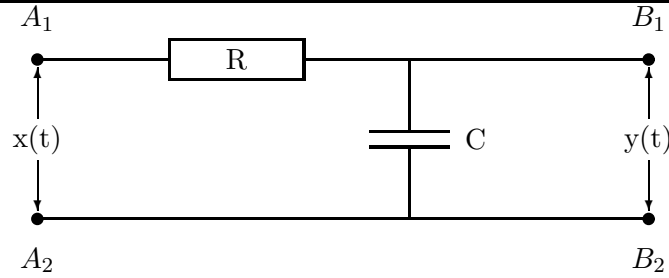
$$RCy'(t) + y(t) = x(t), \quad (4.43)$$

and the equations has the (deterministic) solution

$$y(t) = \frac{1}{RC} \int_{-\infty}^t e^{-(t-u)/(RC)} x(u) du.$$

Thus, the impulse response of the RC-filter is

$$h(u) = \frac{1}{RC} e^{-u/(RC)}, \quad \text{for } u > 0,$$



**Figure 4.1:** Input  $x(t)$  and output  $y(t)$  in an exponential smoother (RC-filter).

with frequency response

$$g(\omega) = \int_0^{\infty} e^{-i\omega u} \frac{1}{RC} e^{-u/(RC)} du = \frac{1}{1 + i\omega RC}.$$

Applying the relation (4.37) we get the spectral density relation between input and output,

$$f_y(\omega) = \frac{f_x(\omega)}{(\omega RC)^2 + 1}. \quad (4.44)$$

The depreciation of high frequencies in the spectrum explains the use of the RC-filter as a smoother.

To precede the general results for covariance function relations we also make the following elementary observation about the covariance functions, where we use the cross-covariances from Theorem 2:13:

$$\begin{aligned} r_x(\tau) &= \text{Cov}(RCy'(t) + y(t), RCy'(t + \tau) + y(t + \tau)) \\ &= (RC)^2 r_{y'}(\tau) + RC r_{y, y'}(t, t + \tau) + RC r_{y', y}(t, t + \tau) + r_y(\tau) \\ &= (RC)^2 r_{y'}(\tau) + RC r'_y(\tau) + RC r'_y(-\tau) + r_y(\tau) \\ &= (RC)^2 r_{y'}(\tau) + r_y(\tau). \end{aligned}$$

Using the spectral density  $\omega^2 f_y(\omega)$  for  $\{y'(t), t \in \mathbb{R}\}$ , according to Example 4:6, we find

$$\begin{aligned} r_x(t) &= (RC)^2 \int e^{i\omega t} \omega^2 f_y(\omega) d\omega + \int e^{i\omega t} f_y(\omega) d\omega \\ &= \int e^{i\omega t} \{(\omega RC)^2 + 1\} f_y(\omega) d\omega, \end{aligned}$$

and get the spectral density for  $\{x(t), t \in \mathbb{R}\}$ ,

$$f_x(\omega) = \{(\omega RC)^2 + 1\} f_y(\omega), \quad (4.45)$$

in accordance with (4.44).

As a final observation we note that the impulse response function satisfies the differential equation

$$RCh'(u) + h(u) = 0, \quad (4.46)$$

for  $u > 0$  with the initial condition  $h(0) = 1/(RC)$ .

### 4.4.3.2 Linear stochastic differential equations

Suppose we have a stationary process  $\{x(t), t \in \mathbb{R}\}$ , sufficiently differentiable, and assume that the process  $\{y(t), t \in \mathbb{R}\}$  is a solution to an ordinary linear differential equation with constant coefficients,

$$\sum_{k=0}^p a_{p-k} y^{(k)}(t) = x(t). \quad (4.47)$$

or, seemingly more generally,

$$\sum_{k=0}^p a_{p-k} y^{(k)}(t) = \sum_{j=0}^q b_{q-j} x^{(j)}(t). \quad (4.48)$$

By “solution” we mean either that (almost all) sample functions satisfy the equations or that there exists a process  $\{y(t), t \in \mathbb{R}\}$  such that the two sides are equivalent. Note that (4.48) is only marginally more general than (4.47), since both right hand sides are stationary processes without any further assumption.

What can then be said about the solution to these equations: when does it exist and when is it a stationary process; and in that case, what is its spectrum and covariance function?

For the linear differential equation (4.47),

$$a_0 y^{(p)}(t) + a_1 y^{(p-1)}(t) + \dots + a_{p-1} y'(t) + a_p y(t) = x(t). \quad (4.49)$$

we define the *generating function*,

$$A(r) = a_0 + a_1 r + \dots + a_p r^p,$$

and the corresponding *characteristic equation*

$$r^p A(r^{-1}) = a_0 r^p + a_1 r^{p-1} + \dots + a_{p-1} r + a_p = 0. \quad (4.50)$$

The existence of a stationary process solution depends on the solutions to the characteristic equation. The differential equation (4.49) is called *stable* if the roots of the characteristic equation all have negative real part.

One can work with (4.49) as a special case of a multivariate first order differential equation. Dividing both sides by  $a_0$  the form is

$$\mathbf{y}' = \mathbf{A}\mathbf{y} + \mathbf{x}, \quad (4.51)$$

with  $\mathbf{y}(t) = (y(t), y'(t), \dots, y^{(p-1)}(t))'$ ,  $\mathbf{x}(t) = (0, 0, \dots, x(t))'$ , and

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_p & -a_{p-1} & a_{p-2} & \dots & -a_1 \end{pmatrix}.$$

This is the formulation which is common in linear and non-linear systems theory; cf. for example [14, Ch. 8], to which we refer for part of the following theorem.

**Theorem 4:9** a) If the differential equation (4.49) is stable, and the right hand side  $\{x(t), t \in \mathbb{R}\}$  is a stationary process, then there exists a stationary process  $\{y(t), t \in \mathbb{R}\}$  that solves the equation. The solution can be written as the output of a linear filter

$$y(t) = \int_{-\infty}^t h(t-u)x(u) du, \quad (4.52)$$

where the function  $h(u)$  solves the equation

$$a_0 h^{(p)}(t) + a_1 h^{(p-1)}(t) + \dots + a_{p-1} h'(t) + a_p h(t) = 0, \quad (4.53)$$

with initial conditions  $h(0) = h'(0) = \dots = h^{(p-2)}(0) = 0$ ,  $h^{(p-1)}(0) = 1/a_p$ . Further  $\int_{-\infty}^{\infty} |h(u)| du < \infty$ .

b) If  $\{x(t), t \in \mathbb{R}\}$  is a  $p$  times differentiable stationary process with spectral density  $f_x(\omega)$ , then also  $\sum_0^p a_j x^{(j)}(t)$  is a stationary process, and it has the spectral density

$$\left| \sum_0^p a_j (i\omega)^j \right|^2 f_x(\omega). \quad (4.54)$$

c) If  $\{x(t), t \in \mathbb{R}\}$  and  $\{y(t), t \in \mathbb{R}\}$  are two stationary processes that solve the differential equation (4.48), then their spectral densities obey the relation

$$\left| \sum a_k (i\omega)^k \right|^2 f_y(\omega) = \left| \sum b_j (i\omega)^j \right|^2 f_x(\omega). \quad (4.55)$$

**Proof:** a) If  $\{x(t), t \in \mathbb{R}\}$  has  $q$  times differentiable sample paths (with probability one), we use a standard result in ordinary differential equations to get a solution for almost every sample path; see [14, Ch. 8].

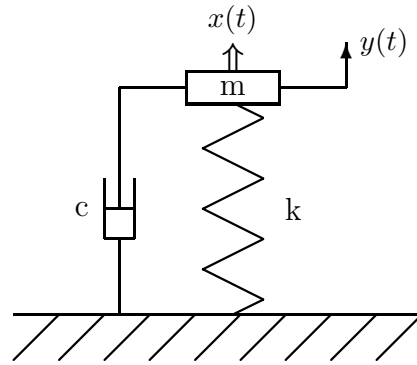
If we work only with second order properties one can take (4.52) as the definition of a process  $\{y(t), t \in \mathbb{R}\}$  and then show that it is  $p$  times differentiable (in quadratic mean) and that the two sides of (4.49) are equivalent.

Part b) and c) are easy consequences of the spectral process property (4.14). Just write the differential form in the right hand sides by means of the spectral representation and perform the integration.  $\square$

#### 4.4.3.3 The linear oscillator

The linear random oscillator is the basic ingredient in many engineering applications of stationary processes. We will examine two formulations from mechanical and electrical engineering, respectively.

**Example 4:7** First consider a spring-and-damper system as in Figure 4.2, with mass  $m$ , stiffness  $k$  and damping coefficient  $c$ . When the mass is subject to a regular or irregular varying force  $x(t)$  it moves more or less periodically, and we denote the displacement from the equilibrium by  $Y(t)$ ; see Figure 4.2.



**Figure 4.2:** A simple damped oscillator.

The relation between the force  $X(t)$  and the resulting displacement is described by the following differential equation,

$$my''(t) + cy'(t) + ky(t) = x(t). \quad (4.56)$$

Here  $\omega_0 = \sqrt{k/m}$  is called the *response frequency* or *eigenfrequency*, and  $\zeta = \frac{c}{2\sqrt{mk}}$  the *relative damping*. Expressed in terms of the damping and eigenfrequency the fundamental equation is

$$y''(t) + 2\zeta\omega_0 y'(t) + \omega_0^2 y(t) = m^{-1}x(t).$$

This equation can be solved just like an ordinary differential equation with a continuous  $x(t)$  and, from Theorem 4:9, it has the solution

$$y(t) = \int_{-\infty}^t h(t-u)x(u) du,$$

expressed with the impulse response function

$$h(u) = m^{-1}\tilde{\omega}_0^{-1}e^{-\alpha u} \sin(\tilde{\omega}u), \quad u \geq 0,$$

with the constants

$$\begin{aligned} \alpha &= \zeta\omega_0, \\ \tilde{\omega}_0 &= \omega_0(1 - \zeta^2)^{1/2}. \end{aligned}$$

To find the frequency function  $g(\omega)$  for the linear oscillator we consider each term on the left hand side in (4.56). Since differentiation has frequency function  $i\omega$ , and hence, repeated differentiation has frequency function  $-\omega^2$ , we see that  $g(\omega)$  satisfies the equation

$$\{-m\omega^2 + ci\omega + k\} \cdot g(\omega) = 1,$$

and hence

$$g(\omega) = \frac{1}{-m\omega^2 + ic\omega + k}. \quad (4.57)$$

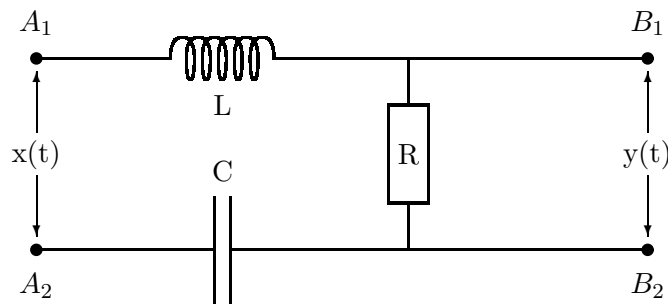
Since

$$|g(\omega)|^2 = \frac{1}{(k - m\omega^2)^2 + c^2\omega^2},$$

the spectral density for the output signal  $y(t)$  is

$$f_x(\omega) = \frac{f_x(\omega)}{(k - m\omega^2)^2 + c^2\omega^2} = \frac{f_x(\omega)/m^2}{(\omega_0^2 - \omega^2)^2 + 4\alpha^2\omega^2}. \quad (4.58)$$

**Example 4:8** A resonance circuit with one inductance, one resistance, and one capacitance in series is an electronic counterpart to the harmonic mechanical oscillator; see Figure 4.3.



**Figure 4.3:** A resonance circuit.

If the input potential between  $A_1$  and  $A_2$  is  $x(t)$ , the current  $I(t)$  through the circuit obeys the equation

$$LI'(t) + RI(t) + \frac{1}{C} \int_{-\infty}^t I(s) ds = x(t).$$

The output potential between  $B_1$  and  $B_2$ , which is  $y(t) = RI(t)$ , therefore follows the same equation (4.56) as the linear mechanical oscillator,

$$Ly''(t) + Ry'(t) + \frac{1}{C}y(t) = Rx'(t), \quad (4.59)$$

but this time with  $x'(t)$  as driving force. The frequency function for the filter between  $x(t)$  and  $y(t)$  is (cf. (4.57)),

$$g(\omega) = \frac{i\omega}{-(L/R)\omega^2 + i\omega + 1/(RC)}.$$

The response frequency  $\omega_0 = 1/\sqrt{LC}$  is here called the *resonance frequency*. The relative damping  $\zeta$  corresponds to the *relative bandwidth*  $1/Q = 2\zeta$ , where

$$1/Q = \Delta\omega/\omega_0 = R\sqrt{C/L},$$

and  $\Delta\omega = \omega_2 - \omega_1$  is such that  $|g(\omega_1)| = |g(\omega_2)| = |g(\omega_0)|/\sqrt{2}$ .

**Example 4:9** As a final example we consider

$$y''(t) + 2y'(t) + y(t) = x(t),$$

with the stationary process  $\{x(t), t \in \mathbb{R}\}$  as input. The frequency function for the filter in (4.52) is the solution to  $h''(u) + 2h'(u) + h(u) = 0$ , and is of the form

$$h(u) = e^{-u}(C_1 + C_2u),$$

where the boundary conditions give  $C_1 = 0, C_2 = 1$ . The solution

$$y(t) = \int_{-\infty}^t (t-u)e^{-(t-u)}x(t) du,$$

has the spectral density

$$f_y(\omega) = \frac{f_x(\omega)}{|1 + 2(i\omega) + (i\omega)^2|^2} = \frac{f_x(\omega)}{(1 + \omega^2)^2}.$$

#### 4.4.4 White noise in linear systems

##### 4.4.4.1 White noise in a linear differential equation

The Wiener process was constructed as a mathematical model for the Brownian motion of particles suspended in a viscous fluid in which the erratic particle movements are the results of bombardment by the fluid molecules. The Wiener process model requires that the fluid has zero viscosity and infinite mass.

A more realistic model gives room also for the viscosity and particle mass. If  $x(t)$  denotes the force acting on the particle and  $y(t)$  is the velocity, we get the Ornstein-Uhlenbeck differential equation (4.32) from Example 4:3,

$$a_0y'(t) + a_1y(t) = x(t), \quad (4.60)$$

where  $a_1$  depends on the viscosity and  $a_0$  is the particle mass. If the force  $x(t)$  is caused by collisions from independent molecules it is reasonable that different  $x(t)$  be independent. Adding the assumption that they are Gaussian leads us to take  $x(t) = \sigma w'(t)$  as the “derivative of a Wiener process”, i.e. Gaussian white noise,

$$a_0y'(t) + a_1y(t) = \sigma w'(t).$$

This equation can be solved as an ordinary differential equation by

$$y(t) = \frac{1}{a_0} \int_{-\infty}^t e^{-\alpha(t-u)} w'(u) du = \frac{1}{a_0} \int_{-\infty}^t e^{-\alpha(t-u)} dw(u). \quad (4.61)$$

Here the last integral is well defined, Example 2:7 on page 51, even if the differential equation we started out from is not.

By carrying out the integration it is easy to see that the process  $y(t)$  defined by (4.61) satisfies

$$a_1 \int_{u=t_0}^t y(u) du = a_0(y(t) - y(t_0)) + \sigma(w(t) - w(t_0)),$$

which means that, instead of equation (4.60), we could have used the integral equation

$$a_0(y(t) - y(t_0)) + a_1 \int_{u=t_0}^t y(u) du = \sigma(w(t) - w(t_0)), \quad (4.62)$$

to describe the increments of  $y(t)$ .

The general differential equation

$$a_0 y^{(p)}(t) + a_1 y^{(p-1)}(t) + \dots + a_{p-1} y'(t) + a_p y(t) = \sigma w'(t)$$

can be solved in a similar way, and expressed as a stochastic integral,

$$y(t) = \sigma \int_{-\infty}^t h(t-u) dw(u), \quad (4.63)$$

The impulse response function  $h(u)$  is the solution to

$$a_0 h^{(p)}(t) + a_1 h^{(p-1)}(t) + \dots + a_{p-1} h'(t) + a_p h(t) = 0,$$

as in Theorem 4:9. The formal differential equation can be replaced by the well defined differential-integral equation

$$\begin{aligned} & a_0(y^{(p-1)}(t) - y^{(p-1)}(t_0)) + a_1(y^{(p-2)}(t) - y^{(p-2)}(t_0)) + \\ & \quad + \dots + a_{p-1}(y(t) - y(t_0)) + a_p \int_{t_0}^t y(u) du \\ & = \sigma(w(t) - w(t_0)). \end{aligned}$$

Stochastic differential equations involving Gaussian white noise are often written as

$$a_0 dy(t) dt + a_1 y(t) dt = \sigma dw(t),$$

or more generally as

$$dy(t) = a(t)y(t) dt + \sigma(t) dw(t),$$

with variable deterministic coefficients.

In the most general form, with random coefficients,

$$dy(t) = a(y(t), t) y(t) dt + \sigma(y(t), t) dw(t),$$

a complete new theory is needed, namely *stochastic calculus*; the reader is referred to [39] for a good introduction.

#### 4.4.4.2 White noise and constant spectral density

A linear systems equation defines the relation between an input signal  $x(t)$  and a response process  $y(t)$ . The linear system acts as a frequency dependent amplifier and phase modifier on the input. Of special importance is the case when the input is *white noise*. This idealized type of process is strictly defined only in the context of linear systems. The characteristic feature of white noise, it may be denoted  $n(t)$  or, if it is Gaussian,  $w'(t)$ , is that all frequencies are represented in equal amount, i.e. it has *constant spectral density*

$$f_n(\omega) = \frac{\sigma^2}{2\pi}, \quad -\infty < \omega < \infty.$$

Strictly, this is not a proper spectral density of a stationary process since it has infinite integral, but used as input in a linear system with a impulse response function that satisfies  $\int |h(u)|^2 du < \infty$ , it produces a stationary output process.

**Theorem 4:10** a) *The stationary process  $\{x(t), t \in \mathbb{R}\}$ , defined as a stochastic integral from a standard Wiener process  $\{w(t), t \in \mathbb{R}\}$  by*

$$x(t) = \int_{-\infty}^{\infty} h(t-u) dw(u),$$

*has covariance function*

$$r_x(t) = \int_{-\infty}^{\infty} h(t-u)h(-u) du$$

*and spectral density*

$$f_x(\omega) = \frac{|g(\omega)|^2}{2\pi},$$

where  $g(\omega) = \int_{-\infty}^{\infty} e^{-i\omega u} h(u) du$  is the frequency response function corresponding to the impulse response  $h(u)$ .

b) *If  $x(t) = \int_{-\infty}^t h(t-u) dw(u)$  is a solution to a stable stochastic differential equation*

$$a_0 y^{(p)}(t) + a_1 y^{(p-1)}(t) + \dots + a_{p-1} y'(t) + a_p y(t) = \sigma w'(t), \quad (4.64)$$

*then its spectral density is*

$$f_x(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{\left| \sum_{k=0}^p a_k (i\omega)^{p-k} \right|^2}.$$

**Proof:** a) The covariance function is a direct consequence of Theorem 2:16. Now the right hand side of the integral expression for  $r_x(t)$  is the convolution of  $h(u)$  with  $h(-v)$  and their Fourier transforms are  $g(\omega)$  and  $\overline{g(\omega)}$ , respectively.

Since convolution corresponds to multiplication of the Fourier transforms, the spectral density of  $r_x(t)$  is, as stated,

$$f_x(\omega) = g(\omega)\overline{g(\omega)}/(2\pi) = |g(\omega)|^2/(2\pi).$$

b) The relation between the impulse response and the frequency response function  $g(\omega)$  is a property of the systems equation (4.64) and does not depend on any stochastic property. One can therefore use the established relation

$$g(\omega) = \frac{1}{\sum_{k=0}^p a_k (i\omega)^{p-k}}$$

to get the result.  $\square$

Part (b) of the theorem finally confirms our claim that the Gaussian white noise  $\sigma w'(t)$  can be treated as if it has constant spectral density  $\sigma^2/(2\pi)$ .

A stationary process with spectral density of the form  $\frac{C}{|P(i\omega)|^2}$  where  $P(\omega)$  is a complex polynomial can be generated as the output from a stable linear system with white noise input; this is a very convenient way to produce a stationary process with suitable spectral properties.

**Example 4:10** The linear oscillator

$$y''(t) + 2\zeta\omega_0 y'(t) + \omega_0^2 y(t) = n(t)$$

where the white noise input  $n(t)$  has constant spectral density,  $f_n(\omega) = \frac{\sigma^2}{2\pi}$ , has spectral density (cf. (4.58)),

$$f_y(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{(\omega_0^2 - \omega^2)^2 + 4\alpha^2\omega^2}.$$

The covariance function is found, for example by residue calculus, from  $r_y(t) = \int e^{i\omega t} f_y(\omega) d\omega$ . With  $\alpha = \zeta\omega_0$  and  $\tilde{\omega}_0 = \omega_0\sqrt{1-\zeta^2}$  one gets the covariance function

$$r_y(t) = \frac{\sigma^2}{4\alpha\omega_0^2} e^{-\alpha|t|} \left( \cos \tilde{\omega}_0 t + \frac{\alpha}{\tilde{\omega}_0} \sin \tilde{\omega}_0 |t| \right). \quad (4.65)$$

## 4.4.5 The Hilbert transform and the envelope

### 4.4.5.1 The Hilbert transform

By the spectral representation we have expressed a stationary process  $x(t)$  in terms of random cosine and sine functions with positive and negative frequencies. The complex form of the spectral representation,

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega),$$

yielded a real process by the requirement  $dZ(-\omega) = \overline{dZ(\omega)}$ . In fact,  $x(t)$  is then expressed as the sum of two complex processes, of which one is the complex conjugate of the other.

If we take only the half spectral representation,

$$x^*(t) = 2 \int_{0+}^{\infty} e^{i\omega t} dZ(\omega) + \Delta Z(0),$$

where  $\Delta Z(0)$  is the jump of  $Z(\omega)$  at the origin, we obtain a particularly useful linear transform of  $x(t)$ . One can obtain  $x^*(t)$  as the limit, as  $h \downarrow 0$  through continuity points of  $F(\omega)$ , of the linear operation with frequency function

$$g_h(\omega) = \begin{cases} 0, & \omega < -h, \\ 1, & |\omega| \leq h, \\ 2, & \omega > h. \end{cases}$$

The process  $\hat{x}(t)$ , defined by

$$x^*(t) = x(t) + i\hat{x}(t),$$

is the result of a linear filter on  $x(t)$  with frequency function

$$g(\omega) = \begin{cases} i, & \omega < 0, \\ 0, & \omega = 0, \\ -i, & \omega > 0. \end{cases}$$

It is called the *Hilbert transform* of  $x(t)$ .<sup>3</sup>

When  $x(t)$  is real, with  $dZ(-\omega) = \overline{dZ(\omega)}$ , with the real spectral representation (4.23),

$$x(t) = \int_{0+}^{\infty} \cos \omega t du(\omega) + \int_0^{\infty} \sin \omega t dv(\omega) + du(0),$$

it follows that also  $\hat{x}(t) = i(x(t) - x^*(t))$  is real, and that it is given by

$$\hat{x}(t) = \int_0^{\infty} \sin \omega t du(\omega) - \int_{0+}^{\infty} \cos \omega t dv(\omega). \quad (4.66)$$

Thus,  $x^*(t)$  is a complex process with  $x(t)$  as real part and  $\hat{x}(t)$  as imaginary part. All involved processes can be generated by the same real spectral processes  $\{u(\lambda), 0 \leq \lambda < \infty\}$  and  $\{v(\lambda), 0 \leq \lambda < \infty\}$ .

**Theorem 4:11** *Let  $\{x(t), t \in \mathbb{R}\}$  be stationary and real, with mean 0, covariance function  $r(t)$  and spectral distribution function  $F(\omega)$ , with a possible jump  $\Delta F(0)$  at  $\omega = 0$ . Denote the Hilbert transform of  $x(t)$  by  $\hat{x}(t)$ . Then, with  $G(\omega)$  denoting the one-sided spectrum,*

<sup>3</sup>MATLAB, Signal processing toolbox, contains a routine for making Hilbert transforms. Try it!

a)  $\{\widehat{x}(t), t \in \mathbb{R}\}$  is stationary and real, with mean 0, and covariance function

$$\widehat{r}(t) = r(t) - \Delta F(0) = \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega) - \Delta F(0) = \int_{0+}^{\infty} \cos \omega t dG(\omega).$$

b) The process  $\widehat{x}(t)$  has the same spectrum  $F(\omega)$  as  $x(t)$ , except that any jump at  $\omega = 0$  has been removed.

c) The cross-covariance function between  $x(t)$  and  $\widehat{x}(t)$  is

$$r^*(t) = E(x(s) \cdot \widehat{x}(s+t)) = \int_0^{\infty} \sin \omega t dG(\omega).$$

In particular,  $x(t)$  and  $\widehat{x}(t)$  are uncorrelated, taken at the same time instant.

**Proof:** Part (a) and (c) follow from (4.23), (4.66), and the correlation properties (4.24-4.26) of the real spectral processes. Then part (b) is immediate.  $\square$

#### 4.4.5.2 The envelope

Assume that  $F(\omega)$  is continuous at  $\omega = 0$  so there is no "constant" random component in  $x(t)$ , and consider the joint behavior of  $x(t)$  and its Hilbert transform  $\widehat{x}(t)$ . Also assume  $x(t)$ , and hence  $\widehat{x}(t)$ , to be Gaussian processes, with common covariance function  $r(t)$ , and consider the complex process

$$x^*(t) = x(t) + i\widehat{x}(t) = 2 \int_{0+}^{\infty} e^{i\omega t} dZ(\omega).$$

The *envelope*  $R(t)$  of  $x(t)$  is the absolute value of  $x^*(t)$ ,

$$R(t) = \sqrt{x(t)^2 + \widehat{x}(t)^2}.$$

In particular,  $|x(t)| \leq R(t)$ , with equality when  $\widehat{x}(t) = 0$ .

Since, for Gaussian processes,  $x(t)$  and  $\widehat{x}(t)$  are independent with the same Gaussian distribution, the envelope has a Rayleigh distribution with density

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, \quad r \geq 0.$$

The envelope, as defined here, does always exist, and it always has the stated statistical properties. The physical meaning of an envelope is however not clear from the mathematical definition, and we must turn to special types of processes before we can identify any particularly interesting properties of  $R(t)$ .

### 4.4.5.3 The envelope of a narrow band process

For processes with spectrum concentrated to a narrow frequency band, the sample functions have a characteristic "fading" look, as a wave with one dominating frequency. Then the envelope represents the slowly varying random amplitude.

We consider a stationary process  $x(t)$  with spectral density  $f(\omega)$  concentrated around some frequency  $\omega_0$ ,

$$f(\omega) = \frac{1}{2}f_0(\omega - \omega_0) + \frac{1}{2}f_0(\omega + \omega_0),$$

for some function  $f_0(\omega)$  such that  $f_0(\omega) = f_0(-\omega)$ , and  $f_0(\omega) = 0$  for  $|\omega| > d$ , some  $d < \omega_0$ . Express  $x(t)$  in spectral form, using the normalized form (4.22), to obtain

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} e^{i\omega t} \sqrt{f(\omega)} dW_C(\omega) \\ &= \int_{\omega_0-d}^{\omega_0+d} \sqrt{f_0(\omega - \omega_0)/2} e^{i\omega t} dW_C(\omega) \\ &\quad + \int_{-\omega_0-d}^{-\omega_0+d} \sqrt{f_0(\omega + \omega_0)/2} e^{i\omega t} dW_C(\omega) \\ &= I_1(t) + I_2(t), \quad \text{say.} \end{aligned}$$

By change of variables in  $I_1(t)$  and  $I_2(t)$  this gives,

$$\begin{aligned} I_1(t) &= e^{i\omega_0 t} \int_{-d}^d \sqrt{f_0(\omega)/2} e^{i\omega t} d_\omega W_C(\omega + \omega_0) = e^{i\omega_0 t} Y(t), \\ I_2(t) &= e^{-i\omega_0 t} \int_{-d}^d \sqrt{f_0(\omega)/2} e^{i\omega t} d_\omega W_C(\omega - \omega_0) \\ &= e^{-i\omega_0 t} \int_{-d}^d \sqrt{f_0(\omega)/2} e^{-i\omega t} d_\omega W_C(-\omega - \omega_0) = \overline{I_1(t)}, \end{aligned}$$

and, in combination,

$$x(t) = 2 \Re e Y(t) e^{i\omega_0 t}.$$

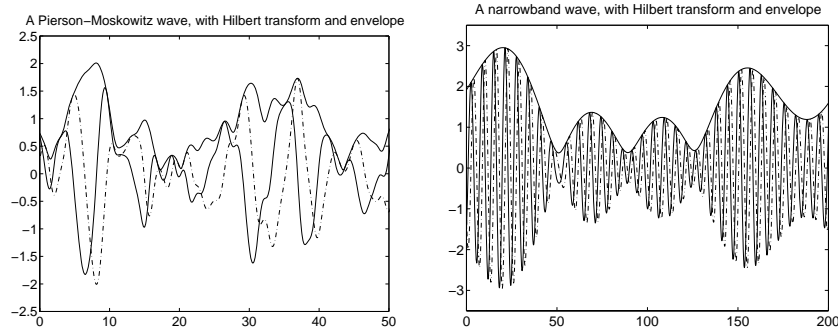
Here,  $Y(t)$  is a complex process

$$Y(t) = Y_1(t) + iY_2(t) = \int_{-d}^d \sqrt{f_0(\omega)/2} e^{i\omega t} d_\omega W_C(\omega + \omega_0)$$

with only low frequencies. With  $R(t) = 2|Y(t)|$  and  $\Theta(t) = \arg Y(t)$ , we obtain

$$x(t) = R(t) \Re e e^{i(\omega_0 t + \Theta(t))} = R(t) \cos(\omega_0 t + \Theta(t)).$$

The envelope  $R(t)$  has here a real physical meaning as the slowly varying amplitude of the narrow band process.



**Figure 4.4:** *Gaussian processes and their Hilbert transforms and envelopes. Left: Pierson-Moskowitz waves, Right: process with triangular spectrum over  $(0.8, 1.2)$ .*

#### 4.4.6 The sampling theorem

The spectral form expresses a stationary process as an integral, in quadratic mean, of elementary cosine functions with random amplitude and phase. If all these amplitudes and phases are known, the process can be reconstructed. For *band limited* processes, this is particularly simple. A process is band limited to frequency  $\omega_0$  if

$$F(-\omega_0+) - F(-\infty) = F(\infty) - F(\omega_0-) = 0,$$

i.e. its spectrum is restricted to the interval  $[-\omega_0, \omega_0]$ . We require that there is no spectral mass at the points  $\pm\omega_0$ .

**Theorem 4:12** *If the stationary process  $\{x(t), t \in \mathbb{R}\}$  is band limited to  $\omega_0$ , then it is perfectly specified by its values at discrete time points spaced  $t_0 = \pi/\omega_0$  apart. More specifically, with probability one,*

$$x(t) = \sum_{k=-\infty}^{\infty} x(\alpha + kt_0) \cdot \frac{\sin \omega_0(t - \alpha - kt_0)}{\omega_0(t - \alpha - kt_0)}, \quad (4.67)$$

where  $\alpha$  is an arbitrary constant.

**Proof:** The spectral representation says that

$$x(t) = \int_{-\omega_0+}^{\omega_0-} e^{i\omega t} dZ(\omega).$$

For a fixed  $t$ , the function  $g_t(\omega) = e^{i\omega t} \cdot I_{[-\omega_0, \omega_0]}$  is square integrable over  $(-\omega_0, \omega_0)$ , and from the theory of Fourier series, it can be expanded as

$$e^{i\omega t} = \lim_{N \rightarrow \infty} \sum_{k=-N}^N e^{i\omega kt_0} \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)},$$

with convergence in  $\mathcal{H}(F)$ , i.e.

$$\int_{-\omega_0}^{\omega_0} \left| e^{i\omega t} - \sum_{k=-N}^N e^{i\omega kt_0} \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)} \right|^2 dF(\omega) \rightarrow 0.$$

The convergence is also uniform for  $|\omega| < \omega_0 = \pi/t_0$ . For  $\omega = \pm\omega_0$  it converges to

$$\frac{e^{i\omega_0 t} + e^{-i\omega_0 t}}{2} = \cos \omega_0 t.$$

Therefore, if  $dF(\pm\omega_0) = 0$ , then

$$\sum_{k=-N}^N e^{i\omega kt_0} \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)} \rightarrow e^{i\omega t}, \quad (4.68)$$

in  $\mathcal{H}(F)$  as  $N \rightarrow \infty$ .

Inserting this expansion in the spectral representation of  $\{x(t), t \in \mathbb{R}\}$  we obtain,

$$\begin{aligned} E \left( \left| x(t) - \sum_{k=-N}^N x(kt_0) \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)} \right|^2 \right) & \quad (4.69) \\ &= E \left( \left| \int_{-\omega_0}^{\omega_0} \left( e^{i\omega t} - \sum_{k=-N}^N e^{i\omega kt_0} \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)} \right) dZ(\omega) \right|^2 \right) \\ &\leq \int_{-\omega_0}^{\omega_0} \left| e^{i\omega t} - \sum_{k=-N}^N e^{i\omega kt_0} \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)} \right|^2 dF(\omega). \end{aligned}$$

By (4.68), this goes to 0 as  $N \rightarrow \infty$ , i.e.

$$x(t) = \lim_{N \rightarrow \infty} \sum_{k=-N}^N x(kt_0) \cdot \frac{\sin \omega_0(t - kt_0)}{\omega_0(t - kt_0)},$$

in  $\mathcal{H}(x)$ , which is the statement of the theorem for  $\alpha = 0$ . For arbitrary  $\alpha$ , apply the just proved result to  $y(t) = x(t + \alpha)$ .  $\square$

**Remark 4:2** If there is spectral mass  $F^+ = dF(\omega_0)$ ,  $F^- = dF(\omega_0)$ , at the endpoints  $\pm\omega_0$ , then (4.69) would tend to

$$\sin^2 \omega_0 t (F^+ + F^-),$$

failing the sampling representation.

An example of this is the simple random cosine process,  $x(t) = \cos(\omega_0 t + \phi)$ , which has covariance function  $r(t) = \frac{1}{2} \cos \omega_0 t$ , and spectrum concentrated at  $\pm\omega_0$ . Then

$$x(\alpha + kt_0) = (-1)^k x(\alpha),$$

which means that for every  $t$ , the sum

$$\sum_k x(\alpha + kt_0) \cdot \frac{\sin \omega_0(t - \alpha - kt_0)}{\omega_0(t - \alpha - kt_0)}$$

is proportional to  $x(\alpha)$ . On the other hand  $x(\alpha + t_0/2)$  is uncorrelated with  $x(\alpha)$  and cannot be represented by the sampling theorem.

## 4.5 Karhunen-Loève expansion

### 4.5.1 Principal components

In a multivariate distribution of a random vector the components may be more or less statistically dependent. If there is strong dependence between the components it is possible that it suffices to specify a few (random) values in order to specify almost the entire outcome of the full random vector. The formal tool to generate such a common behavior is the concept of *principal components*, which is a tool to approximately reduce the dimensionality of a variation space.

Let  $\mathbf{x} = (x_1, \dots, x_n)'$  be a vector of  $n$  random variables with mean zero and a covariance matrix  $\Sigma$ , which is symmetric and non-negative definite by construction. The covariance matrix has  $n$  eigenvalues  $\omega_k$  with corresponding orthonormal eigenvectors  $\mathbf{p}_k$ , decreasingly ordered as  $\omega_1 \geq \omega_2 \geq \dots \geq \omega_n$ , such that

$$\Sigma \mathbf{p}_k = \omega_k \mathbf{p}_k.$$

The transformation

$$z_k = \frac{1}{\sqrt{\omega_k}} \mathbf{p}_k' \mathbf{x}$$

gives us  $n$  new standardized random variables

$$V(z_k) = \frac{\mathbf{p}_k' \Sigma \mathbf{p}_k}{\omega_k} = \frac{\mathbf{p}_k' \mathbf{p}_k \omega_k}{\omega_k} = 1.$$

Furthermore, they are uncorrelated, for  $j \neq k$ ,

$$\text{Cov}(z_j, z_k) = E(z_j z_k) = \frac{1}{\sqrt{\omega_j \omega_k}} E(\mathbf{p}_j' \mathbf{x} \mathbf{x}' \mathbf{p}_k) = \frac{1}{\sqrt{\omega_j \omega_k}} \mathbf{p}_j' \Sigma \mathbf{p}_k = 0,$$

since the eigenvectors  $\mathbf{p}_j$  and  $\mathbf{p}_k$  are orthogonal.

The random variables  $y_k = \sqrt{\omega_k} z_k$ ,  $k = 1, \dots, n$ , are called the *principal components* of the vector  $\mathbf{x}$ . In matrix language, with  $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$  as the matrix with the eigenvectors as columns,

$$\mathbf{y} = \mathbf{P}\mathbf{x} \quad \text{with inverse} \quad \mathbf{x} = \mathbf{P}'\mathbf{y},$$

is a vector of uncorrelated variables with decreasing variances  $\omega_k$ .

Since the matrix  $\mathbf{P}$  is orthogonal,  $\mathbf{P}^{-1} = \mathbf{P}'$ , the original  $x$ -variables can be expressed as a linear combination of the uncorrelated variables  $y_k$  and  $z_k$ ,

$$x_k = \mathbf{p}'_k \mathbf{y} = \sum_{j=1}^n \sqrt{\omega_j} \mathbf{p}_k z_j. \quad (4.70)$$

In practice, when one wants to simulate a large vector of many correlated Gaussian variables, one can use (4.70) to generate successively the most important variational modes and truncate the sum when, for example it describes 99% of the variation of the  $x$ -variables.

#### 4.5.2 Expansion of a stationary process along eigenfunctions

We shall now generalize the finite-dimensional formulation in the previous section to continuous parameter stochastic processes. By the spectral representation,  $x(t) = \int e^{i\omega t} dZ(\omega)$ , every stationary process can be expressed by means of uncountably many orthogonal variables  $dZ(\omega)$ . For processes with discrete spectrum with jumps at  $\pm\omega_k$ ,  $Z(\omega)$  has countable many jumps  $\Delta Z(\omega_k)$  and  $x(t) = \sum e^{i\omega_k t} \Delta_k Z$ .

In fact, every quadratic mean continuous process, stationary or not, can be expressed on a finite interval  $[a, b]$  as a sum of deterministic functions with random *orthogonal* coefficients,

$$x(t) = \lim_{n \rightarrow \infty} \sum_0^n c_k(t) z_k.$$

The convergence is uniform for  $a \leq t \leq b$ , in the sense that

$$E \left( \left| x(t) - \sum_0^n c_k(t) z_k \right|^2 \right) \rightarrow 0,$$

uniformly in  $t$ , as  $n \rightarrow \infty$ .

The functions  $c_k(t)$  depend on the choice of observation interval  $[a, b]$ , and the random variables  $z_k$  are elements in the Hilbert space spanned by  $x(t); t \in [a, b]$ ,

$$z_k \in \mathcal{H}(x(t); t \in [a, b]).$$

Let us first investigate what properties such an expansion should have, if it exists. Write  $\mathcal{H}(x)$  instead of  $\mathcal{H}(x(t); t \in [a, b])$ . Suppose there exists  $z_k$  with

$$\|z_k\|_{\mathcal{H}(x)}^2 = 1, \quad (z_j, z_k) = E(z_j \overline{z_k}) = 0, j \neq k,$$

and assume that the family  $\{z_k\}$  is complete for  $\mathcal{H}(x)$ , i.e. the  $z_k$  form a basis for  $\mathcal{H}(x)$ . In particular this means that for every  $U \in \mathcal{H}(x)$ ,

$$U \perp z_k, \forall k \Rightarrow U = 0.$$

Now take any  $y \in \mathcal{H}(x)$ , and define  $c_k = (y, z_k)$ . Then, by the orthogonality,

$$E \left( \left| y - \sum_0^n c_k z_k \right|^2 \right) = \dots = E(|y|^2) - \sum_0^n |c_k|^2,$$

so  $\sum_0^n |c_k|^2 \leq \|y\|_{\mathcal{H}(x)}^2$  for all  $n$ , and hence  $\sum_0^\infty |c_k|^2 \leq \|y\|_{\mathcal{H}(x)}^2$ . This means that  $\sum_0^\infty c_k z_k$  exists as a limit in quadratic mean, and also that

$$y - \sum_0^\infty c_k z_k \perp z_n$$

for all  $n$ . But since  $\{z_k\}$  is a complete family,  $y = \sum_0^\infty c_k z_k = \sum_0^\infty (y, z_k) z_k$ .

Now replace  $y$  by a fixed time observation of  $x(t)$ . Then, naturally, the  $c_k$  will depend on the time  $t$  and are functions  $c_k(t)$ ,  $x(t) = \sum c_k(t) z_k$ . For the covariance function of  $x(t) = \sum_k c_k(t) z_k$  we have, by the orthogonality of  $z_k$ ,

$$r(s, t) = E(x(s)\overline{x(t)}) = \sum_{j,k} c_j(s) \overline{c_k(t)} E(z_j \overline{z_k}) = \sum_k c_k(s) \overline{c_k(t)}.$$

Thus, we shall investigate the existence and properties of the following pair of expansions,

$$x(t) = \sum_k c_k(t) z_k, \quad (4.71)$$

$$r(s, t) = \sum_k c_k(s) \overline{c_k(t)}. \quad (4.72)$$

Not only can the  $z_k$  be taken as uncorrelated but also the functions  $c_k(t)$  can be chosen as *orthogonal*, i.e.

$$\int_a^b c_j(t) \overline{c_k(t)} dt = 0, \quad j \neq k$$

$$\int_a^b |c_k(t)|^2 dt = \omega_k \geq 0.$$

As a final check on the consequences of an expansion (4.72) and the ortho-

gonality of the functions  $c_k(t)$ , we observe the following arguments<sup>4</sup>,

$$\begin{aligned} \int_a^b r(s, t) c_j(t) dt &= \int_a^b \left\{ \sum_0^\infty c_k(s) \overline{c_k(t)} \right\} c_j(t) dt \\ &= \sum_0^\infty c_k(s) \int_a^b c_j(t) \overline{c_k(t)} dt = \sum_0^\infty c_k(s) \omega_k \delta_{j-k} = \omega_j c_j(s). \end{aligned}$$

Thus, the functions  $c_j(t)$  are eigenfunctions with eigenvalues  $\omega_j$  to the covariance operator  $r(s, t)$ ,

$$c(\cdot) \mapsto \int_a^b r(\cdot, t) c(t) dt.$$

Call the normalized eigenfunctions  $\phi_j(t) = \frac{1}{\sqrt{\omega_j}} c_j(t)$  if  $\omega_j > 0$ , making  $\phi_k(t)$  a family of orthonormal eigenfunctions.

### 4.5.3 The Karhunen-Loève theorem

**Theorem 4:13** *Let  $\{x(t); a \leq t \leq b\}$  be continuous in quadratic mean with mean zero and covariance function  $r(s, t) = E(x(s)\overline{x(t)})$ . Then there exist orthonormal eigenfunctions  $\phi_k(t), k = 0, 1, \dots, N \leq \infty$ , for  $a \leq t \leq b$ , with eigenvalues  $\omega_k \geq 0$ , to the equation*

$$\int_a^b r(s, t) \phi(t) dt = \omega \phi(s),$$

such that the random variables

$$z_k = \frac{1}{\sqrt{\omega_k}} \int_a^b \overline{\phi_k(t)} x(t) dt$$

are uncorrelated, and can represent  $x(t)$  as

$$x(t) = \sum_0^\infty \sqrt{\omega_k} \phi_k(t) z_k. \quad (4.73)$$

The sum is a limit in quadratic mean and  $\|x(t) - \sum_0^n \sqrt{\omega_k} \phi_k(t) z_k\|_{\mathcal{H}(x)}^2 \rightarrow 0$  uniformly for  $t \in [a, b]$ .

The variables  $z_k$  are sometimes called *observables* and they can be used for example to make statistical inference about the distribution of the process  $x(t)$ . Note, that if  $x(t)$  is a normal process, then the  $z_k$  are uncorrelated normal variables, and hence independent, making inference simple.

Before we sketch the proof of the theorem, we present an explicit construction of the Wiener process as an example of the Karhunen-Loève theorem.

<sup>4</sup>The termwise integration is allowed, since  $\sup_{s, t \in [a, b]} |\sum_0^\infty c_k(s) \overline{c_k(t)}| \leq \sup_t r(t, t) < \infty$ . Prove this as an exercise.

**Example 4:11** The standard Wiener process  $w(t)$ , observed over  $[0, T]$  has covariance function  $r(s, t) = \min(s, t)$ , and eigenfunctions can be found explicitly: from

$$\int_0^T \min(s, t) \phi(t) dt = \omega \phi(s),$$

it follows by twice differentiation,

$$\int_0^s t \phi(t) dt + \int_s^T s \phi(t) dt = \omega \phi(s) \quad (4.74)$$

$$s \phi(s) - s \phi(s) + \int_s^T \phi(t) dt = \omega \phi'(s) \quad (4.75)$$

$$-\phi(s) = \omega \phi''(s). \quad (4.76)$$

The initial conditions,  $\phi(0) = 0$ ,  $\phi'(T) = 0$ , obtained from (4.74) and (4.75), imply the solution  $\phi(t) = A \sin \frac{t}{\sqrt{\omega}}$ , with  $\cos \frac{T}{\sqrt{\omega}} = 0$ . Thus, the positive eigenvalues  $\omega_k$  satisfy

$$\frac{T}{\sqrt{\omega_k}} = \frac{\pi}{2} + k\pi, \quad k = 0, 1, 2, \dots$$

The normalized eigenfunctions are

$$\phi_k(t) = \sqrt{\frac{2}{T}} \sin \frac{(k + \frac{1}{2})\pi t}{T},$$

with eigenvalues

$$\omega_k = T^2 \frac{1}{\pi^2(k + \frac{1}{2})^2}.$$

With

$$z_k = \frac{1}{\sqrt{\omega_k}} \int_0^T \phi_k(t) w(t) dt = \frac{\pi(k + \frac{1}{2})}{T} \int_0^T \sqrt{\frac{2}{T}} \cdot \sin \frac{t\pi(k + \frac{1}{2})}{T} \cdot w(t) dt,$$

we have that the Wiener process can be defined as the infinite (uniformly convergent in quadratic mean) sum

$$w(t) = \sqrt{\frac{2}{T}} \sum_{k=0}^{\infty} \frac{\sin \frac{\pi t(k + \frac{1}{2})}{T}}{\frac{\pi(k + \frac{1}{2})}{T}} z_k,$$

with independent standard normal variables  $z_k$ .

The reader should simulate a Wiener process  $w(t)$ , find the variables  $z_k$  for  $k = 0, 1, \dots, n < \infty$  by numerical integration, and reproduce  $w(t)$  as a truncated sum.

**Proof:** (of Theorem 4:13) We only indicate the steps in the proof, following the outline in [35]. One has to show: the mathematical facts about existence and properties of eigenvalues and eigenfunctions, the convergence of the series (4.73), and finally the stochastic properties of the variables  $z_k$ . This is done in a series of steps.

(i) If  $\int_a^b r(s, t) \phi(t) dt = \omega \phi(s)$ , then  $\omega$  is real and non-negative. This follows from

$$0 \leq \int_a^b \int_a^b r(s, t) \overline{\phi(s)} \phi(t) ds dt = \omega \int_a^b |\phi(s)|^2 ds.$$

(ii) There is at least one non-zero eigenvalue. The largest eigenvalue is

$$\omega_0 = \max_{\phi: \|\phi\|=1} \int_a^b \int_a^b r(s, t) \overline{\phi(s)} \phi(t) ds dt,$$

where the maximum is taken over  $\|\phi\|^2 = \int_a^b |\phi(t)|^2 dt = 1$ . As stated in [35], "this is not easily proved". The corresponding eigenfunction is denoted by  $\phi_0(t)$ , and it is continuous.

(iii) The function  $r_1(s, t) = r(s, t) - \omega_0 \phi_0(s) \overline{\phi_0(t)}$  is a continuous covariance function, namely for the process

$$x_1(t) = x(t) - \phi_0(t) \int_a^b \overline{\phi_0(s)} x(s) ds,$$

and it holds that

$$\int_a^b r_1(s, t) \phi_0(t) dt = 0. \quad (4.77)$$

Repeating step (ii) with  $r_1(s, t)$  instead of  $r(s, t)$  we get a new eigenvalue  $\omega_1 \geq 0$  with eigenfunction  $\phi_1(t)$ . Since  $\int_a^b r_1(s, t) \phi_1(t) dt = \omega_1 \phi_1(s)$ , we have

$$\begin{aligned} \int_a^b \phi_1(s) \overline{\phi_0(t)} dt &= \frac{1}{\omega_1} \int_a^b \overline{\phi_0(s)} \left\{ \int_a^b r_1(s, t) \phi_1(t) dt \right\} ds \\ &= \frac{1}{\omega_1} \int_a^b \phi_1(t) \left\{ \int_a^b r_1(s, t) \overline{\phi_0(s)} ds \right\} dt = 0, \end{aligned}$$

according to (4.77) since  $r_1(s, t)$  is real. Thus  $\phi_0$  and  $\phi_1$  are orthogonal.

It also follows that  $\phi_1$  is an eigenfunction to  $r(s, t)$ ,

$$\begin{aligned} \int_a^b r(s, t) \phi(t) dt &= \int_a^b r_1(s, t) \phi_1(t) dt + \omega_0 \phi_0(s) \int_a^b \overline{\phi_0(t)} \phi_1(t) dt \\ &= \omega_1 \phi_1(s) + 0. \end{aligned}$$

(iv) Repeat (ii) and (iii) as long as there is anything remaining of

$$r_n(s, t) = r(s, t) - \sum_{k=0}^n \omega_k \phi_k(s) \overline{\phi_k(t)}.$$

Then, either there is a finite  $n$  such that  $r_n \equiv 0$ , or there is an infinite decreasing sequence of positive eigenvalues  $\omega_k \downarrow 0$  with  $\sum_k \omega_k < \infty$ . Show as an exercise that  $\sum_k \omega_k \leq \int_a^b r(s, s) dt$ .

(v) If there is an infinite number of eigenvalues, then

$$\sup_{a \leq s, t \leq b} \left| r(s, t) - \sum_{k=0}^n \omega_k \phi_k(s) \overline{\phi_k(t)} \right| \rightarrow 0$$

as  $n \rightarrow \infty$ , i.e.

$$r(s, t) = \sum_{k=0}^{\infty} \omega_k \phi_k(s) \overline{\phi_k(t)},$$

with uniform convergence. (This is Mercer's theorem from 1909.)

(vi) For the representation (4.73) we have

$$\begin{aligned} E \left| x(t) - \sum_0^n \sqrt{\omega_k} \phi_k(t) z_k \right|^2 &= E(|x(t)|^2) - \sum_0^n \omega_k |\phi_k(t)|^2 \\ &= r(t, t) - \sum_0^n \omega_k |\phi_k(t)|^2 \rightarrow 0, \end{aligned}$$

uniformly in  $a \leq t \leq b$ , according to (v), as  $n \rightarrow \infty$ .

(vii) The properties of  $z_k$  follow from the orthogonality of the eigenfunctions.  $\square$

**Example 4:12** As a classical engineering problem about signal detection we shall illustrate the use of the Karhunen-Loève expansion we shall show how one can make hypothesis testing on the mean value function for a Gaussian process  $x(t)$  with known covariance function  $r(s, t) = Cov(x(s), x(t))$  but unknown mean value function  $m(t) = E(x(t))$ . Following Grenander, [15], suppose we have observed  $x(t)$  for  $a \leq t \leq b$ , and that we have two alternative hypotheses about  $m(t)$ ,

$$H_0 : m(t) = m_0(t),$$

$$H_1 : m(t) = m_1(t)$$

wanting to ascertain which one is most likely true.

We calculate the independent  $N(0, 1)$ -variables  $z_k$ ,

$$x(t) = m(t) + \sum_0^{\infty} \sqrt{\omega_k} \phi_k(t) z_k = m(t) + \tilde{x}(t),$$

where  $z_k = \frac{1}{\sqrt{\omega_k}} \int_a^b \overline{\phi_k(t)} \tilde{x}(t) dt$ . The  $z_k$  are not observable, since they require  $\tilde{x}(t) = x(t) - m(t)$ , and  $m(t)$  is unknown, but we can introduce observable independent Gaussian variables  $y_k$  by a similar procedure working on  $x(t)$ ,

$$y_k = \int_a^b \overline{\phi_k(t)} x(t) dt = \int_a^b \overline{\phi_k(t)} m(t) dt + z_k \sqrt{\omega_k}.$$

Writing

$$a_k = \int_a^b \overline{\phi_k(t)} m(t) dt$$

$$a_{ik} = \int_a^b \overline{\phi_k(t)} m_i(t) dt, \quad i = 0, 1,$$

we have that  $y_k \in N(a_k, \sqrt{\omega_k})$  and the hypotheses are transformed into hypotheses about  $a_k$ :

$$H_0 : a_k = a_{0k}, \quad k = 0, 1, \dots$$

$$H_1 : a_k = a_{1k}, \quad k = 0, 1, \dots$$

Testing hypotheses about an infinite number of independent Gaussian variables is no more difficult than for finitely many. The *likelihood ratio* (LR) test can be used in any case. Let the parameters be  $\mathbf{T}_0 = (a_{00}, a_{01}, \dots, a_{0n}, \dots)$  and  $\mathbf{T}_1 = (a_{10}, a_{11}, \dots, a_{1n}, \dots)$ , respectively. The LR-test based on  $y_0, \dots, y_n$  rejects  $H_0$  if the likelihood ratio  $p_n(\mathbf{y}) = f_1(\mathbf{y})/f_0(\mathbf{y}) > c$ , where the constant  $c$  determines the significance level. In this case

$$p_n(\mathbf{y}) = \frac{f_{y_0, \dots, y_n}(y_0, \dots, y_n; \mathbf{T}_1)}{f_{y_0, \dots, y_n}(y_0, \dots, y_n; \mathbf{T}_0)} = \frac{\prod_{k=0}^n \frac{1}{\sqrt{2\pi\omega_k}} e^{-(y_k - a_{1k})^2/2\omega_k}}{\prod_{k=0}^n \frac{1}{\sqrt{2\pi\omega_k}} e^{-(y_k - a_{0k})^2/2\omega_k}}$$

$$= \exp \left\{ - \sum_{k=0}^n \frac{y_k(a_{0k} - a_{1k})}{\omega_k} + \frac{1}{2} \sum_{k=0}^n \frac{a_{0k}^2 - a_{1k}^2}{\omega_k} \right\} = \exp \left\{ - \sum_{k=0}^n U_k \right\},$$

say. The LR-test thus rejects  $H_0$  if  $\sum_0^n U_k < c_\alpha$ .

We now let  $n \rightarrow \infty$ , and examines the limit of the test quantity. If

$$\sum_0^\infty \frac{a_{0k}^2}{\omega_k} < \infty \quad \text{and} \quad \sum_0^\infty \frac{a_{1k}^2}{\omega_k} < \infty,$$

then the sums converge; in particular  $\sum_0^\infty U_k$  converges in quadratic mean to a normal random variable.

Since

$$E(U_k) = \begin{cases} +\frac{1}{2} \frac{(a_{1k} - a_{0k})^2}{\omega_k} & \text{if } H_0 \text{ is true} \\ -\frac{1}{2} \frac{(a_{1k} - a_{0k})^2}{\omega_k} & \text{if } H_1 \text{ is true} \end{cases}$$

and

$$V(U_k) = \frac{(a_{1k} - a_{0k})^2}{\omega_k}$$

under both  $H_0$  and  $H_1$ , we have that  $\sum_0^\infty U_k$  is normal with mean

$$E\left(\sum_0^\infty U_k\right) = \begin{cases} m_0 = \sum_0^\infty \frac{1}{2} \frac{(a_{1k} - a_{0k})^2}{\omega_k} & \text{if } H_0 \text{ is true} \\ m_1 = -m_0 = -\sum_0^\infty \frac{1}{2} \frac{(a_{1k} - a_{0k})^2}{\omega_k} & \text{if } H_1 \text{ is true} \end{cases}$$

and with variance

$$V\left(\sum_0^\infty U_k\right) = 2(m_0 - m_1) = 4m_0.$$

Thus,  $H_0$  is rejected if  $\sum_0^\infty U_k < m_0 - \omega_\alpha 2\sqrt{m_0}$ , where  $\omega_\alpha$  is the upper normal  $\alpha$ -quantile.

If  $\sum_0^\infty \frac{(a_{0k} - a_{1k})^2}{\omega_k} < \infty$ , the test can be expressed in a simple way by the observation that

$$\sum_0^\infty U_k = \int_a^b f(t) \left( x(t) - \frac{m_0(t) + m_1(t)}{2} \right) dt,$$

where

$$f(t) = \sum_0^\infty \frac{a_{0k} - a_{1k}}{\omega_k} \phi_k(t).$$

## Exercises

- 4:1. Let  $x(t)$  be a stationary Gaussian process with  $E(x(t)) = 0$ , covariance function  $r_x(t)$  and spectral density  $f_x(\omega)$ . Calculate the covariance function for the process

$$y(t) = x^2(t) - r_x(0),$$

and show that it has the spectral density

$$f_y(\omega) = 2 \int_{-\infty}^{\infty} f_x(\mu) f_x(\omega - \mu) d\mu.$$

- 4:2. Derive the spectral density for  $u(t) = 2x(t)x'(t)$  if  $x(t)$  is a differentiable stationary Gaussian process with spectral density  $f_x(\omega)$ .
- 4:3. Let  $e_t$ ,  $t = 0, \pm 1, \pm 2, \dots$  be independent  $N(0, 1)$ -variables and define the stationary processes

$$x_t = \theta x_{t-1} + e_t = \sum_{n=-\infty}^t \theta^{t-n} e_n,$$

$$y_t = e_t + \psi e_{t-1},$$

with  $|\theta| < 1$ . Find the expressions for the spectral processes  $Z_x(\omega)$  and  $Z_y(\omega)$  in terms of the spectral process  $Z_e(\omega)$ , and derive the cross spectrum between  $x_t$  and  $y_t$ . (Perhaps you should read Chapter 6 first.)

- 4:4. Let  $u_n$  and  $v_n$  be two sequences of independent, identically distributed variables with zero mean and let the stationary sequences  $x_n$  and  $y_n$  be defined by

$$y_n = a_1 + b_1 x_{n-1} + u_n$$

$$x_n = a_2 - b_2 y_n + v_n.$$

Express the spectral processes  $dZ_x$  and  $dZ_y$  as functions of  $u_n$  and  $v_n$ , and derive the spectral densities for  $x_n$  and  $y_n$  and their cross spectrum.

- 4:5. Use the limit

$$\lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin \omega t}{t} dt = \begin{cases} 1 & \text{for } \omega > 0, \\ 0 & \text{for } \omega = 0, \\ -1 & \text{for } \omega < 0. \end{cases}$$

for the following alternative derivation of the spectral representation of a stationary process  $x(t)$  with spectral distribution function  $F(\omega)$ .

- a) First show that the following integral and limit exists in quadratic mean:

$$Z(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-i\omega t} - 1}{-it} x(t) dt.$$

- b) Then show that the process  $Z(\omega)$ ,  $-\infty < \omega < \infty$ , has orthogonal increments and that

$$E|Z(\omega_2) - Z(\omega_1)|^2 = F(\omega_2) - F(\omega_1)$$

for  $\omega_1 < \omega_2$ .

- c) Finally, show that the integral

$$\int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega) = \lim \sum e^{i\omega_k t} (Z(\omega_{k+1}) - Z(\omega_k))$$

exists, and that  $E|x(t) - \int e^{i\omega t} dZ(\omega)|^2 = 0$ .

- 4:6. Complete the proof of Lemma 4.1 on page 102.

- 4:7. Consider the covariance function

$$r_y(t) = \frac{\sigma^2}{4\alpha\omega_0^2} e^{-\alpha|t|} \left( \cos \tilde{\omega}_0 t + \frac{\alpha}{\tilde{\omega}_0} \sin \tilde{\omega}_0 |t| \right),$$

of the linear oscillator in Example 4:10 on page 115.

The covariance function contains some  $|t|$ ; show that the covariance functions fulfil a condition for sample function differentiability, but not for twice differentiability.

Find the relation between the relative damping *zeta* and the spectral width parameter  $\alpha = \omega_2 / \sqrt{\text{omega}_0 \omega_4}$ .

- 4:8. Prove that  $\sup_{s,t \in [a,b]} |\sum_0^\infty c_k(s) \overline{c_k(t)}| \leq \sup_t r(t, t) < \infty$  in the expansions (4.71) and (4.72).

- 4:9. Prove that in step (iv) in the proof of Theorem 4:13,

$$\sum_k \omega_k = \sum_k |\phi_k(t)|^2 dt \leq \max_{a \leq s \leq b} r(s, s) \cdot (b - a)^2.$$



## Chapter 5

# Ergodic theory and mixing

The concept of ergodicity is one of the most fundamental in probability, since it links the mathematical theory of probability to what can be observed in a deterministic mechanical world. It also plays an important role in statistical physics and in fact, the term *ergodic* was coined by Ludwig Boltzmann in 1887 in his study of the time development of mechanical particle systems. The term itself stems from the Greek *ergos* = "work" and *hodos* = "path", possibly meaning that ergodic theory is about how the energy in a system evolves with time. Ergodicity in itself is not a probabilistic concept and it can be studied within a purely deterministic framework, but it is only in terms of probabilities that a sensible interpretation can be given to the basic ergodic results. The main result in this chapter is the Birkhoff ergodic theorem, which in 1931 settled the question of convergence of dynamical systems. For an account of the parallel development in statistical physics and probability theory, see the interesting historical work by von Plato [26]. The account in this chapter is based on [5] and [9]. More results on ergodic behavior of random and non-random sequences can be found in [12], and for general stochastic aspects of dynamical systems, [21].

### 5.1 The basic Ergodic theorem in $L^2$

We met our first ergodic theorem in Section 2.7, Theorem 2:17, stating covariance conditions under which the time average  $T^{-1} \int_0^T x(t) dt$  tends, in quadratic mean and with probability one, to the expectation of the stationary process  $\{x(t), t \in \mathbb{R}\}$ . In the special case when  $x(t)$  is stationary with covariance function  $r(t) = Cov(x(s), x(s+t))$ , the quadratic mean convergence becomes particularly simple. If  $E(x(t)) = 0$  then

$$\frac{1}{T} \int_0^T r(t) dt \rightarrow 0 \quad \text{implies} \quad \frac{1}{T} \int_0^T x(t) dt \xrightarrow{q.m.} 0,$$

as  $T \rightarrow \infty$ . This was proven by elementary calculation in Theorem 2:17(a). Note that it follows from the statements in Theorem 4:7 in Section 4.3.3, that the relation is satisfied if and only if the spectral distribution function is continuous at the origin.

By means of the spectral representation  $x(t) = \int e^{i\omega t} dZ(\omega)$ , we can formulate a more precise theorem and see what happens if this sufficient condition does not hold. In Section 4.3.2 we stated an explicit expression (4.15) for the spectral process  $Z(\omega)$ . In fact, if  $\omega_1$  and  $\omega_2$  are continuity points of the spectral distribution function  $F(\omega)$ , then we have the parallel expressions

$$F(\omega_2) - F(\omega_1) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{-it} r(t) dt,$$

$$Z(\omega_2) - Z(\omega_1) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{-it} x(t) dt,$$

the latter convergence being in quadratic mean.

We repeat the statement from Theorem 4:7, that when the spectral distribution  $F$  is a step function, then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T r(t) e^{-i\omega_k t} dt = \Delta F_k,$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |r(t)|^2 dt = \sum_k (\Delta F_k)^2,$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) e^{-i\omega_k t} dt = \Delta Z_k.$$

## 5.2 Stationarity and transformations

### 5.2.1 Pseudo randomness and transformation of sample space

For strictly stationary processes one can obtain limit theorems of quite different character from those valid for processes which satisfy a covariance stationarity condition. These theorems also require much deeper conditions than simple covariance conditions. Remember, however, that a Gaussian (weakly) stationary process is also strictly stationary, so the general ergodicity properties of Gaussian processes can be inferred already from its covariance function. We start by giving all results for stationary sequences  $\{x_n; n \in \mathbb{Z}\}$ .

For a strictly stationary sequence the location of the origin is unessential for the stochastic properties, i.e.  $P((x_1, x_2, \dots) \in B) = P((x_{k+1}, x_{k+2}, \dots) \in B)$  for every Borel set  $B \in \mathcal{B}_\infty$ ; see Section 1.3.2. This also means that we can assume the sequence to be double ended, and to have started in the remote

past.<sup>1</sup> From now on in this chapter, by a stationary process we mean a process that is strictly stationary.

How do we – or nature – construct stationary sequences? Obviously, first we need something (call it "a game") that can go on forever, and second, we need a game where the rules remain the same forever.

**Example 5:1** (*The irrational modulo game*) A simple game, that can go on forever and has almost all interesting properties of a stationary process, is the adding of an irrational number. Take a random  $x_0$  with uniform distribution between 0 and 1, and let  $\theta$  be an irrational number. Define

$$x_{k+1} = x_k + \theta \pmod{1},$$

i.e.  $x_{k+1}$  is the fractional part of  $x_k + \theta$ . It is easy to see that  $x_k$  is a stationary sequence. We shall soon see why this game is more interesting with an irrational  $\theta$  than with a rational one. If computers could handle irrational numbers, this type of pseudorandom number generator would be even more useful in Monte Carlo simulations than it is.

**Example 5:2** One can define other stationary sequences by applying any time invariant rule to a stationary sequence  $x_n$ , e.g. with  $0 \leq x_0 \leq 1$ ,  $y_n = 1$  if  $x_{n+1} > x_n^2$  and  $y_n = 0$  otherwise. A more complicated rule is

$$y_n = x_n + \max_{k>0} e^{-k} x_{n+k}.$$

The well-known quadratic transformation,

$$x_{k+1} = cx_k(1 - x_k),$$

is an example of an interesting transformation giving rise to a stochastic sequence when  $x_0$  is random; in Exercise 5:5 you are asked to find its stationary distribution.

### 5.2.2 Strict stationarity and measure preserving transformations

Deterministic explicit rules, like  $x_{k+1} = x_k + \theta \pmod{1}$ , are just examples of transformations of a sample space which can produce stationary sequences for certain probability distributions. We need to define a general concept of a *measure preserving transformation*, which makes the resulting process stationary.

---

<sup>1</sup>More precisely, for every strictly stationary sequence  $\{x_n, n \in \mathbb{N}\}$  there exists a strictly stationary sequence  $\{\tilde{x}_n; n \in \mathbb{Z}\}$  such that  $\tilde{x}_n; n \geq 0$  have the same finite-dimensional distributions as  $x_n; n \geq 0$ ; prove this in Exercise 5:4.

**Definition 5:1** a) Consider a probability space  $(\Omega, \mathcal{F}, P)$ . A measurable transformation<sup>2</sup>  $T$  on  $(\Omega, \mathcal{F}, P)$  is called measure preserving if

$$P(T^{-1}A) = P(A) \quad \text{for all } A \in \mathcal{F}. \quad (5.1)$$

b) Given a measurable space  $(\Omega, \mathcal{F})$  and a measurable transformation  $T$  on  $(\Omega, \mathcal{F})$ , a probability measure  $P$  is called invariant if (5.1) holds.

In statistical physics or dynamical systems theory, a probability space is regarded as a model for the "universe", with the outcomes  $\omega$  representing all its different "states", e.g. the location and velocity of all its particles. The measure  $P$  defines the probabilities for the universe to be in such and such a state that certain events occur. A transformation  $T\omega$  is just a law of nature that changes the state of the universe; that a transformation is measure preserving means that the events occur with the same probability before and after the transformation.

### 5.2.2.1 From a measure preserving transformation to a stationary sequence

Every measure preserving transformation can generate a stationary sequence. To see how, take a random variable  $x(\omega)$  on  $(\Omega, \mathcal{F}, P)$ , i.e. a "measurement" on the state of the universe, for example its temperature. Then  $x(T\omega)$  is the same measurement taken after the transformation. Define the random sequence

$$x_1(\omega) = x(\omega), x_2(\omega) = x(T\omega), \dots, x_n(\omega) = x(T^{n-1}\omega), \dots$$

Since  $T$  is measure preserving, this sequence is strictly stationary: With  $B \in \mathcal{B}_\infty$  and

$$A = \{\omega; (x_1(\omega), x_2(\omega), \dots) \in B\} = \{\omega; (x(\omega), x(T\omega), \dots) \in B\},$$

$P(A) = P((x_1, x_2, \dots) \in B)$ . Further,

$$\begin{aligned} T^{-1}A &= \{\omega; T\omega \in A\} = \{\omega; (x(T\omega), x(T^2\omega), \dots) \in B\} \\ &= \{\omega; (x_2(\omega), x_3(\omega), \dots) \in B\}, \end{aligned}$$

and thus  $P(T^{-1}A) = P((x_2, x_3, \dots) \in B)$ . Since  $T$  is measure preserving,  $P(A) = P(T^{-1}A)$ , and hence  $(x_1, x_2, \dots)$  and  $(x_2, x_3, \dots)$  have the same distribution, that is,  $x_n$  is a stationary sequence.

---

<sup>2</sup>A measurable transformation  $T$  on a measurable space  $(\Omega, \mathcal{F})$  is a function defined on  $\Omega$  such that the inverse images under  $T$  of all sets in  $\mathcal{F}$  are again in  $\mathcal{F}$ ; that is  $T^{-1}A = \{\omega; T\omega \in A\} \in \mathcal{F}$  for all  $A \in \mathcal{F}$ .

### 5.2.2.2 From a stationary process to a measure preserving transformation

We have just seen how to construct a stationary sequence from a measure preserving transformation. Conversely, every stationary sequence generates a measure preserving transformation on  $\mathbb{R}^\infty$ , the space of realizations for the stationary sequence.

Take the probability space  $(\mathbb{R}^\infty, \mathcal{B}_\infty, P)$  with outcomes

$$\omega = (x_1, x_2, \dots),$$

and define the *shift transformation*  $T$  by

$$T\omega = (x_2, x_3, \dots).$$

As an example, take the set  $A = \{\omega; x_1 < x_2\}$  to be the outcomes for which the first coordinate is smaller than the second one. Then  $T^{-1}A = \{\omega; T\omega \in A\}$  is the set  $\{(x_1, x_2, \dots); (x_2, x_3, \dots) \in A\} = \{(x_1, x_2, \dots); x_2 < x_3\}$ , that is, the second coordinate is smaller than the third. The transformation just shifts the event criterion one step to the right.

Take the coordinate process  $\mathbf{x}(\omega) = \omega = (x_1, x_2, \dots)$ , for which the  $n^{\text{th}}$  variable in the sequence is equal to the  $n^{\text{th}}$  coordinate in the outcome, and assume that  $P$  is such that  $\mathbf{x}$  is a stationary sequence. Then the shift transformation  $T$  is measure preserving – the shifted sequence has the same distribution as the original one.

**Remark 5:1** *The property that a transformation is measure preserving, of course depends on the probability measure  $P$  and not only on the transformation itself. In probability theory, where the probability measure is often given a priori, it is natural to put the request on the transformation.*

*In the mathematical study of dynamical systems, one often starts with a transformation  $T$  and seeks a measure under which  $T$  is measure preserving. Such a measure is called invariant. Thus, invariant measures in the dynamical systems theory are equivalent to the strictly stationary processes in probability theory.*

## 5.3 The Ergodic theorem, transformation view

The classical ergodic theory for dynamical systems, deals with what happens in the long run when one observes some characteristic of the system, i.e. when one takes observation  $x(\omega)$ . The state of the system changes by the transformation  $T$ , and our interest is with the time average of the sequence of measurements  $x(\omega), x(T\omega), x(T^2\omega), \dots$ , taken for a fixed initial outcome  $\omega$ ,

$$\frac{1}{n} \sum_1^n x(T^{k-1}\omega)$$

as  $n \rightarrow \infty$ . Think of  $\Omega$  as all the possible states our universe can be in, and, to be concrete, think of  $x(\omega)$  as the temperature at one specific location. The universe changes states from day to day, and if  $\omega$  is the state of the universe today, and  $T\omega$  its state tomorrow, then  $\frac{1}{n} \sum_1^n x(T^{k-1}\omega)$  is the average temperature observed over an  $n$ -day period.

In this section we shall take the *transformation view* on ergodic theory, and prove the ergodic theorem in that vein. In the next section we shall do exactly the same thing, but take a probabilistic *strictly stationary process* view, and prove the ergodic theorem in terms of random variables and expectations.

### 5.3.1 Invariant sets and invariant random variables

To motivate the introduction of invariant sets and invariant random variables, we consider the possible limits of the average of a sequence of random variables,

$$S_n/n = \frac{1}{n} \sum_1^n x_k.$$

If  $S_n/n$  converges, as  $n \rightarrow \infty$ , what are the possible limits? Say, if  $S_n/n \rightarrow y$ , a random variable, then obviously  $x_n/n = S_n/n - S_{n-1}/n \rightarrow y - y = 0$ , and<sup>3</sup>

$$\begin{aligned} y &= \lim \frac{x_2 + x_3 + \dots + x_{n+1}}{n} + \lim \frac{x_1 - x_{n+1}}{n} \\ &= \lim \frac{x_2 + x_3 + \dots + x_{n+1}}{n} - \lim \frac{x_{n+1}}{n}. \end{aligned}$$

We see that the limit of  $S_n/n$  is the same for the sequence  $x_1, x_2, \dots$  as it would be for the shifted sequence  $x_2, x_3, \dots$ .

**Definition 5:2** Consider a probability space  $(\Omega, \mathcal{F}, P)$  and a measure preserving transformation  $T$  of  $\Omega$  onto itself.

(a) A random variable  $x$  on  $(\Omega, \mathcal{F}, P)$  is called invariant under  $T$  if  $x(\omega) = x(T\omega)$ , for almost every  $\omega \in \Omega$ .

(b) A set  $A \in \mathcal{F}$  is called invariant under  $T$  if  $T^{-1}A = A$ .

**Example 5:3** The limit of  $S_n/n$  is invariant (when it exists) under the shift transformation  $(x_1, x_2, \dots) \mapsto (x_2, x_3, \dots)$  of  $\mathbb{R}^\infty$ . The random variables  $y = \limsup x_n$  and  $\limsup S_n/n$  are always invariant.

---

<sup>3</sup>Show as an exercise, if you have not done so already, that  $x_n$  is a stationary sequence with  $E(|x_n|) < \infty$ , then  $P(x_n/n \rightarrow 0, \text{ as } n \rightarrow \infty) = 1$ ; Exercise 5:6.

**Example 5:4** Let  $x_n$  be a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 2/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix},$$

with starting distribution  $\mathbf{p}^{(0)} = (2p/5, 3p/5, 3q/5, 2q/5)$  is stationary for every  $p + q = 1$ . The sets  $A_1 = \{\omega = (x_1, x_2, \dots); x_k \in \{1, 2\}\}$  and  $A_2 = \{\omega = (x_1, x_2, \dots); x_k \in \{3, 4\}\}$  are both invariant under the shift transformation.

The proof of the following simple theorem is left to the reader.

**Theorem 5:1** (a) *The family of invariant sets,*

$$\mathcal{J} = \{\text{invariant sets } A \in \mathcal{F}\},$$

*is a  $\sigma$ -field.*

(b) *A random variable  $y$  is invariant if and only if it is measurable with respect to the family  $\mathcal{J}$  of invariant sets.*

### 5.3.2 Ergodicity

The fundamental property of ergodic systems  $\omega \mapsto T\omega$  with a stationary (or invariant) distribution  $P$ , is that the  $T^k\omega$ , with increasing  $k$ , visits every corner of the state space, exactly with the correct frequency as required by  $P$ . Another way of saying this is that the "histogram", counting the number of visits to any neighborhood of states, converges to a limiting "density", namely the density for the invariant distribution  $P$  over the state space. If we make a measurement  $x(\omega)$  on the system, then the expected value  $E(x)$  is the "ensemble average" with respect to the measure  $P$ ,

$$E(x) = \int_{\omega \in \Omega} x(\omega) dP(\omega),$$

and – here is the ergodicity – this is exactly the limit of the "time average"

$$\frac{1}{n} \sum_1^n x(T^{k-1}\omega).$$

In the Markov Example 5:4, if  $p$  and  $q$  are not 0, there is no possibility for the process to visit every state the correct number of times, since either it starts in the invariant set  $A_1$  and then it always takes the values 1, 2, or it starts in  $A_2$  and then it stays there forever and takes only values 3, 4. This is the key to the definition of ergodicity.

**Definition 5:3** A measure preserving transformation  $T$  on  $(\Omega, \mathcal{F}, P)$  is called ergodic if every invariant set  $A \in \mathcal{J}$  has either  $P(A) = 0$  or  $P(A) = 1$ , that is, all invariant sets are trivial. The term "metrically transitive" is sometimes used instead of "ergodic".

Here is a nice example of a transformation that can be ergodic.

**Example 5:5** The modulo game in Example 5:1, considered as a transformation on the unit interval,  $([0, 1], \mathcal{B}, \ell)$ , ( $\ell$  is the Lebesgue measure, i.e. the uniform distribution),

$$Tx = x + \theta \pmod{1},$$

is ergodic when  $\theta$  is irrational, and non-ergodic for rational  $\theta$ .

For  $\theta = m/n$  one has  $T^n x = x$  and every set of the form  $A \cup \{A + 1/n\} \cup \{A + 2/n\} \dots \cup \{A + (n-1)/n\}$  is invariant but can have probability  $\in (0, 1)$ .

To see what happens when  $\theta$  is irrational we need some further results for ergodic transformations.

**Theorem 5:2** Let  $T$  be a measure preserving transformation of  $(\Omega, \mathcal{F}, P)$ . Then  $T$  is ergodic if and only if every invariant random variable  $x$  is a.s. a constant. It is sufficient that every bounded invariant random variable is a.s. constant.

**Proof:** First assume that every bounded invariant variable is constant, and take an invariant set  $A$ . Its indicator function  $\chi_A(\omega) = 1$  if  $\omega \in A$  is then an invariant random variable, and by assumption it is an a.s. constant. This means that the sets where it is 0 and 1, respectively, have probability either 0 or 1, i.e.  $P(A) = 0$  or 1. Hence  $T$  is ergodic.

Conversely, take an ergodic  $T$ , and consider an arbitrary invariant random variable  $x$ . We shall show that  $x$  is a.s. constant. Define, for real  $x_0$ ,

$$A_{x_0} = \{\omega; x(\omega) \leq x_0\}.$$

Then  $T^{-1}A_{x_0} = \{\omega; T\omega \in A_{x_0}\} = \{\omega; x(T\omega) \leq x_0\} = A_{x_0}$ , since  $x$  is invariant. But then, by ergodicity,  $P(A_{x_0}) = 0$  or 1, depending on  $x_0$ , and it is easy to see that then there is an  $\tilde{x}_0$  such that  $P(x = \tilde{x}_0) = 1$ , hence  $x$  is constant.  $\square$

**Example 5:6** (*Irrational modulo game*) We can now show that  $Tx = x + \lambda \pmod{1}$  is ergodic if  $\lambda$  is irrational. Take any Borel-measurable function  $f(x)$  on  $[0, 1)$  with  $\int_0^1 f^2(x) dx < \infty$ . It can be expanded in a Fourier series

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i k x},$$

with convergence in quadratic mean for almost all  $x$  and  $\sum |c_k|^2 < \infty$ . Then  $y = f(\omega)$  is a random variable. We assume it is invariant, and prove that it is a.s. constant. But  $f$  invariant means that

$$f(x) = f(Tx) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i k x} \cdot e^{2\pi i k \lambda},$$

which implies  $c_k(1 - e^{2\pi i k \lambda}) = 0$  for all  $k$ . But  $e^{2\pi i k \lambda} \neq 1$  for  $k \neq 0$  when  $\lambda$  is irrational, and hence  $c_k = 0$  for all  $k \neq 0$ , which means that  $f(x) = c_0$ , constant. By Theorem 5:2 we conclude that  $T$  is ergodic.

The study of ergodic transformations is an important part of the dynamic systems and chaos theory; see [21].

### 5.3.3 The Birkhoff Ergodic theorem

In the introductory remarks, Section 5.3.1, we noted that any limit of the time average

$$\frac{1}{n} \sum_1^n x(T^{k-1}\omega)$$

is invariant. The limit could be a constant, and it could be a random variable, but it needs to be invariant. If it is constant, we need to find the value of the constant, and if it is a random variable, we want to find out as much as possible about its distribution.

In fact, for measure preserving transformations, the limit *always* exists, and is equal to the conditional expectation of  $x$  given the invariant sets. Since we have not dealt with this concept previously in this course, we state the basic properties of conditional expectations by giving its definition.

**Definition 5:4** *If  $y$  is a random variable on  $(\Omega, \mathcal{F}, P)$  with  $E(|y|) < \infty$ , and  $\mathcal{A}$  a sub- $\sigma$ -field of  $\mathcal{F}$ , then by the conditional expectation of  $y$  given  $\mathcal{A}$ ,  $E(y | \mathcal{A})$ , is meant any  $\mathcal{A}$ -measurable random variable  $u$  that satisfies*

$$\int_{\omega \in A} y(\omega) dP(\omega) = \int_{\omega \in A} u(\omega) dP(\omega),$$

*for all  $A \in \mathcal{A}$ . Note that the value of  $E(y | \mathcal{A}) = u$  is defined only almost surely, and that any  $\mathcal{A}$ -measurable variable which has the same integral as  $y$  when integrated over  $\mathcal{A}$ -sets, works equally well as the conditional expectation.*

*In particular, if  $\mathcal{A}$  only contains sets which have probability 0 or 1,  $E(y | \mathcal{A})$  is a.s. constant and equal to  $E(y)$ .*

In particular, we consider the conditional expectation  $E(x | \mathcal{J})$ , given the  $\sigma$ -field  $\mathcal{J}$  of sets which are invariant under a measure preserving transformation.

### 5.3.3.1 Time averages always converge

**Theorem 5:3** (*Birkhoff ergodicity theorem*) Let  $T$  be a measure preserving transformation on  $(\Omega, \mathcal{F}, P)$ . Then for any random variable  $x$  with  $E(|x|) < \infty$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_0^{n-1} x(T^k \omega) = E(x | \mathcal{J}), \text{ a.s.}$$

The proof is based on the following lemma, shown by Adriano M. Garsia in 1965.

**Lemma 5.1** Let  $T$  be a measure preserving transformation and  $x$  a random variable with  $E(|x|) < \infty$ . Define

$$S_k(\omega) = x(\omega) + \dots + x(T^{k-1}\omega), \quad \text{and} \quad M_n(\omega) = \max(0, S_1, S_2, \dots, S_n).$$

Then

$$\int_{\omega; M_n > 0} x(\omega) dP(\omega) \geq 0.$$

**Proof of lemma:** Consider  $S'_k = x(T\omega) + \dots + x(T^k\omega) = S_k - x(\omega) + x(T^k\omega)$ , and  $M_n(T\omega) = M'_n = \max(0, S'_1, \dots, S'_n)$ . For  $k = 1, \dots, n$ ,  $M'_n \geq S'_k$ , so

$$x + M'_n \geq x + S'_k = S_{k+1},$$

and (for  $k = 0$ )  $x + M'_n \geq S_1 (= x)$ . Hence

$$x \geq S_k - M'_n, \quad \text{for } k = 1, 2, \dots, n+1,$$

which implies

$$x \geq \max(S_1, \dots, S_n) - M'_n.$$

Thus (with  $M'_n = M_n(T\omega)$ ),

$$\int_{M_n > 0} x(\omega) dP(\omega) \geq \int_{M_n > 0} \{\max(S_1(\omega), \dots, S_n(\omega)) - M_n(T\omega)\} dP(\omega).$$

But on the set  $\{\omega; M_n(\omega) > 0\}$ , one has that  $M_n = \max(S_1, \dots, S_n)$ , and thus

$$\begin{aligned} \int_{M_n > 0} x(\omega) dP(\omega) &\geq \int_{M_n > 0} \{M_n(\omega) - M_n(T\omega)\} dP(\omega) \\ &\geq \int \{M_n(\omega) - M_n(T\omega)\} dP(\omega) = 0, \end{aligned}$$

since increasing the integration area does not change the integral of  $M_n(\omega)$  while it can only make the integral of  $M_n(T\omega)$  larger. Further,  $T$  is measure

preserving, i.e. shifting the variables one step does not change the distribution, nor the expectation.  $\square$

**Proof of theorem:** We first assume that  $E(x | \mathcal{J}) = 0$  and prove that the average converge to 0, a.s. For the general case consider  $x - E(x | \mathcal{J})$  and use that

$$E(x | \mathcal{J})(T\omega) = E(x | \mathcal{J})(\omega),$$

since  $E(x | \mathcal{J})$  is invariant by Theorem 5:1(b), page 139

We show that  $\bar{x} = \limsup S_n/n \leq 0$  and, similarly,  $\underline{x} = \liminf S_n/n \geq 0$ , giving  $\lim S_n/n = 0$ . Take an  $\epsilon > 0$  and denote  $D = \{\omega; \bar{x} > \epsilon\}$ : we shall show that  $P(D) = 0$ . Since, from Example 5:3,  $\bar{x}$  is an invariant random variable, also the event  $D$  is invariant. Define a new random variable,

$$x^*(\omega) = \begin{cases} x(\omega) - \epsilon & \text{if } \omega \in D, \\ 0 & \text{otherwise,} \end{cases}$$

and set  $S_n^*(\omega) = \sum_{k=1}^n x^*(T^{k-1}\omega)$ , with  $M_n^*$  defined from  $S_k^*$ . From Lemma 5.1 we know

$$\int_{M_n^* > 0} x^*(\omega) dP(\omega) \geq 0. \quad (5.2)$$

We now only have to replace this inequality by an inequality for a similar integral over the set  $D$  to be finished. The sets

$$F_n = \{M_n^* > 0\} = \left\{ \max_{1 \leq k \leq n} S_k^* > 0 \right\}$$

increase towards the set

$$F = \left\{ \sup_{k \geq 1} S_k^* > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \epsilon \right\} \cap D.$$

But since  $\sup_{k \geq 1} S_k/k \geq \limsup S_k/k = \bar{x}$ , we have that  $F = D$ . In order to take the limit of (5.2) we must be sure the expectations are finite, i.e.  $E(|x^*|) \leq E(|x|) + \epsilon$ , and bounded convergence gives

$$0 \leq \int_{M_n^* > 0} x^*(\omega) dP(\omega) \rightarrow \int_D x^*(\omega) dP(\omega). \quad (5.3)$$

Here the right hand side is

$$\begin{aligned} \int_D x^*(\omega) dP(\omega) &= \int_D x(\omega) dP(\omega) - \epsilon P(D) \\ &= \int_D E(x | \mathcal{J}) dP - \epsilon P(D) = -\epsilon P(D), \end{aligned}$$

since  $\int_D E(x | \mathcal{J}) dP = 0$ , by assumption. Together with (5.3) this implies  $P(D) = 0$ , and hence  $\bar{x} \leq \epsilon$ . But  $\epsilon > 0$  was arbitrary so  $\bar{x} \leq 0$ , a.s. The same chain of arguments leads to  $\underline{x} \geq 0$ , and hence  $\limsup S_n/n \leq 0 \leq \liminf S_n/n$ . The limit therefore exists and is 0, which was to be shown.  $\square$

### 5.3.3.2 Time averages of ergodic transformations go to the mean

It is now a simple corollary that time averages for ergodic transformations converge to the expected value.

**Corollary 5.1** *If  $T$  is a measure preserving ergodic transformation on the probability space  $(\Omega, \mathcal{F}, P)$ , then for any random variable  $x$  with  $E(|x|) < \infty$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_0^{n-1} x(T^k \omega) = E(x), \text{ a.s.}$$

**Proof:** When  $T$  is ergodic every invariant set has probability 0 or 1 and therefore the conditional expectation is constant,  $E(x | \mathcal{J}) = E(x)$ , a.s.  $\square$

**Remark 5:2** *If  $x$  is non-negative with  $E(x) = \infty$  then  $S_n/n \rightarrow \infty$  if  $T$  is ergodic. Show that as an exercise; Exercise 5:10.*

### 5.3.3.3 Ergodic non-random walks

One can regard a measure preserving transformation  $T$  as a non-random walk  $\omega, T\omega, T^2\omega, \dots$ , over the sample space. In the beginning of this section we interpreted the ergodic statements as convergence results for the number of visits to any neighborhood of a fixed outcome. This can now be made precise. Take a set  $A \in \mathcal{F}$  and consider its indicator function  $\chi_A(\omega)$ . The ergodic theorem says, that if  $T$  is ergodic,

$$\frac{1}{n} \sum_0^{n-1} \chi_A(T^k \omega) \xrightarrow{\text{a.s.}} P(A).$$

**Example 5:7** In the irrational modulo game,  $Tx = x + \theta \pmod{1}$ , the relative number of points falling in an interval  $[a, b)$  converges to the length of the interval if  $\theta$  is irrational. Thus the number of points becomes asymptotically equidistributed over  $[0, 1)$ . This is the weak Weyl's equidistribution theorem.

## 5.4 The Ergodic theorem, process view

It is easy to formulate the convergence theorem for transformations in terms of time averages of stationary sequences. First we need to define invariant events and ergodicity in  $(\Omega, \mathcal{F}, P)$ .

**Definition 5:5** *(Reformulation of Definition 5:2) Let  $\{x_n\}$  be a stationary sequence. An event  $A \in \mathcal{F}$  is called invariant for  $\{x_n\}$ , if there exists a  $B \in \mathcal{B}_\infty$  such that for any  $n \geq 1$ ,*

$$A = \{(x_n, x_{n+1}, \dots) \in B\}.$$

The sequence is called ergodic if every invariant set has probability 0 or 1.

A random variable  $z$  is called invariant for  $\{x_n\}$ , if it is a function of  $x_1, x_2, \dots$  and remains unchanged under the shift transformation, i.e. if there exists a random variable  $\phi$  on  $(\mathbb{R}^\infty, \mathcal{B}_\infty)$  such that  $z = \phi(x_n, x_{n+1}, \dots)$  for all  $n \geq 1$ .

From the correspondence between transformations of the sample space and stationary sequences, it is easy to formulate an ergodicity theorem for a stationary sequence  $\{x_n, n \in \mathbb{Z}\}$ .

**Theorem 5:4** (a) If  $\{x_n, n \in \mathbb{Z}\}$  is a stationary sequence with  $E(|x_1|) < \infty$ , and  $\mathcal{J}$  denotes the  $\sigma$ -field of invariant sets,

$$\frac{1}{n} \sum_1^n x_k \xrightarrow{\text{a.s.}} E(x_1 | \mathcal{J}).$$

(b) If  $\{x_n, n \in \mathbb{Z}\}$  is stationary and ergodic, then

$$\frac{1}{n} \sum_1^n x_k \xrightarrow{\text{a.s.}} E(x_1).$$

(c) If  $\{x_n, n \in \mathbb{Z}\}$  is stationary and ergodic, and  $\phi(x_1, x_2, \dots)$  is measurable on  $(\mathbb{R}^\infty, \mathcal{B}_\infty)$ , then the process

$$y_n = \phi(x_n, x_{n+1}, \dots)$$

is stationary and ergodic. The special case that  $\phi(x_n, x_{n+1}, \dots) = \phi(x_n)$  only depends on  $x_n$  should be noted in particular.

**Proof:** It is easy to reformulate the results of Theorem 5:3 to yield the theorem. However, we shall give the direct proof once more, but use the process formulation, in order to get a slightly better understanding in probabilistic terms. We give a parallel proof of part (a), and leave the rest to the reader.

(a) First the counterpart of Lemma 5.1: From the sequence  $\{x_n, n \in \mathbb{Z}\}$ , define  $S_k = \sum_{j=1}^k x_j$  and  $M_n = \max(0, S_1, S_2, \dots, S_n)$ . We prove that

$$E(x_1 | M_n > 0) \geq 0. \quad (5.4)$$

To this end, write  $S'_k = \sum_{j=2}^{k+1} x_j = S_k - x_1 + x_{n+1}$ , and define the corresponding maximum,  $M'_n = \max(0, S'_1, S'_2, \dots, S'_n)$ , and note that, for  $k = 1, \dots, n$ ,

$$x_1 + M'_n \geq x_1 + S'_k = S_{k+1}.$$

Since  $M'_n \geq 0$  we also have  $x_1 + M'_n \geq S_1 (= x_1)$ , so

$$x_1 \geq \max(S_1, \dots, S_n) - M'_n.$$

Now, when  $M_n > 0$  we can replace  $\max(S_1, \dots, S_n)$  by  $M_n$ , so taking the conditional expectation, given  $M_n > 0$ , we get

$$\begin{aligned} E(x_1 | M_n > 0) &\geq E(\max(S_1, \dots, S_n) - M'_n | M_n > 0) \\ &= E(M_n - M'_n | M_n > 0). \end{aligned}$$

Further, since  $M'_n \geq 0$ , one easily argues that

$$E(M_n - M'_n | M_n > 0) \geq E(M_n - M'_n) / P(M_n > 0),$$

which is 0, since  $M_n$  and  $M'_n$  have the same expectation. This proves (5.4).

We continue with the rest of the proof of part (a). Suppose  $E(x_1 | \mathcal{J}) = 0$  and consider the invariant random variables  $\bar{x} = \limsup S_n/n$ . Take an  $\epsilon > 0$  and introduce the invariant event  $D = \{\bar{x} > \epsilon\}$ . Then, by the assumption,

$$E(x_1 | \bar{x} > \epsilon) = 0, \tag{5.5}$$

a fact which is basic in the proof. We intend to prove that  $P(D) = 0$  for every  $\epsilon > 0$ , thereby showing that  $\bar{x} \leq 0$ .

Similarly,  $\underline{x} = \liminf S_n/n$  can be shown to be non-negative, and hence  $\bar{x} = \underline{x} = 0$ .

However, before we prove that  $P(D) = P(\bar{x} > \epsilon) = 0$ , we need to discuss the meaning of (5.5). A conditional expectation is defined as a random variable, measurable with respect to the conditioning  $\sigma$ -field, in this case  $\mathcal{J}$ . In (5.5) we conditioned on one of the events  $D \in \mathcal{J}$  and that is fine if  $P(D) > 0$ , but if  $P(D) = 0$ , the claim (5.5) makes no sense. The conditional expectation given an event of probability 0 can be given any value we like, since the only requirement on the conditional expectation is that it should give a correct value when integrated over a  $\mathcal{J}$ -event. If that event has probability 0 the integral is 0 regardless of how the expectation is defined. We return to this at the end of the proof.

From  $(x_1, x_2, \dots)$ , define a new sequence of variables

$$x_k^* = \begin{cases} x_k - \epsilon & \text{if } \bar{x} > \epsilon, \\ 0 & \text{otherwise,} \end{cases}$$

and define  $S_k^* = \sum_{j=1}^k x_j^*$  and  $M_n^* = \max(0, S_1^*, S_2^*, \dots, S_n^*)$ , in analogy with  $S_k$  and  $M_n$ . The sequence  $\{x_k^*\}$  is stationary, since  $\bar{x}$  is an invariant random variable, so we can apply (5.4) to get

$$E(x_1^* | M_n^* > 0) \geq 0. \tag{5.6}$$

On the other hand, from the definition of  $x_k^*$  we have

$$E(x_1^* | \bar{x} > \epsilon) = E(x_1 | \bar{x} > \epsilon) - \epsilon = -\epsilon < 0, \tag{5.7}$$

since  $E(x_1 | \bar{x} > \epsilon) = 0$  by the assumption. These two inequalities go in opposite directions, and in fact they will turn out to be in conflict, unless  $P(D) = 0$ , proving the assertion.

So we would like to have a relation between the events  $D = \{\bar{x} > \epsilon\}$  and  $F_n = \{M_n^* > 0\} = \{\max_{1 \leq k \leq n} S_k^* > 0\}$ , and we see that as  $n$  increases, then  $F_n$  increases to

$$F = \left\{ \sup_{k \geq 1} S_k^* > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \epsilon \right\} \cap \{\bar{x} > \epsilon\}.$$

But  $\bar{x} = \limsup S_k/k \leq \sup_{k \geq 1} S_k/k$ , so the right hand side is just  $D = \{\bar{x} > \epsilon\}$ ;  $F_n \uparrow D$ . This implies (here  $E(|x_1^*|) \leq E(|x_1|) + \epsilon < \infty$  is needed),

$$0 \leq \lim_{n \rightarrow \infty} E(x_1^* | F_n) = E(x_1^* | D) = -\epsilon < 0,$$

which obviously is impossible.

Where did it go wrong, and where is the contradiction? By definition, for the outcomes where  $\bar{x} > \epsilon$ , the variables  $x_k^*$  ARE equal to  $x_k - \epsilon$ , and  $S_k^*/k = S_k/k - \epsilon$ . But  $F_n \uparrow D$  does not imply  $\lim_{n \rightarrow \infty} E(x_1^* | F_n) = \lim_{n \rightarrow \infty} E(x_1^* | D)$ , since these expressions are not well defined. If  $P(D) > 0$  our reasoning makes sense and lead to a contradiction, if  $P(D) = 0$  we have argued with undefined quantities in (5.6) and (5.7). The reader who wants to be on safest possible grounds should return to the formulation and proof of Theorem 5:3.  $\square$

#### 5.4.0.4 Ergodic stationary processes

For continuous time processes  $\{x(t), t \in \mathbb{R}\}$ , one defines the shift transformation  $U_\tau$  by

$$(U_\tau x)(t) = x(t + \tau).$$

If  $x(t)$  is stationary,  $U_\tau$  is measure preserving. For a set of functions  $B$ , the shifted set  $U_\tau B$  is the set of functions  $U_\tau x$  for  $x \in B$ . A Borel set  $B \in \mathcal{B}_{\mathbb{R}}$  is called a.s. invariant, if  $B$  and  $U_\tau B$  differ by, at most, sets of  $P$ -measure 0. Let  $\mathcal{J}$  denote the  $\sigma$ -field of invariant sets. The process  $\{x(t), t \in \mathbb{R}\}$  is called ergodic if all invariant sets have probability 0 or 1.

**Theorem 5:5** (a) For any stationary process  $\{x(t), t \in \mathbb{R}\}$  with  $E(|x(t)|) < \infty$  and integrable sample paths, as  $T \rightarrow \infty$ ,

$$\frac{1}{T} \int_0^T x(t) dt \xrightarrow{a.s.} E(x(0) | \mathcal{J}).$$

(b) If further  $\{x(t), t \in \mathbb{R}\}$  is ergodic, then

$$\frac{1}{T} \int_0^T x(t) dt \xrightarrow{a.s.} E(x(0)).$$

**Proof:** One first has to show that  $x_n = \int_{n-1}^n x(t) dt$  is a stationary sequence, and use the ergodic theorem to get convergence for integer  $n$ ,

$$\frac{1}{n} \int_0^n x(t) dt \xrightarrow{a.s.} E\left(\int_0^1 x(t) dt \mid \mathcal{J}\right),$$

as  $n \rightarrow \infty$ . By invariance and stationarity the last expectation is equal to  $E(x(0) \mid \mathcal{J})$ . Finally, with  $n = [T]$ ,

$$\frac{1}{T} \int_0^T x(t) dt = \frac{n}{T} \frac{1}{n} \int_0^n x(t) dt + \frac{1}{T} \int_n^T x(t) dt.$$

The first term has the same limit as  $\frac{1}{n} \int_0^n x(t) dt$ , while the second term is bounded by

$$\frac{1}{T} \int_n^{n+1} |x(t)| dt.$$

But also  $|x(t)|$  is a stationary process, to which we can apply Theorem 5:4(a), getting convergence, and hence can conclude that the last term tends to 0. Thus we obtain the desired limit.  $\square$

## 5.5 Ergodic Gaussian sequences and processes

We shall now give simple conditions for ergodicity for Gaussian stationary processes, characterized by their covariance function  $r(t)$  in continuous or discrete time.

**Theorem 5:6** *Let  $\{x(t), t \in \mathbb{R}\}$  be stationary and Gaussian with  $E(x(t)) = 0$  and  $V(x(t)) = 1$ , and let its covariance function be  $r(t)$ . Then  $x(t)$  is ergodic if and only if its spectral distribution function  $F(\omega)$  is continuous everywhere. If the spectral distribution has a density  $f(\omega)$ ,  $F(\omega) = \int_{-\infty}^{\omega} f(l) dl$ , the  $F(\omega)$  is obviously continuous, but that is by no means necessary. It suffices that  $F(\omega)$  is a continuous function.*

**Proof of "only if" part:** If  $x(t)$  is ergodic, so is  $x^2(t)$ , and therefore the time average of  $x^2(t)$  tends to  $E(x^2(0)) = 1$ ,

$$\frac{S_T}{T} = \frac{1}{T} \int_0^T x^2(t) dt \xrightarrow{a.s.} 1,$$

as  $T \rightarrow \infty$ . It is a property of the Gaussian distribution that  $E((S_T/T)^4) \leq K$  is bounded for large  $T$ , and therefore the almost sure convergent  $S_T/T$  also converges in quadratic mean, i.e.  $E((S_T/T - 1)^2) \rightarrow 0$ . But this expectation can be calculated. Since, for a standard Gaussian process,  $E(x(s)^2 x(t)^2) =$

$1 + 2r(t - s)^2$ , one gets,

$$\begin{aligned} E((S_T/T - 1)^2) &= \frac{1}{T^2} E \left( \int_0^T \int_0^T x(s)^2 x(t)^2 ds dt \right) - 1 \\ &= \frac{2}{T^2} \int_0^T \int_0^T r^2(t - s) ds dt = \frac{4}{T^2} \int_0^T t \cdot \left\{ \frac{1}{t} \int_0^t r^2(s) ds \right\} dt. \end{aligned} \quad (5.8)$$

But according to Theorem 4:7, page 94, relation (4.20),  $\frac{1}{t} \int_0^t r^2(s) ds$  tends to the sum of squares of all jumps of the spectral distribution function  $F(\omega)$ ,  $\sum(\Delta F_k)^2$ . Hence, if this sum is strictly positive, the right hand side in (5.8) has a positive limit, which contradicts what we proved above, and we have concluded the "only if" part of the theorem.  $\square$

The "if" part is more difficult, and we can prove it here only under the additional condition that the process has a spectral density, i.e. the spectral distribution is  $F(\omega) = \int_{-\infty}^{\omega} f(x) dx$ , because then

$$r(t) = \int e^{i\omega t} f(\omega) d\omega \rightarrow 0,$$

as  $t \rightarrow \infty$ , by Riemann-Lebesgue's Lemma. (The full statement and proof can be found in [24, 15].) So what we have to show is that if  $r(t) \rightarrow 0$ , then  $x(t)$  is ergodic. Since this is worth to remember, we formulate it as a lemma.

**Lemma 5.2** *A Gaussian stationary process is ergodic if its covariance function  $r(t) \rightarrow 0$  as  $t \rightarrow \infty$ .*

**Proof of lemma, and the "if" part of theorem:** We show that if  $r(t) \rightarrow 0$ , then every invariant set has probability 0 or 1. Let  $S$  be an a.s. invariant set for the  $x(t)$ -process, i.e. the translated event  $S_\tau$  differs from  $S$  by an event of probability zero. But every event in  $\mathcal{F}$  can be approximated arbitrarily well by a finite-dimensional event,  $B$ , depending only on  $x(t)$  for a finite number of time points  $t_k, k = 1, \dots, n$ ; cf. Section 1.3.3. From stationarity, also  $S_\tau$  can be approximated by the translated event  $B_\tau = U_\tau B$ , with the same error, and combining  $S$  with  $S_\tau$  can at most double the error. Thus, we have

$$\begin{aligned} |P(S) - P(B)| &< \epsilon, \\ |P(S \cap S_\tau) - P(B \cap B_\tau)| &< 2\epsilon. \end{aligned}$$

Here  $P(S \cap S_\tau) = P(S)$  since  $S$  is invariant, so  $P(S)$  can be approximated arbitrarily well by both  $P(B)$  and by  $P(B \cap B_\tau)$ .

But  $B$  depends on  $x(t_i), i = 1, \dots, n$ , while  $B_\tau$  is defined from  $x(\tau + t_j), j = 1, \dots, n$ , and they are multivariate normal with covariances

$$\text{Cov}(x(t_i), x(t_j + \tau)) = r(\tau + t_j - t_i) \rightarrow 0$$

as  $\tau \rightarrow \infty$ . Thus these two groups of random variables become asymptotically independent,

$$P(B \cap B_\tau) - P(B) \cdot P(B_\tau) \rightarrow 0,$$

and by stationarity,  $P(B \cap B_\tau) \rightarrow P(B)^2$ . Thus both  $P(B)$  and  $P(B)^2$  approximate  $P(S)$ , and we conclude that  $P(S) = P(S)^2$ . This is possible only if  $P(S)$  is either 0 or 1, i.e.  $x(t)$  is ergodic.  $\square$

## 5.6 Mixing and asymptotic independence

How much of the future development of a stochastic process is pre-determined from what has already happened, and how much information about the future is there in a piece of observation of a stochastic process and how closely dependent are disjoint segments of the process?

In this section we will briefly mention some criteria on a stationary process that guarantee perfect predictability and un-predictability, respectively. A differentiable process can be predicted locally by means of a Taylor expansion. The linear prediction theory gives the Cramér-Wold decomposition (Theorem 5:7) into a singular component that can be predicted linearly without error and one regular component for which the predictable part tends to zero with increasing prediction horizon. The spectral representation in Chapter 4 relates the spectrum to the number of harmonic components which are needed to build a stationary process. The ergodic theorem in Chapter 5 touched upon the problem of asymptotic independence; for a Gaussian process with (absolutely) continuous spectrum and asymptotically vanishing covariance function, values far apart are asymptotically independent and a "law of large numbers" holds. In this section we shall try to relate these scattered results to each other. Some comments are for Gaussian processes only, while others are of general nature. A new concept, *mixing*, will also be introduced.

### 5.6.1 Singularity and regularity

When predicting from  $x(s), s \leq t$ , a question of both practical and theoretical (or perhaps philosophical) interest is *from where* does the information about  $x(t+h)$  originate, and how much new information is added to  $\mathcal{H}(x, t) = \mathcal{S}(x(s); s \leq t)$ , with increasing  $t$ .

When  $t \rightarrow -\infty$ , obviously

$$\mathcal{H}(x, t) \downarrow \mathcal{H}(x, -\infty) = \cap_{t \leq t_0} \mathcal{H}(x, t),$$

and

$$\mathcal{H}(x, -\infty) \subseteq \mathcal{H}(x, t) \subseteq \mathcal{H}(x).$$

The subspace  $\mathcal{H}(x, t)$  is the space of random variables that can be obtained as limits of linear combinations of variables  $x(t_k)$  with  $t_k \leq t$ , and  $\mathcal{H}(x, -\infty)$  is

what can be obtained from old variables, regardless of how old they may be. It can be called *the infinitely remote past*, or the *the primordial randomness*.<sup>4</sup>

Two extremes may occur, depending on the size of  $\mathcal{H}(x, -\infty)$ :

- if  $\mathcal{H}(x, -\infty) = \mathcal{H}(x)$ , then  $\{x(t), t \in \mathbb{R}\}$  is purely deterministic, or singular,
- if  $\mathcal{H}(x, -\infty) = \mathbf{0}$ , then  $\{x(t), t \in \mathbb{R}\}$  is purely non-deterministic, or regular.

A process is deterministic if all information about the future that can be obtained from the past at time  $t$ ,  $x(s), s \leq t$ , can be obtained already from  $x(s), s \leq \tau < t$ , arbitrarily far back. An example of this is the band-limited white noise Gaussian process, which we studied in Chapter 4. Such a process is infinitely differentiable – this follows from the general rules for differentiability – and the entire sample functions can be reconstructed from the values in an arbitrarily small interval located anywhere on the time axis. A summary of facts pertaining reconstruction and prediction is given later in these notes; see Section 5.6 on general results on asymptotic independence and its opposite, complete dependence.

The following theorem was proved by Cramér (1962) in the general parameter case; the discrete stationary case was given by Wold (1954).

**Theorem 5:7** (*The Cramér-Wold decomposition*) *Every stochastic process*

$$\{x(t), t \in \mathbb{R}\}$$

*with  $E(|x(t)|^2) < \infty$ , is the sum of two uncorrelated processes,*

$$x(t) = y(t) + z(t),$$

*where  $\{y(t), t \in \mathbb{R}\}$  is regular (purely non-deterministic) and  $\{z(t), t \in \mathbb{R}\}$  is singular (deterministic).*

**Proof:** Construct  $\mathcal{H}(x, t)$  and  $\mathcal{H}(x, -\infty) = \lim_{t \downarrow -\infty} \mathcal{H}(x, t)$ , and define

$$z(t) = P_{-\infty}(x(t)) = \text{the projection of } x(t) \text{ on } \mathcal{H}(x, -\infty),$$

$$y(t) = x(t) - z(t).$$

To prove the theorem, we have to show that

1.  $\mathcal{H}(z, -\infty) = \mathcal{H}(z, t)$ , and  $z(t)$  is deterministic,
2.  $\mathcal{H}(y, -\infty) = \mathbf{0}$ , and  $y(t)$  is non-deterministic,
3.  $\mathcal{H}(y) \perp \mathcal{H}(z)$ , and  $y(s)$  and  $z(t)$  are uncorrelated.

---

<sup>4</sup>Nowadays called "the primordial soup".

Number (3) follows from the projection properties, since the residual  $y(s) = x(s) - P_{-\infty}(x(s))$  is uncorrelated with every element in  $\mathcal{H}(x, -\infty)$ , i.e.  $y(s) \perp \mathcal{H}(x, -\infty)$ . Since  $z(t) \in \mathcal{H}(x, -\infty)$ , number (3) follows.

Further,  $\mathcal{H}(y, t) \subseteq \mathcal{H}(x, t)$  and  $\mathcal{H}(y, t) \perp \mathcal{H}(x, -\infty)$ . Therefore  $\mathcal{H}(y, -\infty)$  is equal to  $\mathbf{0}$ , because if  $y$  is an element of  $\mathcal{H}(y, -\infty)$  then both  $y \in \mathcal{H}(y, t) \subseteq \mathcal{H}(x, t)$  for all  $t$ , i.e.  $y \in \mathcal{H}(x, -\infty)$ , and at the same time  $y \perp \mathcal{H}(x, -\infty)$ . The only element that is both in  $\mathcal{H}(x, -\infty)$  and is orthogonal to  $\mathcal{H}(x, -\infty)$  is the zero element, showing (2).

Finally,  $\mathcal{H}(z, t) = \mathcal{H}(x, -\infty)$  for every  $t$ . To see this, note that

$$\mathcal{H}(x, t) \subseteq \mathcal{H}(y, t) \oplus \mathcal{H}(z, t)$$

for all  $t$ , and therefore also

$$\mathcal{H}(x, -\infty) \subseteq \mathcal{H}(y, -\infty) \oplus \mathcal{H}(z, t).$$

Since  $\mathcal{H}(y, -\infty) = \mathbf{0}$ ,

$$\mathcal{H}(x, -\infty) \subseteq \mathcal{H}(z, t) \subseteq \mathcal{H}(x, -\infty).$$

Thus  $\mathcal{H}(x, -\infty) = \mathcal{H}(z, t)$ , and (1) is proved.  $\square$

As the band-limited white noise example shows, there are natural deterministic processes. Other common process models are regular. Examples of processes combining the two properties seem to be rather artificial.

**Example 5:8** An AR(1)-process with an added component,

$$x(t) = ax(t-1) + e + e(t),$$

with uncorrelated  $e$  and  $e(t)$ -variables, can be decomposed into

$$y(t) = \sum_{k=0}^{\infty} a^k e(t-k),$$

which is regular, and

$$z(t) = \frac{1}{1-a}e,$$

which is singular. The common ARMA-process is regular.

## 5.6.2 Asymptotic independence, regularity and singularity

As we saw in the proof of Lemma 5.2, if  $r(t) \rightarrow 0$  as  $t \rightarrow \infty$ , then in a Gaussian process, finitely many variables taken sufficiently far apart are almost independent. But this does definitely not mean that the process observed in an entire interval,  $x(s), s \in I$  is asymptotically independent of  $x(s+t), s \in I$ . These two

segments can be completely dependent of each other in a deterministic way. For example, the realizations can be infinitely differentiable and part of an analytic function that can be reconstructed from its derivatives in an arbitrarily small interval. It was shown by Belyaev [3] that if the covariance function of a separable process is an entire function, i.e. analytic in the entire complex plane, then the sample functions are a.s. also entire functions, which can be expressed as a convergent power series

$$x(t) = \sum_{k=0}^{\infty} x^{(k)}(0) \frac{t^k}{k!}.$$

Examples of such covariance functions are  $r(t) = e^{-t^2/2}$  and  $r(t) = \frac{\sin t}{t}$ . Yaglom [38, Ch. 8] contains a readable account of prediction in this case; [18, 11] give more mathematical details.

### 5.6.2.1 Regularity, singularity and the spectrum

The Cramér-Wold decomposition deals with the problem of predicting future values by means of linear combinations of past observations. A singular (or purely deterministic) process can be perfectly predicted linearly from old values. In a regular process the predictable part tends to zero with increasing prediction horizon. Simple conditions for singularity/regularity can be formulated in terms of the spectral distribution function  $F(\omega)$ .

Let  $f(\omega) = \frac{d}{d\omega} F(\omega)$  be the derivative of the bounded and non-decreasing function  $F(\omega)$ . For almost all  $\omega$  this derivative exists and is non-negative, and its integral is bounded,  $\int_{-\infty}^{\infty} f(\omega) d\omega \leq F(\infty) - F(-\infty)$ . Write

$$F^{(ac)}(\omega) = \int_{-\infty}^{\omega} f(x) dx \leq F(\omega).$$

The spectrum is called absolutely continuous with spectral density  $f(\omega)$  if

$$F(\omega) = F^{(ac)}(\omega).$$

In general,  $F^{(ac)}(\omega)$  need not be equal to  $F(\omega)$ . In particular, this is of course the case when the spectrum has jumps  $\Delta F_k$  at frequencies  $\omega_k$ . Write

$$F^{(d)}(\omega) = \sum_{\omega_k \leq \omega} \Delta F_k,$$

so the spectrum is discrete if  $F(\omega) = F^{(d)}(\omega)$ . The part of the spectrum which is neither absolutely continuous nor discrete is called the *singular* part:

$$F^{(s)}(\omega) = F(\omega) - F^{(ac)}(\omega) - F^{(d)}(\omega).$$

Note that both  $F^{(d)}(\omega)$  and  $F^{(s)}(\omega)$  are bounded non-decreasing functions, differentiable almost everywhere, with zero derivative. The question of singularity or regularity of the process  $\{x(t), t \in \mathbb{R}\}$  depends on the behavior of the spectrum for large  $|\omega|$ .

### 5.6.2.2 Conditions for stationary sequences

Since  $\int_{-\pi}^{\pi} f(\omega) d\omega < \infty$  and  $-\infty \leq \log f(\omega) \leq f(\omega) = \frac{d}{d\omega} F(\omega)$ , the integral

$$P = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega \quad (5.9)$$

is either finite or equal to  $-\infty$ .

**Theorem 5:8** *For a stationary sequence  $x_t, t \in \mathbb{Z}$  the following cases can occur.*

- a) *If  $P = -\infty$ , then  $x_t$  is singular.*
- b) *If  $P > -\infty$  and the spectrum is absolutely continuous with  $f(\omega) > 0$  for almost all  $\omega$ , then  $x_t$  is regular.*
- c) *If  $P > -\infty$ , but  $F(\omega)$  is either discontinuous, or is continuous with non-vanishing singular part,  $F^{(s)}(\omega) \neq 0$ , then  $x_t$  is neither singular nor regular.*

This theorem is quite satisfying, and it is worth making some comments on its implications. First, if the spectrum is *discrete* with a finite number of jumps, then  $f(\omega) = 0$  for almost all  $\omega$  and  $P = -\infty$ , so the process is singular. As seen from the spectral representation (4.18), the process then depends only on a finite number of random quantities which can be recovered from a finite number of observed values.

If the spectrum is *absolutely continuous* with density  $f(\omega)$ , singularity and regularity depends on whether  $f(\omega)$  comes close to 0 or not. For example, if  $f(\omega)$  vanishes in an interval, then  $P = -\infty$  and  $x(t)$  is singular.<sup>5</sup> If  $f(\omega) \geq c > 0$  for  $-\pi < \omega \leq \pi$ , then the integral is finite and the process regular.

A stationary sequence  $x(t)$  is regular, if and only if it can be represented as a *one-sided moving average*

$$x_t = \sum_{k=-\infty}^t h_{t-k} y_k,$$

with uncorrelated  $y_k$ ; cf. Theorem 4:8 which also implies that it must have a spectral density.

A sequence that is neither singular nor regular can be represented as sum of two uncorrelated sequences,

$$x_t = x_t^{(s)} + x_t^{(r)} = x_t^{(s)} + \sum_{k=-\infty}^t h_{t-k} y_k.$$

---

<sup>5</sup>Singularity also occurs if  $f(\omega) = 0$  at a single point  $\omega_0$  and is very close to 0 nearby, such as when  $f(\omega) \sim \exp(-\frac{1}{(\omega-\omega_0)^2})$  when  $\omega \rightarrow \omega_0$ .

The regular part  $x_t^{(r)}$  has absolutely continuous spectrum,

$$F^{(ac)}(\omega) = \int_{-\pi}^{\omega} f(x) dx,$$

while the singular part  $x_t^{(s)}$  has spectral distribution  $F^{(d)}(\omega) + F^{(s)}(\omega)$ .

It is also possible to express the prediction error in terms of the integral

$$P = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega.$$

In fact, the one step ahead prediction error

$$\sigma_0^2 = \inf_{h_0, h_1, \dots} E \left( \left| x_{t+1} - \sum_{k=0}^{\infty} h_k x_{t-k} \right|^2 \right) = 2\pi \exp(P).$$

### 5.6.2.3 Conditions for stationary processes

Conditions for regularity and singularity for stationary processes  $x(t)$  with continuous parameter, can be expressed in terms of the integral

$$Q = \int_{-\infty}^{\infty} \frac{\log f(\omega)}{1 + \omega^2} d\omega,$$

where as before,  $f(\omega) = \frac{d}{d\omega} F(\omega)$  is the a.s. existing derivative of the spectral distribution function.

**Theorem 5:9** For a stationary process  $\{x(t), t \in \mathbb{R}\}$ , one has that

- a) if  $Q = -\infty$ , then  $x(t)$  is singular,
- b) if  $Q > -\infty$  and the spectrum is absolutely continuous then  $x(t)$  is regular.

The decomposition of  $x(t)$  into one singular component which can be predicted, and one regular component which is a moving average, is analogous to the discrete time case,

$$x(t) = x^{(s)}(t) + \int_{u=-\infty}^t h(t-u) d\zeta(u),$$

where  $\{\zeta(t), t \in \mathbb{R}\}$  is process with uncorrelated increments.

We apply the theorem to processes with covariance functions  $r(t) = e^{-t^2/2}$  and  $r(t) = \frac{\sin t}{t}$ . They have absolutely continuous spectra with spectral densities,

$$f(\omega) = \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} \quad \text{and} \quad f(\omega) = 1/2 \quad \text{for } |\omega| < 1,$$

respectively. Obviously  $Q = -\infty$  is divergent in both cases, and we have verified the statement that these processes are deterministic, although their

covariance functions tend to 0 as  $t \rightarrow \infty$ . The Ornstein-Uhlenbeck process with covariance function  $r(t) = e^{-\alpha|t|}$  and spectral density  $f(\omega) = \frac{\alpha}{\pi(\alpha^2 + \omega^2)}$  is an example of a regular process with  $Q > -\infty$ .

Note that in all three examples the covariance function  $r(t)$  tends to 0 as  $t \rightarrow \infty$ , but with quite different rates. For the regular Ornstein-Uhlenbeck process it tends to 0 exponentially fast, while for the two singular (and hence predictable) processes, the covariance falls off either much faster, as  $e^{-t^2/2}$ , or much slower, as  $1/t$ . Hence, we learn that stochastic determinism and non-determinism are complicated matters, even if they are entirely defined in terms of correlations and distribution functions.

### 5.6.3 Uniform, strong, and weak mixing

Predictability, regularity, and singularity are probabilistic concepts, defined and studied in terms of prediction error moments and correlations. When it comes to ergodicity, we have seen examples of totally deterministic sequences which exhibit ergodic behavior, in the sense that they obey the law of large numbers. To complicate matters, we have seen that a Gaussian process, with covariance function tending to 0 at infinity, implying asymptotic independence, is always ergodic, even if the remote past is in a deterministic sense determined by the arbitrarily remote.

Ergodicity is a law of large numbers, time averages converge to a limit. In statistics, one would also like to have some idea of the asymptotic distribution; in particular, one would like a central limit theorem for normalized sums or integrals,

$$\frac{\sum_{k=1}^N x(k) - A_N}{B_N}, \quad \frac{\int_{t=0}^N x(t) dt - A_N}{B_N},$$

as  $N \rightarrow \infty$ . This asks for general concepts, that guarantee the asymptotic independence of *functionals* of a stochastic process.

For a general stationary process, only the ergodic theorem can be expected to hold. For example, if  $x(t) \equiv x$ , then  $B_N^{-1} \sum_{k=1}^N x(k) = x$  with  $A_N = 0$  and  $B_N = N$ , and this has the same distribution as  $x$ , regardless of  $N$ . Here, there is a strong dependence between the  $x$ -variables. But even a very weak dependence would not help us much to obtain any interesting limiting distributions. For example, if  $y(k)$  is a sequence of independent, identically distributed random variables, and

$$x(t) = y(t+1) - y(t),$$

then

$$\frac{1}{B_N} \sum_{k=1}^N x(k) = \frac{y(N+1) - y(1)}{B_N}.$$

This has a non-trivial asymptotic distribution only if  $B_N$  has a non-zero limit. Central limit theorems for normalized sums typically require  $B_N \rightarrow \infty$  and

then some mixing condition has to be imposed. (Martingale arguments are another class of "almost independence" conditions for asymptotic normality; see [36, 37].)

For any stochastic process  $\{x(t), t \in \mathbb{R}\}$  define the  $\sigma$ -field  $\mathcal{M}_a^b$  as the  $\sigma$ -field generated by  $x(t); a \leq t \leq b$ . Taking  $a$  and  $b$  as  $\mp$  infinity, we get  $\mathcal{M}_{-\infty}^b$  and  $\mathcal{M}_a^{\infty}$  as the  $\sigma$ -fields generated by  $\{x(t); t \leq b\}$  and by  $\{x(t); t \geq a\}$ , respectively. The following mixing conditions represent successively milder conditions on the asymptotic independence.

**uniform mixing:**  $x(t)$  is called uniformly mixing (or  $\phi$ -mixing) if there is a non-negative function  $\phi(n)$  such that  $\phi(n) \rightarrow 0$  as  $n \rightarrow \infty$ , and for all  $t$  and events  $A \in \mathcal{M}_{-\infty}^t$  and  $B \in \mathcal{M}_{t+n}^{\infty}$ ,

$$|P(A \cap B) - P(A)P(B)| \leq \phi(n)P(A).$$

**strong mixing:**  $x(t)$  is called strongly mixing (or  $\alpha$ -mixing) if there is a non-negative function  $\alpha(n)$  such that  $\alpha(n) \rightarrow 0$  as  $n \rightarrow \infty$ , and for all  $t$  and events  $A \in \mathcal{M}_{-\infty}^t$  and  $B \in \mathcal{M}_{t+n}^{\infty}$ ,

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(n).$$

**weak mixing:**  $x(t)$  is called weakly mixing if, for all events  $A \in \mathcal{M}_{-\infty}^t$  and  $B \in \mathcal{M}_t^{\infty}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n |P(A \cap U^{-k}B) - P(A)P(B)| = 0.$$

Here,  $U^{-k}B = \{x(\cdot); x(k + \cdot) \in B\} \in \mathcal{M}_k^{\infty}$ .

Of these, uniform mixing is the most demanding and weak mixing the least.

**Theorem 5:10** *Let  $\{x(t); t \in \mathbb{Z}\}$  be a stationary Gaussian sequence. Then*

a)  $x(t)$  is uniformly mixing if and only if it is  $m$ -dependent, i.e. there is an  $m$  such that the covariance function  $r(t) = 0$  for  $|t| > m$ .

b)  $x(t)$  is strong mixing if it has a continuous spectral density  $f(\omega) \geq c > 0$  on  $-\pi < \omega \leq \pi$ .

**Proof:** a) That  $m$ -dependence implies uniform mixing is obvious. To prove the necessity, assume that  $r(t)$  is not identically 0 for large  $t$ . That  $x(t)$  is uniformly mixing implies that for all  $A \in \mathcal{M}_{-\infty}^0$ , with  $P(A) > 0$  and all  $B \in \mathcal{M}_n^{\infty}$ ,

$$|P(B | A) - P(B)| < \phi(n) \rightarrow 0 \tag{5.10}$$

as  $n \rightarrow \infty$ . This directly implies that  $r(t) \rightarrow 0$  as  $t \rightarrow \infty$ , since otherwise there would be an infinite number of  $t = t_k$  for which  $r(t_k) \geq r_0 > 0$ , say. Take  $A = \{x(0) > 0\} \in \mathcal{M}_0^0$  and  $B = \{x(t_k) > 0\} \in \mathcal{M}_{t_k}^{t_k}$ . It is easy to see that

$$P(B | A) - P(B) \geq c_0 > 0$$

is bounded away from 0 by a positive constant  $c_0$ , and hence can not tend to 0.

For simplicity, assume  $E(x(t)) = 0$ ,  $E(x(t)^2) = 1$  and assume there is an infinite number of  $t = t_k$  for which  $\rho_k = r(t_k) > 0$ , but still  $r(t_k) \rightarrow 0$ . Define  $A = \{x(0) > 1/\rho_k\}$ , (obviously  $P(A) > 0$ ) and let  $B = \{x(t_k) > 1\}$ . Since  $x(0), x(t_k)$  have a bivariate normal distribution, the conditional distribution of  $x(t_k)$  given  $x(0) = x$  is normal with mean  $\rho_k x$  and variance  $1 - \rho_k^2$ . As  $\rho_k \rightarrow 0$ , the conditional distribution of  $x(0)$  given  $x(0) > 1/\rho_k$  will be concentrated near  $1/\rho_k$  and then  $x(t_k)$  will be approximately  $N(1, 1)$ . Therefore, as  $\rho_k \rightarrow 0$ ,  $P(B | A) \rightarrow 1/2$ . On the other hand,  $P(B) = (2\pi)^{-1/2} \int_1^\infty \exp(-y^2/2) dy < 0.2$ . Hence (5.10) does not hold for  $\phi(t_k) \rightarrow 0$ .

b) For a proof of this, see [18] or [17].  $\square$

**Theorem 5:11** *A stationary process  $x(t)$  that is strong mixing is ergodic.*

**Proof:** The proof is almost a complete repetition of the proof of Lemma 5.2. Take an invariant event  $S$  and approximate it by a finite-dimensional event  $B$  with  $|P(S) - P(B)| < \epsilon$ . Suppose  $B \in \mathcal{M}_a^b \subset \mathcal{M}_{-\infty}^b$ . Then the translated event

$$B_n = U_n B \in \mathcal{M}_{a+n}^{b+n} \subset \mathcal{M}_{a+n}^\infty,$$

and hence

$$P(B \cap B_n) - P(B) \cdot P(B_n) \rightarrow 0.$$

As in the proof of Lemma 5.2, it follows that  $P(B)^2 = P(B)$  and hence  $P(S)^2 = P(S)$ .  $\square$

The reader could take as a challenge to prove that also weak mixing implies that  $x(t)$  is ergodic; see also Exercise 14.

## Exercises

- 5:1. Show that the following transformation of  $\Omega = [0, 1)$ ,  $\mathcal{F} = \mathcal{B}$ ,  $P =$  Lebesgue measure, is measurable and measure preserving,

$$Tx = 2x \pmod{1}.$$

- 5:2. (Continued.) Define the random variable  $x(\omega) = 0$  if  $0 \leq \omega < 1/2$ ,  $x(\omega) = 1$  if  $1/2 \leq \omega < 1$ . Show that the sequence  $x_n(\omega) = x(T^{n-1}\omega)$  consists of independent zeros and ones, with probability  $1/2$  each.
- 5:3. Show that if  $T$  is measure preserving on  $(\Omega, \mathcal{F}, P)$  and  $x$  is a random variable, then  $E(x(\omega)) = E(x(T\omega))$ .
- 5:4. Show that every one-sided stationary sequence  $\{x_n, n \geq 0\}$  can be extended to a two-sided sequence  $\{x_n, n \in \mathbb{Z}\}$  with the same finite-dimensional distributions.
- 5:5. Find a distribution for  $x_0$  which makes  $x_{k+1} = 4x_k(1 - x_k)$  a stationary sequence.
- 5:6. Prove that if  $E(|x_n|) < \infty$ , then  $P(x_n/n \rightarrow 0, \text{ as } n \rightarrow \infty) = 1$ ; this was used on page 138.
- 5:7. For any sequence of random variables,  $x_n$ , and event  $B \in \mathcal{B}$ , show that the event  $\{x_n \in B, \text{ infinitely often}\}$  is invariant under the shift transformation.
- 5:8. Give an example of an ergodic transformation  $T$  on  $(\Omega, \mathcal{F}, P)$  such that  $T^2$  is not ergodic.
- 5:9. Show that  $x_n$  is ergodic if and only if for every  $A \in \mathcal{B}_k$ ,  $k = 1, 2, \dots$ ,

$$\frac{1}{n} \sum_{j=1}^n \chi_A(x_j, \dots, x_{j+k}) \rightarrow P((x_1, \dots, x_{k+1}) \in A).$$

- 5:10. Show that if  $x$  is non-negative with  $E(x) = \infty$  and  $x_n(\omega) = x(T^{n-1}\omega)$ ,  $S_n = \sum_1^n x_n$ , then  $S_n/n \rightarrow \infty$  if  $T$  is ergodic.
- 5:11. Prove Theorem 5:4.
- 5:12. Take two stationary and ergodic sequences  $x_n$  and  $y_n$ . Take one of the two sequences at random with equal probability,  $z_n = x_n, n = 1, 2, \dots$  or  $z_n = y_n, n = 1, 2, \dots$ . Show that  $z_n$  is not ergodic.
- 5:13. Let  $\{x_n\}$  and  $\{y_n\}$  be two ergodic sequences, both defined on  $(\Omega, \mathcal{F}, P)$ , and consider the bivariate sequence  $z_n = (x_n, y_n)$ . Construct an example that shows that  $z_n$  need not be ergodic, even if  $\{x_n\}$  and  $\{y_n\}$  are independent.

- 5:14. Prove that a sufficient condition for  $z(n) = (x(n), y(n))$  to be ergodic, if  $\{x(n)\}$  and  $\{y(n)\}$  are independent ergodic sequences, is that one of  $\{x(n)\}$  and  $\{y(n)\}$  is weakly mixing.

## Chapter 6

# Vector processes and random fields

### 6.1 Cross-spectrum and spectral representation

The internal correlation structure of a stationary process is defined by the covariance function; the spectrum distributes the correlation over different frequencies and in the spectral representation the process is actually built by individual components, with independent amplitudes and phases, like in the discrete spectrum case (4.18), on page 94. The phases are all independent and uniformly distributed over  $(0, 2\pi)$ . The spectrum does not contain any phase information!

When we have two stationary processes  $\{x(t), t \in \mathbb{R}\}$  and  $\{y(t), t \in \mathbb{R}\}$  which are correlated, their individual amplitudes and phases are still, of course, independent, but for every frequency the amplitudes in the two processes can be dependent, and there can exist a complicated dependence between the phases. This cross dependence is described by the cross-covariance function and the cross-spectrum.

A stationary vector-valued process is a vector of  $p$  stationary processes,

$$\mathbf{x}(t) = (x_1(t), \dots, x_p(t)),$$

with stationary cross-covariances, for  $E(x_j(t)) = 0$ ,

$$r_{jk}(t) = E(x_j(s+t) \cdot \overline{x_k(s)}) = \overline{r_{kj}(-t)}.$$

If the processes are real, which we usually assume,

$$r_{jk}(t) = E(x_j(s+t) \cdot x_k(s)) = r_{kj}(-t).$$

The covariance function

$$\mathbf{R}(t) = (r_{jk}(t))$$

is a matrix function of covariances, where each auto-covariance function  $r_{kk}(t)$  has its marginal spectral representation  $r_{kk}(t) = \int e^{i\omega t} dF_{kk}(\omega)$ .

### 6.1.1 Spectral distribution

**Theorem 6:1** (a) To every continuous covariance matrix function  $\mathbf{R}(t)$  there exists a spectral distribution  $\mathbf{F}(\omega)$  such that

$$\mathbf{R}(t) = \int_{-\infty}^{\infty} e^{i\omega t} d\mathbf{F}(\omega),$$

where  $\mathbf{F}(\omega)$  is a function of positive type, i.e. for every pair  $j, k$ , complex  $\mathbf{z} = (z_1, \dots, z_n)$ , and frequency interval  $\omega_1 < \omega_2$ ,

$$\sum_{j,k} (F_{jk}(\omega_2) - F_{jk}(\omega_1)) z_j \bar{z}_k \geq 0.$$

This says that  $\Delta\mathbf{F}(\omega) = (F_{jk}(\omega_2) - F_{jk}(\omega_1))$  is a non-negative definite Hermite matrix.

(b) If  $F_{jj}(\omega)$ ,  $F_{kk}(\omega)$  are absolutely continuous with spectral densities  $f_{jj}(\omega)$ ,  $f_{kk}(\omega)$ , then  $F_{jk}(\omega)$  is absolutely continuous, with

$$|f_{jk}(\omega)|^2 \leq f_{jj}(\omega) f_{kk}(\omega).$$

**Proof:** (a) Take the  $\mathbf{z}$  and define the stationary process

$$y(t) = \sum_j z_j x_j(t)$$

with covariance function

$$r^{\mathbf{z}}(t) = \sum_{jk} r_{jk}(t) z_j \bar{z}_k = \int e^{i\omega t} dG^{\mathbf{z}}(\omega),$$

where  $G^{\mathbf{z}}(\omega)$  is a real, bounded and non-decreasing spectral distribution function. We then take, in order,  $z_j = z_k = 1$ , and  $z_j = i$ ,  $z_k = 1$ , the rest being 0. This gives two spectral distribution  $G_1(\omega)$ , and  $G_2(\omega)$ , say, and we have,

$$r_{jj}(t) + r_{kk}(t) + r_{jk}(t) + r_{kj}(t) = \int e^{i\omega t} dG_1(\omega),$$

$$r_{jj}(t) + r_{kk}(t) + ir_{jk}(t) - ir_{kj}(t) = \int e^{i\omega t} dG_2(\omega).$$

Together with  $r_{jj}(t) = \int e^{i\omega t} dF_{jj}(\omega)$ , and  $r_{kk}(t) = \int e^{i\omega t} dF_{kk}(\omega)$ , we get

$$r_{jk}(t) + r_{kj}(t) = \int e^{i\omega t} (dG_1(\omega) - dF_{jj}(\omega) - dF_{kk}(\omega)),$$

$$ir_{jk}(t) - ir_{kj}(t) = \int e^{i\omega t} (dG_2(\omega) - dF_{jj}(\omega) - dF_{kk}(\omega)),$$

which implies

$$\begin{aligned} r_{jk}(t) &= \overline{r_{kj}(t)} = \int e^{i\omega t} \cdot \frac{1}{2} (dG_1(\omega) - idG_2(\omega) - (1-i)(dF_{jj}(\omega) + dF_{kk}(\omega))) \\ &= \int e^{i\omega t} dF_{jk}(\omega), \end{aligned}$$

say, which is the spectral representation of  $r_{jk}(t)$ .

It is easy to see that  $\Delta \mathbf{F}(\omega)$  has the stated properties; in particular that

$$\sum_{jk} \Delta F_{jk}(\omega) z_j \overline{z_k} \geq 0. \quad (6.1)$$

(b) From (6.1), by taking only  $z_j$  and  $z_k$  to be non-zero, it follows that

$$\Delta F_{jj} |z_j|^2 + \Delta F_{kk} |z_k|^2 + 2 \Re e(\Delta F_{jk} z_j \overline{z_k}) \geq 0,$$

which in turn implies that for any  $\omega$ -interval,  $|\Delta F_{jk}|^2 \leq \Delta F_{jj} \cdot \Delta F_{kk}$ . Thus, if  $F_{jj}$  and  $F_{kk}$  have spectral densities, so does  $F_{jk}$  and

$$|f_{jk}(\omega)|^2 \leq f_{jj}(\omega) f_{kk}(\omega).$$

□

For real-valued vector processes, the spectral distributions may be put in real form as for one-dimensional processes. In particular, the cross-covariance function can be written

$$r_{jk}(t) = \int_0^\infty \{ \cos \omega t dG_{jk}(\omega) + \sin \omega t dH_{jk}(t) \}, \quad (6.2)$$

where  $G_{jk}(\omega)$  and  $H_{jk}(\omega)$  are functions of bounded variation.

## 6.1.2 Spectral representation of $\mathbf{x}(t)$

### 6.1.2.1 The spectral components

Each component  $x_j(t)$  in a stationary vector process has its spectral representation  $x_j(t) = \int e^{i\omega t} dZ_j(\omega)$  in terms of a spectral process  $Z_j(\omega)$  with orthogonal increments. Further, for  $j \neq k$ , the increments of  $Z_j(\omega)$  and  $Z_k(\omega)$  over disjoint  $\omega$ -intervals are orthogonal, while for equal  $\omega$ -intervals,

$$\begin{aligned} E(dZ_j(\omega) \cdot \overline{dZ_k(\omega)}) &= dF_{jk}(\omega), \\ E(dZ_j(\omega) \cdot \overline{dZ_k(\mu)}) &= 0, \quad \text{for } \omega \neq \mu. \end{aligned}$$

The cross-correlation between the components of  $\mathbf{x}(t)$  are determined by the correlations between the spectral components. To see how it works, consider

processes with discrete spectrum, for which the spectral representation are sums of the form (4.29),

$$x_j(t) = \sum_n \sigma_j(n) (U_j(n) \cos \omega(n)t + V_j(n) \sin \omega(n)t).$$

Here  $U_j(n)$  and  $V_j(n)$  are real random variables with mean 0, variance 1, uncorrelated for different  $n$ -values. The correlation between the  $x_j$ - and the  $x_k$ -process is caused by a correlation between the  $U$ :s and  $V$ :s in the two representations:

$$\begin{aligned} E(U_j(n)U_k(n)) &= E(V_j(n)V_k(n)) = \rho_{jk}(n), \quad j \neq k, \\ E(U_j(n)V_k(n)) &= -E(V_j(n)U_k(n)) = -\tilde{\rho}_{jk}(n), \quad j \neq k, \\ E(U_j(n)V_j(n)) &= 0, \end{aligned}$$

for some  $\rho_{jk}(n)$  and  $\tilde{\rho}_{jk}(n)$  such that  $0 \leq \rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2 \leq 1$ . Direct calculation of auto- and cross-covariances gives

$$\begin{aligned} r_{jk}(t) &= \sum_n \sigma_j(n)\sigma_k(n) (\rho_{jk}(n) \cos \omega(n)t + \tilde{\rho}_{jk}(n) \sin \omega(n)t), \\ &= \sum_n A_{jk}(n) \cos(\omega(n)t - \Phi_{jk}(n)), \quad j \neq k, \end{aligned} \quad (6.3)$$

$$\begin{aligned} r_{kj}(t) &= \sum_n \sigma_j(n)\sigma_k(n) (\rho_{jk}(n) \cos \omega(n)t - \tilde{\rho}_{jk}(n) \sin \omega(n)t), \\ &= \sum_n A_{kj}(n) \cos(\omega(n)t - \Phi_{kj}(n)), \quad j \neq k, \end{aligned} \quad (6.4)$$

$$r_j(t) = \sum_n \sigma_j(n)^2 \cos \omega(n)t, \quad j = 1, \dots, p. \quad (6.5)$$

Here,  $A_{jk}(n) = A_{kj}(n) = \sigma_j(n)\sigma_k(n)\sqrt{\rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2}$ , represent the correlation between the amplitudes, while  $\Phi_{jk}(n) = -\Phi_{kj}(n)$ , with

$$\cos \Phi_{jk}(n) = \frac{\rho_{jk}(n)}{\sqrt{\rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2}}, \quad \sin \Phi_{jk}(n) = \frac{\tilde{\rho}_{jk}(n)}{\sqrt{\rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2}},$$

represent the phase relations.

The corresponding spectral distributions  $F_{kk}(\omega)$  are symmetric with mass

$$\Delta F_k = \sigma_k(n)^2/2$$

at the frequencies  $\pm\omega(n)$ , while for  $j \neq k$ , the cross spectrum  $F_{jk}(\omega)$  is skewed if  $\tilde{\rho}_{jk}(n) \neq 0$ , with mass

$$\Delta F_{jk} = \begin{cases} \frac{1}{2}A_{jk}(n) e^{-i\Phi_{jk}(n)} = \frac{1}{2}A_{jk}(n) \frac{\rho_{jk}(n) + i\tilde{\rho}_{jk}(n)}{\sqrt{\rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2}}, & \text{at } \omega = \omega_n, \\ \frac{1}{2}A_{jk}(n) e^{i\Phi_{jk}(n)} = \frac{1}{2}A_{jk}(n) \frac{\rho_{jk}(n) - i\tilde{\rho}_{jk}(n)}{\sqrt{\rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2}}, & \text{at } \omega = -\omega_n. \end{cases}$$

### 6.1.2.2 Phase, amplitude, and coherence spectrum

The frequency dependent function  $\Phi_{jk}(n)$  is called the *phase spectrum* and it defines the time delay between components in  $x_k(s)$  and  $x_j(s+t)$ ; the correlation between the components  $U_j(n) \cos(\omega(n)(s+t)) + V_j(n) \sin(\omega(n)(s+t))$  and  $U_k(n) \cos(\omega(n)s) + V_k(n) \sin(\omega(n)s$  have their maxima at  $t = \Phi_{jk}(n)/\omega(n)$ .

Further,  $A_{jk}(n)/2$  is called the *amplitude spectrum* and the *squared coherence spectrum* is defined as

$$\frac{|\Delta F_{jk}(n)|^2}{\Delta F_{jj}(n)\Delta F_{kk}(n)} = \rho_{jk}(n)^2 + \tilde{\rho}_{jk}(n)^2.$$

For processes with continuous spectra and cross spectral density

$$f_{jk}(\omega) = \frac{1}{2}A_{jk}(\omega) e^{i\Phi_{jk}(\omega)},$$

the amplitude spectrum, phase spectrum, and squared coherence spectrum are defined as  $A_{jk}(\omega)/2$ ,  $\Phi_{jk}(\omega)$ , and  $|f_{jk}(\omega)|^2/(f_{jj}(\omega)f_{kk}(\omega))$ , respectively.

### 6.1.2.3 Cross-correlation in linear filters

In a linear filter, the cross-covariance and cross spectrum describe the relation between the input process  $x(t) = \int e^{i\omega t} dZ_x(\omega)$  and the output process  $y(t) = \int h(t-u)x(u) du = \int g(\omega)e^{i\omega t} dZ_x(\omega)$ . One calculates easily,

$$\begin{aligned} r_{xy}(t) &= E(x(s+t) \cdot y(s)) \\ &= E\left(\int e^{i\omega(s+t)} dZ_x(\omega) \cdot \overline{\int g(\omega')e^{i\omega's} dZ_x(\omega')}\right) = \int e^{i\omega t} \overline{g(\omega)} dF_x(\omega), \end{aligned}$$

so the cross spectral distribution is

$$dF_{xy}(\omega) = \overline{g(\omega)} dF_x(\omega).$$

By estimating the cross-covariance or the cross spectrum between input and output in an unknown filter, one can estimate the frequency function as  $g^*(\omega) = \overline{f_{xy}^*(\omega)}/f_x^*(\omega) = f_{xy}^*(-\omega)/f_x^*(\omega)$ , when  $f_x^*(\omega) > 0$ .

## 6.2 Some random field theory

A random field is a stochastic process  $x(\mathbf{t})$  with multi-dimensional parameter

$$\mathbf{t} = (t_1, \dots, t_p).$$

For example, if  $\mathbf{t} = (t_1, t_2)$  is two-dimensional we can think of  $(t_1, t_2, x(\mathbf{t}))$  as a random surface. A time-dependent random surface is a field  $(s_1, s_2, x(t, s_1, s_2))$  with  $\mathbf{t} = (t, s_1, s_2)$ , where  $t$  is time and  $(s_1, s_2) \in \mathbb{R}^2$  is location. In the general

theory we use  $\mathbf{t}$  as generic notation for the parameter; in special applications to random time dependent surfaces we use  $(t, (s_1, s_2))$  as parameter.

From being mainly used in geoscience, like geology under the name of *geo-statistics*, and in marine science as models for a random sea, cf. Section 1.6.3, random fields are now widely used in all sorts of applications with spatial or spatial-temporal variability.

### 6.2.1 Homogeneous fields

Define the mean value and covariance functions for random fields in the natural way as  $m(\mathbf{t}) = E(x(\mathbf{t}))$  and  $r(\mathbf{t}, \mathbf{u}) = C(x(\mathbf{t}), x(\mathbf{u}))$ .

The analogue of a stationary process is a homogeneous field. The field is called *homogeneous* if  $m(\mathbf{t})$  is constant  $m$  and  $r(\mathbf{t}, \mathbf{u})$  depends only on the difference  $\mathbf{t} - \mathbf{u}$ , i.e. assuming a real field with  $m = 0$ ,

$$r(\mathbf{t}) = r(\mathbf{u} + \mathbf{t}, \mathbf{u}) = E(x(\mathbf{u} + \mathbf{t}) \cdot x(\mathbf{u})).$$

The covariance of the process values at two parameter points depends on *distance* as well as on *direction* of the vector between the two points.

In spatial applications it is popular to use the *variogram* defined by

$$2\gamma(\mathbf{u}, \mathbf{v}) = E(|x(\mathbf{u}) - x(\mathbf{v})|^2)$$

or the *semi-variogram*  $\gamma(\mathbf{t})$ .

The variogram plays the same role as the incremental variance does in a Wiener process. There  $E(|w(s+t) - w(s)|^2) = t \cdot \sigma^2$  is independent of  $s$ . A field for which the variogram only depends on the vector  $\mathbf{u} - \mathbf{v}$  is called *intrinsically stationary*. The variogram for a homogeneous field is  $\gamma(\mathbf{t}) = r(\mathbf{0}) - r(\mathbf{t})$ . A homogeneous field is also intrinsically stationary, but as seen from the Wiener process the converse is not sure.

**Theorem 6:2** (a) *The covariance function  $r(\mathbf{t})$  of a homogeneous random field has a spectral distribution*

$$r(\mathbf{t}) = \int e^{i\boldsymbol{\omega} \cdot \mathbf{t}} dF(\boldsymbol{\omega}),$$

where  $\boldsymbol{\omega} \cdot \mathbf{t} = \omega_1 t_1 + \dots + \omega_p t_p$ , and  $F(\boldsymbol{\omega})$  is a  $p$ -dimensional spectral distribution function depending on the frequency parameter  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)$ .<sup>1</sup>

(b) *There exists a stochastic spectral process  $Z(\boldsymbol{\omega})$  with orthogonal increments  $\Delta Z(\boldsymbol{\omega})$  over rectangles  $\Delta\boldsymbol{\omega} = [\omega_1, \omega_1 + \Delta_1] \times \dots \times [\omega_p, \omega_p + \Delta_p]$ , such that  $E(\Delta Z(\boldsymbol{\omega})) = 0$  and*

$$E(|\Delta Z(\boldsymbol{\omega})|^2) = \Delta F(\boldsymbol{\omega}), \quad \text{and} \quad E(\Delta Z(\boldsymbol{\omega}_1) \cdot \overline{\Delta Z(\boldsymbol{\omega}_2)}) = 0,$$

<sup>1</sup>  $F(\boldsymbol{\omega})$  is equal to a regular  $p$ -dimensional probability distribution function multiplied by a positive constant, equal to the variance of the process.

for disjoint rectangles  $\Delta\omega_1$  and  $\Delta\omega_2$ , and

$$x(\mathbf{t}) = \int e^{i\omega \cdot \mathbf{t}} dZ(\omega). \quad (6.6)$$

### 6.2.2 Isotropic fields

In an homogeneous isotropic field, the correlation properties are the same in all directions, i.e. the covariance function  $r(\mathbf{t})$  depends only on the distance  $\|\mathbf{t}\| = \sqrt{t_1^2 + \dots + t_p^2}$ . This type of process model is natural when one can not identify any special directional dependent stochastic properties in the field, but rather it keeps its distributions after rotation (and translation). Many phenomena in the natural world share this property while other naturally do not. Ocean waves are non-isotropic, chemical concentration field in the bottom sediment in the ocean or the disturbance field in mobile phone communication might well be isotropic.

As we have seen, a spectral distribution for a homogeneous field need only satisfy the requirement that it is non-negative, integrable and symmetric. The spectral distribution for an isotropic field needs to satisfy a special invariance condition, giving it a particularly simple structure.

**Theorem 6:3** *The covariance function  $r(\mathbf{t})$  to a homogeneous isotropic field a mixture of Bessel functions,*

$$r(\mathbf{t}) = r(\|\mathbf{t}\|) = \int_0^\infty \frac{J_{(p-2)/2}(\omega \cdot \|\mathbf{t}\|)}{(\omega \cdot \|\mathbf{t}\|)^{(p-2)/2}} dG(\omega),$$

where  $G(\omega)$  is a bounded, non-decreasing function, and  $J_m(\omega)$  is a Bessel function of the first kind of order  $m$ ,

$$J_m(\omega) = \sum_{k=0}^{\infty} (-1)^k \frac{(\omega/2)^{2k+m}}{k! \Gamma(k+m+1)}.$$

**Proof:** We have  $r(\mathbf{t}) = \int e^{i\omega \cdot \mathbf{t}} dF(\omega)$ . Introduce the spherical coordinates,  $\omega = \omega \cdot (\ell_1, \dots, \ell_p)$ ,  $\omega = \|\omega\|$ ,  $\sum \ell_k^2 = 1$ , and let  $\ell_1 = \cos \theta_{p-1}$ . For every  $\theta_{p-1}$ ,  $(\ell_2, \dots, \ell_p)$ , defines a point on the sphere with radius  $\sqrt{1 - \cos^2 \theta_{p-1}} = \sin \theta_{p-1}$ .

Since  $r(\mathbf{t})$  depends only on  $\|\mathbf{t}\|$ , we can calculate the integral for the special point  $\mathbf{t} = (t, 0, \dots, 0)$ , to get

$$r(\mathbf{t}) = \int e^{i\omega_1 t} dF(\omega) = \int e^{i\omega t \cos \theta_{p-1}} dF(\omega).$$

With  $G(\omega) = \int_{\|\omega\| \leq \omega} dF(\omega)$  we find that

$$r(t) = \int_{\omega=0}^{\infty} \left\{ \int_{\theta} e^{i\omega t \cos \theta_{p-1}} d\sigma(\theta) \right\} dG(\omega),$$

where  $d\sigma$  is the rotation invariant measure on the  $(p-1)$ -dimensional unit sphere. For fixed  $\theta_{p-1}$  we integrate  $(\theta_1, \dots, \theta_{p-2})$  over the sphere with radius  $\sin \theta_{p-1}$ , with area  $C_{p-1} \sin^{p-2} \theta_{p-1}$ . We find

$$r(t) = \int_{\omega=0}^{\infty} \left\{ \int_{\theta=0}^{\pi} C_{p-2} e^{i\omega t \cos \theta} \sin^{p-2} \theta d\theta \right\} dG(\omega),$$

which is the stated form, if we incorporate the constant  $C_{p-2}$  into  $G$ .  $\square$

### 6.2.2.1 Isotropic fields with special structure

Of course, a homogeneous field  $x(t_1, t_2)$  with two-dimensional parameter and spectral density  $f_x(\omega_1, \omega_2)$  and covariance function  $r_x(t_1, t_2)$  gives rise to a stationary process when observed along a single straight line, for example along  $t_2 = 0$ . Then  $x(t_1, 0)$  has covariance function  $r(t_1) = r_x(t_1, 0)$  and spectral density  $f(\omega_1) = \int_{\omega_2} f_x(\omega_1, \omega_2) d\omega_2$ .

It is important to realise that even if any non-negative definite function can act as covariance function for a stationary process, not all stationary processes, and corresponding covariance functions, can occur as an observation of a section in an isotropic field with two-dimensional parameter. Only those functions that can be expressed as

$$r(\mathbf{t}) = \int_0^{\infty} J_0(\omega \|\mathbf{t}\|) dG(\omega),$$

with bounded non-decreasing  $G(\omega)$  are possible. A simple *sufficient* condition for a one-dimensional spectral density  $f(\omega)$  to be obtained from a section of an isotropic field is the following:

- Any function  $f(\omega)$ ,  $\omega \in \mathbb{R}$  that is non-increasing for  $\omega \geq 0$  and integrable over  $\mathbb{R}$  can act as the spectral density for a section, e.g.  $x(t_1, 0)$ , of an isotropic field in two dimensions.

One should also note the particularly simple form of the covariance function for the case  $p = 3$ ,

$$r(\|\mathbf{t}\|) = \int_0^{\infty} \frac{\sin(\omega \|\mathbf{t}\|)}{\omega \|\mathbf{t}\|} dG(\omega).$$

Another special case that needs special treatment is when the parameter  $\mathbf{t}$  is composed of both a time and a space parameter,  $\mathbf{t} = (t, s_1, \dots, s_p)$ , one could hope for isotropy in  $(s_1, \dots, s_p)$  only, in which case the spectral form becomes

$$r(t, \|(s_1, \dots, s_p)\|) = \int_{\nu=-\infty}^{\infty} \int_{\omega=0}^{\infty} e^{i\nu t} \cdot H_p(\omega \|(s_1, \dots, s_p)\|) dG(\nu, \omega),$$

where

$$H_p(x) = ((2/x))^{(p-2)/2} \Gamma(p/2) J_{(p-2)/2}(x).$$

The form of the covariance function for an isotropic field as a mixture of Bessel functions is useful for non-parametric estimation from data. It is only

the weight function  $G(\omega)$  that needs to be estimated, for example as a discrete distribution.

The theorem gives all possible isotropic covariance functions valid in the special dimension  $p$ . Some functions can be used as covariance function in any dimension, for example

$$r(\|\mathbf{t}\|) = \sigma^2 \exp(-\phi\|\mathbf{t}\|^\alpha),$$

for any  $\alpha \in (0, 2)$ .

Another popular class of covariance functions valid in any dimension is the *Whittle-Matérn family*, which is

$$r(\|\mathbf{t}\|) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}\|\mathbf{t}\|\phi)^\nu K_\nu(2\sqrt{\nu}\|\mathbf{t}\|\phi),$$

where  $K_\nu$  is a modified Bessel function of order  $\nu$ . The value of  $\nu$  determines the smoothness of the field. When  $\nu \rightarrow \infty$ , the Whittle-Matérn covariance functions tends to a Gaussian shape, in which case the field is infinitely differentiable; cf. Chapter 2.

In all these forms,  $\phi$  is a spatial scale parameter. Often one adds an extra variance term  $\sigma_0^2$  at the center, for  $\mathbf{t} = \mathbf{0}$ , to account for the so called *nugget effect*, a variance component with a covariance range too short to be observed.

A modern account of spatial data modeling is given in [2], also by means of spectral and correlation models.

### 6.2.3 Randomly moving surfaces

The spectral representation makes it easy to imagine the structure of a randomly moving surface, homogeneous in time as well as in space. The spectral representation (6.6) is a representation of real field  $x(\mathbf{t})$  as a packet of directed waves,  $A_{\mathbf{t}} \cos(\boldsymbol{\omega} \cdot \mathbf{t} + \phi_{\boldsymbol{\omega}})$ , with random amplitude and phase, and constant on each plane parallel to  $\boldsymbol{\omega} \cdot \mathbf{t} = 0$ . For example, with  $\mathbf{t} = (s_1, s_2, t)$  and  $t$  is time and  $(s_1, s_2)$  is space, and  $\boldsymbol{\omega} = (\kappa_1, \kappa_2, \omega)$ , the elementary waves are

$$A_{\boldsymbol{\omega}} \cos(\kappa_1 s_1 + \kappa_2 s_2 + \omega t + \phi_{\boldsymbol{\omega}}).$$

For fixed  $t$  this is a cosine-function in the plane, which is zero along lines  $\kappa_1 s_1 + \kappa_2 s_2 + \omega t + \phi_{\boldsymbol{\omega}} = \pi/2 + k\pi$ ,  $k$  integer. For fixed  $(s_1, s_2)$  it is a cosine wave with frequency  $\omega$ . The parameters  $\kappa_1$  and  $\kappa_2$  are called the wave numbers.

In general there is no particular relation between the time frequency  $\omega$  and the space frequencies  $\nu$ , except for water waves, which we shall deal with later. However, one important application of space-time random fields is the modeling of environmental variables, like the concentration of a hazardous pollutant. Over a reasonably short period of time the concentration variation may be regarded as statistically stationary in time, at least averaged over a 24 hour

period. But it is often unlikely that the correlation structure in space is independent of the absolute location. Topography, location of cities and pollutant sources, makes the process inhomogeneous in space.

One way to overcome the inhomogeneity is to make a transformation of the space map and move each observation point  $(s_1, s_2)$  to a new location  $(\hat{s}_1, \hat{s}_2)$  so that the field  $\hat{x}(t, \hat{s}_1, \hat{s}_2) = x(t, s_1, s_2)$  is homogeneous. This may not be exactly attainable, but the technique is often used in environmental statistics for planning of measurements.

### 6.2.4 Stochastic water waves

Stochastic water waves are special cases of homogeneous random fields, for which there is a special relation between time and space frequencies (wave numbers). For a one-dimensional time dependent Gaussian wave  $x(t, s)$ , where  $s$  is distance along an axis, the elementary waves have the form

$$A_\omega \cos(\omega t - \kappa s + \phi_\omega).$$

By physical considerations one can derive an explicit relation, the *dispersion relation*, between wave number  $\kappa$  and frequency  $\omega$ . If  $h$  is the water depth,

$$\omega^2 = \kappa g \tanh(h\kappa),$$

which for infinite depth<sup>2</sup> reduces to  $\omega^2 = \kappa g$ . Here  $g$  is the constant of gravity.

A Gaussian random wave is a mixture of elementary waves of this form, in spectral language, with  $\kappa > 0$  solving the dispersion relation,

$$\begin{aligned} x(t, s) &= \int_{\omega=-\infty}^{\infty} e^{i(\omega t - \kappa s)} dZ_+(\omega) + \int_{\omega=-\infty}^{\infty} e^{i(\omega t + \kappa s)} dZ_-(\omega) \\ &= x_+(t, s) + x_-(t, s). \end{aligned}$$

Here is a case when it is important to use both positive and negative frequencies; cf. the comments in Section 4.3.3.4. Waves described by  $x_+(t, s)$  move to the right and waves in  $x_-(t, s)$  move to the left with increasing  $t$ .

Keeping  $t = t_0$  or  $s = s_0$  fixed, one obtains a *space wave*,  $x(t_0, s)$ , and a *time wave*,  $x(t, s_0)$  respectively. The spectral density of the time wave,  $x(t, s_0) = x_+(t, s_0) + x_-(t, s_0)$  is called the *wave frequency spectrum*,

$$f_x^{freq}(\omega) = f_+(\omega) + f_-(\omega),$$

and we see again that it is not possible to distinguish between the two wave directions by just observing the time wave.

---

<sup>2</sup>  $\tanh x = (e^x - e^{-x})/(e^x + e^{-x})$ .

The space wave has a wave number spectrum given by the equation, for infinite water depth, with  $\omega^2 = g\kappa > 0$ ,

$$f_x^{time}(\omega) = \frac{2\omega}{g} f_x^{space}(\omega^2/g),$$

$$f_x^{space}(\kappa) = \frac{1}{2} \sqrt{\frac{g}{\kappa}} f_x^{time}(\sqrt{g\kappa}).$$

One obvious effect of these relations is that the space process seems to have more short waves than can be inferred from the time observations. Physically this is due to the fact that short waves travel with lower speed than long waves, and they are therefore not observed as easily in the time process. Both the time wave observations and the space registrations are in a sense “biased” as representatives for the full time-space wave field.

In Chapter 3, Remark 3:3, we introduced the mean period  $2\pi\sqrt{\omega_0/\omega_2}$  of a stationary time process. The corresponding quantity for the space wave is the *mean wave length*, i.e. the average distance between two upcrossings of the mean level by the space process  $x(t_0, s)$ . It is expressed in terms of the spectral moments of the wave number spectrum, in particular

$$\kappa_0 = \int_{\kappa} f_x^{space}(\kappa) d\kappa = \int_{\omega} f_x^{time} d\omega = \omega_0,$$

$$\kappa_2 = \int_{\kappa} \kappa^2 f_x^{space}(\kappa) d\kappa = \int_{\omega} \frac{\omega^4}{2g^2} \sqrt{\frac{g^2}{\omega^2}} f_x^{time}(\omega) \frac{2\omega}{g} d\omega = \omega_4/g^2.$$

The average wave length is therefore  $2\pi g\sqrt{\omega_0/\omega_4}$ . We see that the average wave length is more sensitive to the tail of the spectral density than is the average wave period. Considering the difficulties in estimating the high frequency part of the wave spectrum, all statements that rely on high spectral moments are unreliable.

The case of a two-dimensional time dependent Gaussian wave  $x(t, s_1, s_2)$ , the elementary waves with frequency  $\omega$  and direction  $\theta$  become

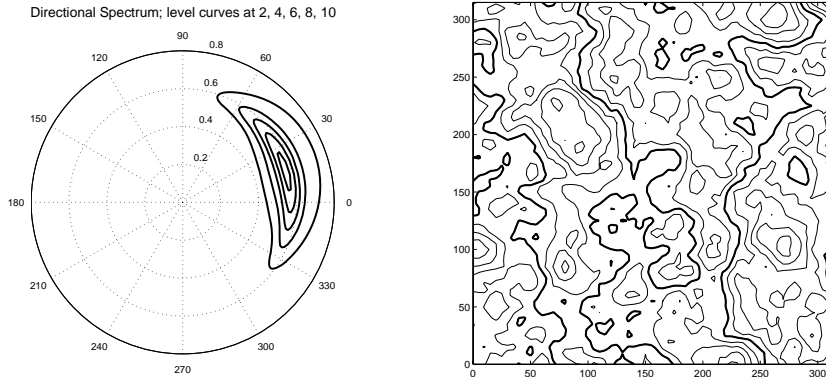
$$A_{\omega} \cos(\omega t - \kappa(s_1 \cos \theta + s_2 \sin \theta) + \phi_{\omega}), \quad (6.7)$$

where  $\omega > 0$  and  $\kappa > 0$  is given by the dispersion relation. With this choice of sign,  $\theta$  determines the direction in which the waves move.

The spectral density for the time-space wave field specifies the contribution to  $x(t, (s_1, s_2))$  from elementary waves of the form (6.7). Summed (or rather integrated) over all directions  $0 \leq \theta < 2\pi$ , they give the time wave  $x(t, \mathbf{s}_0)$ , in which one cannot identify the different directions. The spectral distribution, called the *directional spectrum*, is therefore often written in polar form, based on the spectral density  $f_x^{time}(\omega)$  for the time wave, as

$$f(\omega, \theta) = f_x^{time}(\omega)g(\omega, \theta).$$

The *spreading function*  $g(\omega, \theta)$ , with  $\int_0^{2\pi} g(\omega, \theta) d\theta = 1$ , specifies the relative contribution of waves from different directions. It may be frequency dependent.



**Figure 6.1:** *Left: Level curves for directional spectrum with frequency dependent spreading. Right: Level curves for simulated Gaussian space sea.*

**Example 6:1** Wave spectra for the ocean under different weather conditions are important to characterize the input to (linear or non-linear) ship models. Much effort has been spent on design and estimation of typical wave spectra. One of the most popular is the *Pierson-Moskowitz* spectrum,

$$f_{PM}^{time}(\omega) = \frac{\alpha}{\omega^5} e^{-1.25(\omega_m/\omega)^4}.$$

or the variant, the *JONSWAP* spectrum, in which an extra factor  $\gamma > 1$  is introduced to enhance the peak of the spectrum,

$$f_J^{time}(\omega) = \frac{\alpha}{\omega^5} e^{-1.25(\omega_m/\omega)^4} \gamma^{\exp(-(1-\omega/\omega_m)^2/2\sigma_m^2)}.$$

In both spectra,  $\alpha$  is a main parameter for the total variance, and  $\omega_m$  defines the “peak frequency”. The parameters  $\gamma$  and  $\sigma_m$  determine the peakedness of the spectrum. The spectrum and a realization of a Gaussian process with *JONSWAP* spectrum was shown in Example 1:4.

Figure 6.1 shows to the right, the level curves for a simulated Gaussian wave surface with the directional spectrum with frequency dependent spreading, shown on the left. Frequency spectrum is of *JONSWAP* type.

As mentioned in the historical Section 1.6.3, Gaussian waves have been used since the early 1950, with great success. However, since Gaussian processes are symmetric,  $x(t)$  has the same distribution as  $-x(t)$  and as  $x(-t)$ , they are not very realistic for actual water waves except in special situations; deep water, no strong wind. Much research is presently devoted to development of “non-linear” stochastic wave models, where elementary waves with different frequencies can interact, in contrast to the “linear” Gaussian model, where they just add up.

## Exercises

6:1. To be written.



# Appendix A

## The axioms of probability

Here is a collection of the basic probability axioms, together with proofs of the extension theorem for probabilities on a field (page 5), and of Kolmogorov's extension theorem, from finite-dimensional probabilities to infinite-dimensional ones (page 10).

### A.1 The axioms

A probability  $P$  is a countably additive measure on a probability space, as defined here.

**Definition A:1** (a) A family of subsets  $\mathcal{F}_0$  to an arbitrary space  $\Omega$  is called a field (or algebra) if it contains the whole set  $\Omega$  and is closed under the set operations complement,  $A^*$ , union,  $A \cup B$ , and intersection,  $A \cap B$ , i.e. if  $A$  and  $B$  are sets in the family  $\mathcal{F}_0$ , then also the complement  $A^*$  and the union  $A \cup B$  belong to  $\mathcal{F}_0$ , etc. It then also contains all unions of finitely many sets  $A_1, \dots, A_n$  in  $\mathcal{F}_0$ .

(b) A field  $\mathcal{F}$  of subsets is called a  $\sigma$ -field (or  $\sigma$ -algebra) if it contains all countable unions and intersections of its sets, i.e. if it is a field and furthermore,

$$A_1, A_2, \dots \in \mathcal{F} \quad \text{implies} \quad \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

(c) A probability measure  $P$  on a sample space  $\Omega$  with a  $\sigma$ -field  $\mathcal{F}$  of events, is a function defined for every  $A \in \mathcal{F}$ , with the properties

(1)  $0 \leq P(A) \leq 1$  for all  $A \in \mathcal{F}$ .

(2)  $P(\Omega) = 1$ .

(3) For disjoint sets  $A_k \in \mathcal{F}$ ,  $k = 1, 2, \dots$  one has  $P(\bigcup_1^{\infty} A_k) = \sum_1^{\infty} P(A_k)$ .

(d) Equivalent with (c) is

- (1)  $0 \leq P(A) \leq 1$  for all  $A \in \mathcal{F}$ .
- (2)  $P(\Omega) = 1$ .
- (3') If  $A_1, A_2 \in \mathcal{F}$  are disjoint, then  $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ .
- (3'') If  $A_1 \supseteq A_2 \supseteq \dots \in \mathcal{F}$  with  $\bigcap_1^\infty A_n = \emptyset$ , then  $\lim_{n \rightarrow \infty} P(A_n) = 0$ .

A typical field  $\mathcal{F}_0$  in  $\mathbb{R}$  is the family of sets which are unions of a finite number of intervals. The smallest  $\sigma$ -field that contains all sets in  $\mathcal{F}_0$  is the family  $\mathcal{B}$  of Borel sets.

## A.2 Extension of a probability from field to $\sigma$ -field

How do we define probabilities? For real events, of course, via a statistical distribution function  $F(x)$ . Given a distribution function  $F$ , we can define a probability  $P$  for finite unions of real half open disjoint intervals,

$$P(\cup_1^n (a_k, b_k]) = \sum_1^n (F(b_k) - F(a_k)). \quad (\text{A.1})$$

If we define the probability of a single point as  $P([a]) = F(a) - \lim_{x \uparrow a} F(x)$ , we can define a probability to every finite union of real intervals.

Sets which are unions of a finite number real intervals,  $(a, b]$ ,  $(a, b)$ ,  $[a, b)$ ,  $[a, b]$  with  $-\infty \leq a < b \leq \infty$ , form a field on  $\mathbb{R}$ , and they can be given probabilities via a distribution function. The question is, does this also give probabilities to the more complicated events (Borel sets) in the  $\sigma$ -field  $\mathcal{F}$  generated by the intervals. The answer is yes, as stated in the following extension theorem, *Carathéodory's extension theorem*, which is valid not only for intervals and Borel sets, but for any field  $\mathcal{F}_0$  and generated  $\sigma$ -field  $\mathcal{F}$ . For a proof, the reader is referred to any text book in probability or measure theory, e.g. [36].

**Theorem A:1** *Suppose  $P$  is a function which is defined for all sets in a field  $\mathcal{F}_0$ , there satisfying the probability axioms, i.e. (with three equivalent formulations of Condition 4),*

- (1)  $0 \leq P(A) \leq 1$  for all  $A \in \mathcal{F}_0$ .
- (2)  $P(\Omega) = 1$ .
- (3) If  $A_1, A_2 \in \mathcal{F}_0$  are disjoint, then  $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ .
- (4a) If  $A_1 \supseteq A_2 \supseteq \dots \in \mathcal{F}_0$  with  $\bigcap_1^\infty A_n = \emptyset$ , then  $\lim_{n \rightarrow \infty} P(A_n) = 0$ .
- (4b) If  $A_1, A_2, \dots \in \mathcal{F}_0$  are disjoint and  $\cup_{k=1}^\infty A_k \in \mathcal{F}_0$ , then

$$P(\cup_{k=1}^\infty A_k) = \sum_{k=1}^\infty P(A_k).$$

(4c) If  $A_1, A_2, \dots \in \mathcal{F}_0$  are disjoint and  $\cup_{k=1}^{\infty} A_k = \Omega$  then  $\sum_{k=1}^{\infty} P(A_k) = 1$ .

Then one can extend  $P$  to be defined, in one and only one way, for all sets in the  $\sigma$ -field  $\mathcal{F}$  generated by  $\mathcal{F}_0$ , so that it still satisfies the probability axioms.

We can now state and prove the existence of probability measures on the real line with a given distribution function.

**Theorem A:2** Let  $F(x)$  be a statistical distribution function on  $\mathbb{R}$ , i.e. a non-decreasing, right-continuous function with  $F(-\infty) = 0$ ,  $F(\infty) = 1$ . Then there exists exactly one probability measure  $P$  on  $(\mathbb{R}, \mathcal{B})$ , such that  $P((a, b]) = F(b) - F(a)$ .

**Proof:** We shall use the extension Theorem A:1. The Borel sets  $\mathcal{B}$  equal the  $\sigma$ -field generated by the field  $\mathcal{F}_0$  of unions of finitely many intervals. Equation (A.1) extended by singletons, defines  $P$  for each set in  $\mathcal{F}_0$ , and it is easily checked that properties (1), (2), and (3) hold. The only difficult part is (4a), which we prove by contradiction.

The idea is to use Cantor's theorem that every decreasing sequence of compact, non-empty sets, has a non-empty intersection. Assume, for a decreasing sequence of sets  $A_n \in \mathcal{F}_0$ , that  $P(A_n) \downarrow h > 0$ . We show that then the intersection of the  $A_n$ -sets is not empty. Each  $A_n$  consists of finitely many half open intervals. It is then possible to remove from  $A_n$  a short piece from the left end, to make it closed and bounded, i.e. there exists a compact, nonempty,  $K_n \subset A_n$ , such

$$P(A_n - K_n) \leq \varepsilon/2^n.$$

(Convince yourself that  $P(A_n - K_n)$  is defined.) Then

$$L_m = \cap_1^m K_n \subseteq K_m \subseteq A_m,$$

form a decreasing sequence,

$$L_1 \supseteq L_2 \supseteq \dots$$

If we can prove that the  $L_m$  can be taken nonempty, we can use Cantor's theorem, and conclude that they have a nonempty intersection, i.e. there exists at least one point  $x \in \cap_1^{\infty} L_m$ , which also implies  $x \in \cap_1^{\infty} A_m$ , so the  $A_k$  do not decrease to the empty set. The proof would be finished.

Hence, it remains to prove that we can choose each  $L_m$  nonempty. Take  $\varepsilon < h$ . Then

$$\begin{aligned} P(A_m - L_m) &= P(A_m - \cap_1^m K_n) = P(\cup_1^m (A_m - K_n)) \\ &\leq \sum_1^m P(A_m - K_n) \leq \sum_1^m P(A_n - K_n) \leq \sum_1^m \varepsilon/2^n \leq \varepsilon, \end{aligned}$$

$$P(L_m) = P(A_m) - P(A_m - L_m) \geq h - \varepsilon > 0,$$

and  $L_m$  is non-empty. □

### A.3 Kolmogorov's extension to $\mathbb{R}^\infty$

Kolmogorov's existence, or extension, theorem from 1933, allows us to define a stochastic process through its family of finite-dimensional distribution functions. Kolmogorov's book [20] appeared after a period of about 30 years attempts to give probability theory a solid mathematical foundation; in fact, Hilbert's sixth problem (1900) asked for a logical investigation of the axioms of probability.

**Theorem A:3** *Extension formulation: Every consistent family  $\{P_n\}$  of probability measures on  $(\mathbb{R}^n, \mathcal{B}_n)$ ,  $n = 1, 2, \dots$ , can be uniquely extended to a probability measure  $P$  on  $(\mathbb{R}^\infty, \mathcal{B}_\infty)$ , i.e. in such a way that*

$$\begin{aligned} P((a_1, b_1] \times (a_2, b_2] \times \dots \times (a_n, b_n] \times \mathbb{R}^\infty) \\ = P_n((a_1, b_1] \times (a_2, b_2] \times \dots \times (a_n, b_n]). \end{aligned}$$

*Existence formulation: To every consistent family of finite-dimensional distribution functions,  $\mathbf{F} = \{F_{\mathbf{t}^n}\}_{n=1}^\infty$ , there exists one and only one probability measure  $P$  on  $(\mathbb{R}^\infty, \mathcal{B}_\infty)$  with*

$$P(x_1 \leq b_1, \dots, x_n \leq b_n) = F_n(b_1, \dots, b_n).$$

**Proof:** We prove first the Extension formulation. We are given one probability measure  $P_n$  on each one of  $(\mathbb{R}^n, \mathcal{B}_n)$ ,  $n = 1, 2, \dots$ . Consider the intervals in  $\mathbb{R}^\infty$ , i.e. the sets of the form

$$I = \{\mathbf{x} = (x_1, x_2, \dots); x_i \in (a_i, b_i], i = 1, \dots, n\},$$

for some  $n$ , and unions of a finite number of intervals. Define  $P$  for each interval by

$$P(I) = P_n \left( \prod_1^n (a_i, b_i] \right).$$

Let  $I_1$  and  $I_2$  be two disjoint intervals. They may have different dimensions ( $n_1 \neq n_2$ ), but setting suitable  $a_i$  or  $b_i$  equal to  $\pm\infty$ , we may assume that they have the same dimension. The consistency of the family  $\{P_n\}$  guarantees that this does not change their probabilities, and that the additivity property holds, that is, if also  $I_1 \cup I_2$  is an interval, then  $P(I_1 \cup I_2) = P(I_1) + P(I_2)$ . It is easy to extend  $P$  with additivity to all finite unions of intervals. By this we have defined  $P$  on the field  $\mathcal{F}_0$  of finite unions of intervals, and checked that properties (1), (2), and (3) of Theorem A:1 hold.

Now check property (4a) in the same way as for Theorem A:2, for a decreasing sequence of non-empty intervals with empty intersection,

$$I_1 \supseteq I_2 \supseteq \dots, \quad \text{with} \quad \bigcap_1^\infty I_n = \emptyset,$$

and suppose  $P(I_n) \downarrow h > 0$ .<sup>1</sup> We can always assume  $I_n$  to have dimension  $n$ ,

$$I_n = \{\mathbf{x} \in \mathbb{R}^\infty; a_j^{(n)} < x_j \leq b_j^{(n)}, j = 1, \dots, n\},$$

and we can always assume the  $a_j$  and  $b_j$  to be bounded. As in the proof of Theorem A:2, remove a small piece of the lower side of each interval  $I_n$  to get a compact  $K_n$  and define  $L_m = \cap_1^m K_n$ . By removing a small enough piece one can obtain that  $P(L_m) \geq h/2 > 0$  so  $L_m$  is non-empty.

If we write

$$\begin{aligned} L_1 &: \alpha_1^{(1)} \leq x_1 \leq \beta_1^{(1)} \\ L_2 &: \alpha_1^{(2)} \leq x_1 \leq \beta_1^{(2)}, \alpha_2^{(2)} \leq x_2 \leq \beta_2^{(2)} \\ L_3 &: \alpha_1^{(3)} \leq x_1 \leq \beta_1^{(3)}, \alpha_2^{(3)} \leq x_2 \leq \beta_2^{(3)}, \alpha_3^{(3)} \leq x_3 \leq \beta_3^{(3)} \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \end{aligned}$$

For each  $j$ ,  $[\alpha_j^{(n)}, \beta_j^{(n)}]$ ,  $n = j, j+1, \dots$ , is a decreasing sequence of non-empty, closed and bounded intervals, and by Cantor's theorem they have at least one common point,  $\bar{x}_j \in \cap_{n=j}^\infty [\alpha_j^{(n)}, \beta_j^{(n)}]$ . Then,  $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots) \in \bar{L}_n$  for all  $n$ . Hence  $\bar{\mathbf{x}} \in L_n \subseteq I_n$  for all  $n$  and the intersection  $\cap_1^\infty I_n$  is not empty. This contradiction shows that  $P(I_n) \downarrow 0$ , and (3'') is shown to hold.

The conditions (1), (2), (3), and (4a) of Theorem A:1 are all satisfied, and hence  $P$  can be extended to the  $\sigma$ -field  $\mathcal{F}$  generated by the intervals.

To get the Existence formulation, just observe that the family of finite-dimensional distributions uniquely defines  $P_n$  on  $(\mathbb{R}^n, \mathcal{B}_n)$ , and use the extension.  $\square$

## Exercises

A:1. Let  $\mathbb{Z}$  be the integers, and  $\mathcal{A}$  the family of subsets  $A$ , such that either  $A$  or its complement  $A^c$  is finite. Let  $P(A) = 0$  in the first case and  $P(A) = 1$  in the second case. Show that  $P$  can not be extended to a probability to  $\sigma(\mathcal{A})$ , the smallest  $\sigma$ -field that contains  $\mathcal{A}$ .

---

<sup>1</sup>Property (4a) deals with a decreasing sequence of finite unions of intervals. It is easy to convince oneself that it suffices to show that (4a) holds for a decreasing sequence of intervals.



# Appendix B

## Stochastic convergence

Here we summarize the basic types of stochastic convergence and the ways we have to check the convergence of a random sequence with specified distributions.

**Definition B:1** Let  $\{x_n\}_{n=1}^{\infty}$  be a sequence of random variables  $x_1(\omega)$ ,  $x_2(\omega)$ , ... defined on the same probability space, and let  $x = x(\omega)$  be a random variable, defined on the same probability space. Then, the convergence  $x_n \rightarrow x$  as  $n \rightarrow \infty$  can be defined in three ways:

- **almost surely, with probability one** ( $x_n \xrightarrow{a.s.} x$ ):  $P(\{\omega; x_n \rightarrow x\}) = 1$ ;
- **in quadratic mean** ( $x_n \xrightarrow{q.m.} x$ ):  $E(|x_n - x|^2) \rightarrow 0$ ;
- **in probability** ( $x_n \xrightarrow{P} x$ ): for every  $\epsilon > 0$ ,  $P(|x_n - x| > \epsilon) \rightarrow 0$ .

Furthermore,  $x_n$  tends **in distribution** to  $x$ , (in symbols  $x_n \xrightarrow{\mathcal{L}} x$ ) if

$$P(x_n \leq a) \rightarrow P(x \leq a)$$

for all  $a$  such that  $P(x \leq u)$  is a continuous function of  $u$  at  $u = a$ .

### B.1 Criteria for convergence almost surely

In order for a random sequence  $x_n$  to converge almost surely (i.e. with probability one), to the random variable  $x$ , it is necessary and sufficient that

$$\lim_{m \rightarrow \infty} P(|x_n - x| > \delta \text{ for at least one } n \geq m) = 0 \quad (\text{B.1})$$

for every  $\delta > 0$ .

To prove this, note that if  $\omega$  is an outcome such that the real sequence  $x_n(\omega)$  does not converge to  $x(\omega)$ , then

$$\omega \in \left\{ \bigcup_{q=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} |x_n(\omega) - x(\omega)| > 1/q \right\}.$$

Here, the innermost event has probability

$$P(\bigcup_{n=m}^{\infty} |x_n(\omega) - x(\omega)| > 1/q) = P(|x_n - x| > 1/q \text{ for at least one } n \geq m),$$

and this is 0 for all  $q$  if and only if (B.1) holds for all  $\delta > 0$ . The reader should complete the argument, using that  $P(\cup_k A_k) = 0$  if and only if  $P(A_k) = 0$  for all  $k$ , and the fact that if  $B_1 \supseteq B_2 \supseteq \dots$  is a non-increasing sequence of events, then  $P(\cap_k B_k) = \lim_{k \rightarrow \infty} P(B_k)$ . Now,

$$P(|x_n - x| > \delta \text{ for at least one } n \geq m) \leq \sum_{n=m}^{\infty} P(|x_n - x| > \delta),$$

and hence a simple *sufficient* condition for (B.1) and a sufficient condition for almost sure convergence is that for all  $\delta > 0$ ,

$$\sum_{n=1}^{\infty} P(|x_n - x| > \delta) < \infty. \quad (\text{B.2})$$

(In fact, the first Borel-Cantelli lemma directly shows that (B.2) is sufficient for almost sure convergence.)

A simple moment condition is obtained from the inequality  $P(|x_n - x| > \delta) \leq E(|x_n - x|^h)/\delta^h$ , giving that a sufficient condition for almost sure convergence is

$$\sum_{n=1}^{\infty} E(|x_n - x|^h) < \infty, \quad (\text{B.3})$$

for some  $h > 0$ .

A Cauchy convergence type condition is the following: *sufficient condition for almost sure convergence*: if there exist two sequences of positive numbers  $\delta_n$  and  $\epsilon_n$  such that  $\sum_{n=1}^{\infty} \delta_n < \infty$  and  $\sum_{n=1}^{\infty} \epsilon_n < \infty$ , and such that

$$P(|x_{n+1} - x_n| > \delta_n) < \epsilon_n, \quad (\text{B.4})$$

then there exists a random variable  $x$  such that  $x_n \xrightarrow{a.s.} x$ .

To see this, use the Borel-Cantelli lemma to conclude that

$$P(|x_{n+1} - x_n| > \delta_n \text{ for infinitely many } n) = 0.$$

Thus, for almost all  $\omega$ , there is a number  $N$ , depending on the outcome  $\omega$ , such that

$$|x_{n+1} - x_n| < \delta_n \quad \text{for all } n \geq N.$$

Since  $\sum \delta_n < \infty$ , the sequence  $x_n(\omega)$  converges to a limit  $x(\omega)$  for these outcomes. For  $\omega$  where the limit does not exist, set  $x(\omega) = 0$ , for example. Then  $x_n \xrightarrow{a.s.} x$ , as was to be proved.

**B.1.0.1 Uniform convergence of random functions**

A sequence of random variables can converge almost surely, and we have just given sufficient conditions for this. But we shall also need convergence of a sequence of random functions  $\{x_n(t); t \in T\}$ , where  $T = [a, b]$  is a closed bounded interval.

**Definition B:2** *A sequence of functions  $\{x_n(t); a \leq t \leq b\}$  converges uniformly to the function  $\{x(t); a \leq t \leq b\}$  if*

$$\max_{a \leq t \leq b} |x_n(t) - x(t)| \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

*that is, if  $x_n$  lies close to the limiting function  $x$  in the entire interval  $[a, b]$  for all sufficiently large  $n$ .*

It is a basic result in real analysis that if a sequence of continuous functions converges uniformly in a closed and bounded interval, then the limiting function is also continuous. This fact will be useful when we show the almost sure sample function continuity of a random function.

Condition (B.4) can be restated to deal with almost sure uniform convergence of random functions: if there exist two sequences of positive numbers  $\delta_n$  and  $\epsilon_n$  such that  $\sum_{n=1}^{\infty} \delta_n < \infty$  and  $\sum_{n=1}^{\infty} \epsilon_n < \infty$ , and such that

$$P(\max_{a \leq t \leq b} |x_{n+1}(t) - x_n(t)| > \delta_n) < \epsilon_n, \quad (\text{B.5})$$

then there exists a random function  $x(t); a \leq t \leq b$ , such that  $x_n(t) \xrightarrow{a.s.} x(t)$  uniformly for  $t \in [a, b]$ .

**B.2 Criteria for convergence in quadratic mean**

Some of the representation theorems for stationary processes express a process as a complex stochastic integral, defined as a limit in quadratic mean of approximating sums of complex-valued random variables. To define a quadratic mean integral, or other limit of that kind, one needs simple convergence criteria for when  $x_n \xrightarrow{q.m.} x$  for a sequence of random variables with  $E(|x_n|^2) < \infty$ .

The Cauchy convergence criterion for convergence in quadratic mean states that a necessary and sufficient condition for that there exists a (possibly complex) random variable  $x$  such that  $x_n \xrightarrow{q.m.} x$  is that

$$E(|x_m - x_n|^2) \rightarrow 0, \quad (\text{B.6})$$

as  $n$  and  $m$  tend to infinity, independently of each other. (In mathematical language, this is the completeness of the space  $\mathcal{L}^2$ .)

The limit  $x$  has  $E(|x|^2) = \lim E(|x_n|^2) < \infty$ , and  $E(x_n) \rightarrow E(x)$ . If there are two convergent sequences,  $x_n \xrightarrow{q.m.} x$  and  $y_n \xrightarrow{q.m.} y$ , then

$$E(x_n \bar{y}_n) \rightarrow E(x \bar{y}). \quad (\text{B.7})$$

To show quadratic mean convergence of stochastic integrals, the following criterion is useful:

**the Loève criterion:** the sequence  $x_n$  converges in quadratic mean if and only if

$$E(x_m \bar{x}_n) \text{ has a finite limit } c, \quad (\text{B.8})$$

when  $m$  and  $n$  tend to infinity independently of each other.

The *if* part follows from  $E(|x_m - x_n|^2) = E(x_m \bar{x}_m) - E(x_m \bar{x}_n) - E(x_n \bar{x}_m) + E(x_n \bar{x}_n) \rightarrow c - c - c + c = 0$ . The *only if* part follows from  $E(x_m \bar{x}_n) \rightarrow E(x \bar{x}) = E(|x|^2)$ .

### B.3 Criteria for convergence in probability

Both almost sure convergence and convergence in quadratic mean imply convergence in probability. Further, if  $x_n \xrightarrow{P} x$  then there exists a subsequence  $n_k \rightarrow \infty$  as  $k \rightarrow \infty$ , such that  $x_{n_k} \xrightarrow{a.s.} x$ .

To prove this, we use criterion (B.2). Take any sequence  $\epsilon_k > 0$  such that

$$\sum_{k=1}^{\infty} \epsilon_k < \infty.$$

If  $x_n \xrightarrow{P} x$ , take any  $\delta > 0$  and consider  $P(|x_n - x| > \delta) \rightarrow 0$  as  $n \rightarrow \infty$ . The meaning of the convergence is that for each  $\epsilon_k$  there is an  $N_{\epsilon_k}$  such that

$$P(|x_n - x| > \delta) < \epsilon_k,$$

for all  $n \geq N_{\epsilon_k}$ . In particular, with  $n_k = N_{\epsilon_k}$ , one has

$$\sum_{k=1}^{\infty} P(|x_{n_k} - x| > \delta) < \sum_{k=1}^{\infty} \epsilon_k,$$

which is finite by construction. The sufficient criterion (B.2) gives the desired almost sure convergence of the subsequence  $x_{n_k}$ .

### Exercises

B:1. Prove the Borel-Cantelli lemma:

a) If  $A_k$  are events in a probability space  $(\Omega, \mathcal{F}, P)$ , then

$$\sum_k P(A_k) < \infty,$$

implies  $P(A_k \text{ infinitely often}) = 0$ .

b) If the events  $A_k$  are independent, then

$$\sum_k P(A_k) = \infty,$$

implies  $P(A_k \text{ infinitely often}) = 1$ .

B:2. Let  $x_1, x_2, \dots$  be independent identically distributed random variables. Show that

$$E(|x_k| < \infty \text{ if and only if } P(|x_k| > k \text{ infinitely often}) = 0.$$

B:3. Suppose the random sequences  $x_n$  and  $x'_n$  have the same distribution. Prove that if  $x_n \xrightarrow{a.s.} x$  then there exists a random variable  $x'$  such that  $x'_n \xrightarrow{a.s.} x'$ .



# Appendix C

## Hilbert space and random variables

### C.1 Hilbert space and scalar products

A Hilbert space is a set of elements which can be added and multiplied by complex numbers, and for which there is defined an *inner product*. The inner product in a Hilbert space has the same mathematical properties as the covariance between two random variables with mean zero, and therefore it is natural to think of random variables as elements in a Hilbert space. We summarize here the basic properties of a Hilbert space, for use in Chapters 3 and 4. For further reading on Hilbert spaces and on metric spaces, see e.g. the classical book by Royden [29].

**Definition C:1** *A general Hilbert space  $\mathcal{H}$  over the complex numbers  $\mathbb{C}$  is a set of elements, usually called points or vectors, with the following properties:*

1. *The operations addition and subtraction are defined, and there exists a unique "zero" element  $\mathbf{0} \in \mathcal{H}$  and to each  $x \in \mathcal{H}$  there is a unique inverse  $-x$ :*

$$\begin{aligned}x + y &= y + x \in \mathcal{H}, \\x + \mathbf{0} &= x, \\x + (-x) &= \mathbf{0}.\end{aligned}$$

2. *Multiplication with complex scalar is defined (usually written  $cx = c \cdot x$ ):*

$$\begin{aligned}c \cdot x &\in \mathcal{H}, \\0 \cdot x &= \mathbf{0}, \\1 \cdot x &= x.\end{aligned}$$

3. A scalar (inner) product  $(x, y)$  is defined such that:

$$\begin{aligned}(x, y) &= \overline{(y, x)} \in \mathbb{C}, \\ (ax + by, z) &= a(x, z) + b(y, z), \\ (x, x) &\geq 0, \\ (x, x) &= 0 \text{ if and only if } x = \mathbf{0}.\end{aligned}$$

4. A norm  $\|x\|$  and a distance  $d(x, y) = \|x - y\|$  are defined, and convergence has the standard meaning: if  $x \in \mathcal{H}$  then  $\|x\| = (x, x)^{1/2}$ , and if  $x_n, x \in \mathcal{H}$  then  $\lim_{n \rightarrow \infty} x_n = x$  if and only if  $\|x_n - x\| \rightarrow 0$ .

5. The space is complete in the sense that if  $x_n, x \in \mathcal{H}$  and  $\|x_m - x_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$  then there is a point  $x \in \mathcal{H}$  such that  $\lim_{n \rightarrow \infty} x_n = x$ .

**Remark C:1** If  $\mathcal{H}$  is a space that satisfies (1-3) in the definition, then it can be completed and made a Hilbert space that satisfies also (5).

We list some further properties of Hilbert spaces and scalar products, which will be seen to have parallels as concepts for random variables:

**Schwarz inequality:**  $|(x, y)| \leq \|x\| \cdot \|y\|$  with equality if and only if  $(y, x)x = (x, x)y$ ,

**Triangle inequality:**  $\|x + y\| \leq \|x\| + \|y\|$ ,

**Continuity:** if  $x_n \rightarrow x$  and  $y_n \rightarrow y$  then  $(x_n, y_n) \rightarrow (x, y)$ ,

**Pythagorean theorem:** if  $x$  and  $y$  are orthogonal, i.e.  $(x, y) = 0$ , then

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

### C.1.0.2 Linear subspaces:

Let  $L = \{x_j \in \mathcal{H}; j = 1, 2, \dots\}$  be a set of elements in a Hilbert space  $\mathcal{H}$ , and let

$$M_0 = \{a_1x_1 + \dots + a_kx_k; k = 1, 2, \dots; a_j \in \mathbb{C}\}$$

be the family of all finite linear combinations of elements in  $L$ . Then

$$\mathcal{M} = \overline{M_0} = \mathcal{S}(L) = \left\{ x \in \mathcal{H}; x = \lim_{n \rightarrow \infty} x_n \text{ for some } x_n \in M_0 \right\}$$

is called *the subspace of  $\mathcal{H}$  spanned by  $L$* . It consists of all elements in  $\mathcal{H}$  which are linear combinations of elements in  $L$  or are limits of such linear combinations. It is a *subspace* in the sense that it is closed under addition, multiplication by scalar, and passage to a limit.

## C.2 Projections in Hilbert space

Two elements in a Hilbert space are called *orthogonal*, written  $x \perp y$ , if  $(x, y) = 0$ . Two subsets  $L_1$  and  $L_2$  are said to be orthogonal,  $L_1 \perp L_2$ , if all elements  $x \in L_1$  are orthogonal to all elements  $y \in L_2$ . Similarly, two subspaces  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are orthogonal,  $\mathcal{M}_1 \perp \mathcal{M}_2$ , if all elements in  $\mathcal{M}_1$  are orthogonal to all elements in  $\mathcal{M}_2$ . The reader should check that if  $L_1 \perp L_2$ , then  $\mathcal{S}(L_1) \perp \mathcal{S}(L_2)$ .

For a sequence of subspaces,  $\mathcal{M}_1, \dots, \mathcal{M}_k$  of  $\mathcal{H}$ , write

$$V = \mathcal{M}_1 \oplus \dots \oplus \mathcal{M}_k$$

for the *vector sum* of  $\mathcal{M}_1, \dots, \mathcal{M}_k$ , which is the set of all vectors  $x_1 + \dots + x_k$ , where  $x_j \in \mathcal{M}_j$ , for  $j = 1, \dots, k$ .

### C.2.0.3 The projection theorem

Let  $\mathcal{M}$  be a subspace of a Hilbert space  $\mathcal{H}$ , and let  $x$  be a point in  $\mathcal{H}$  not in  $\mathcal{M}$ . Then  $x$  can be written in exactly one way as a sum

$$x = y + z$$

with  $y \in \mathcal{M}$  and  $z = (x - y) \perp \mathcal{M}$ . Furthermore,  $y$  is the point in  $\mathcal{M}$  which is closest to  $x$ ,

$$d(x, y) = \min_{w \in \mathcal{M}} d(x, w).$$

and equality holds if and only if  $w = y$ .

The most common use of the projection theorem is to approximate a point  $x$  in a general Hilbert space by a linear combination, or a limit thereof, of a finite or infinite number of certain elements in  $\mathcal{H}$ .

### C.2.0.4 Separable spaces and orthogonal bases

A Hilbert space  $\mathcal{H}$  is called *separable* if it contains a countable set of elements  $x_1, x_2, \dots$  such that the subspace spanned by all the  $x_j$  is equal to  $\mathcal{H}$ . If the  $x$ -variables are *linearly independent*, i.e. there is no non-trivial linear combination equal to  $\mathbf{0}$ ,  $a_1x_1 + \dots + a_nx_n = \mathbf{0}$ , it is possible to find *orthogonal* elements  $y_1, y_2, \dots$ , such that

$$\begin{aligned} y_1 &= c_{11}x_1, \\ y_2 &= c_{21}x_1 + c_{22}x_2, \\ &\dots \\ y_n &= c_{n1}x_1 + c_{n2}x_2 + \dots + c_{nn}x_n, \\ &\dots \end{aligned}$$

This is the *Gram-Schmidt orthogonalization process*, and *orthogonal* means that

$$(y_j, y_k) = \delta_{jk} = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

The sequence  $y_1, y_2, \dots$  is called a *complete orthogonal basis* for the Hilbert space  $\mathcal{H}$ . It is a basis, i.e. every element in  $\mathcal{H}$  can be written as a linear combination of  $y_k$ -elements or as a limit of such combinations, and it is orthogonal by construction. It is furthermore complete, i.e. there is no element  $z \in \mathcal{H}$  such that

$$\|z\| > 0, \quad (z, y_j) = 0, \quad \text{for all } j.$$

### C.3 Stochastic processes and Hilbert spaces

A Hilbert space is a set of elements which can be added and multiplied by complex numbers, and for which there is defined an *inner product*. The inner product in a Hilbert space has the same mathematical properties as the covariance between two random variables with mean zero, and therefore it is natural to think of random variables as elements in a Hilbert space; see Appendix C for a summary of elementary properties of Hilbert spaces.

We shall consider a very special Hilbert space, namely the space of all random variables  $X$  on a probability space  $(\Omega, \mathcal{F}, P)$ , which have zero mean and finite variance.

**Theorem C:1** *If  $(\Omega, \mathcal{F}, P)$  is a probability space, then*

$$\mathcal{H} = \{\text{random variables } x \text{ on } (\Omega, \mathcal{F}, P) \text{ such that } E(x) = 0, E(|x|^2) < \infty\}$$

*with the scalar product*

$$(x, y) = E(x\bar{y})$$

*is a Hilbert space; it will be denoted  $\mathcal{H}(\Omega)$ .*

First, it is clear that  $(x, y) = E(x\bar{y})$  has the properties of a scalar product; check that. It is also clear that we can add random variables with mean zero and finite variance to obtain new random variables with the same properties. Also,  $\|x\| = \sqrt{E(|x|^2)}$ , which means that if  $\|x\| = 0$ , then  $P(x = 0) = 1$ , so random variables which are zero with probability one, are, in this context, defined to be equal to the zero element  $\mathbf{0}$ . Convergence in the norm  $\|\cdot\|$  is equal to convergence in quadratic mean of random variables, and if a sequence of random variables  $x_n$  is a Cauchy sequence, i.e.  $\|x_m - x_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$ , then we know that it converges to a random variable  $x$  with finite mean, which means that  $\mathcal{H}(\Omega)$  is complete. Therefore it has all the properties of a Hilbert space.

**C.3.0.5 A stochastic process as a curve in  $\mathcal{H}(\Omega)$** 

A random variable with  $E(x) = 0$  and finite variance is a point in the Hilbert space  $\mathcal{H}(\Omega)$ . Two equivalent random variables  $x$  and  $y$  are represented by the same point in  $\mathcal{H}(\Omega)$ , since  $P(x = y) = 1$  and hence  $\|x - y\|^2 = E(|x - y|^2) = 0$ .

A stochastic process is a family of random variables, and thus a stochastic process  $\{x(t), t \in \mathbb{R}\}$  with one-dimensional parameter  $t$  is a *curve* in  $\mathcal{H}(\Omega)$ . Further, from the definition of the norm  $\|x\| = \sqrt{E(|x|^2)}$ , we see that convergence in this norm is equivalent to convergence in quadratic mean. In other words, if a stochastic process is continuous in quadratic mean, then the corresponding curve in  $\mathcal{H}(\Omega)$  is continuous.

**C.3.0.6 The generated subspace**

A set of points in a Hilbert space generates a subspace, which consists of all finite linear combinations and their limits. If  $\{x(t); t \in T\}$  is a stochastic process, write

$$\mathcal{H}(x) = \mathcal{S}(x)$$

for the subspace spanned by  $x(\cdot)$ . Also, for a process  $\{x(t), t \in \mathbb{R}\}$ , define

$$\mathcal{H}(x, t) = \mathcal{S}(x(s); s \leq t)$$

as the subspace spanned by all variables observed up till time  $t$ . At time  $t$  it contains all variables which can be constructed by linear operations on the available observations. Examples of random variables in  $\mathcal{H}(x, t)$  are

$$\frac{x(t) + x(t-1) + \dots + x(t-n+1)}{n}, \quad \int_{-\infty}^t e^{-(t-u)} x(u) du, \quad x'_-(t) + 3x''_-(t),$$

where  $x'_-(t), x''_-(t)$  denote left derivatives.

**Example C:1** Take an MA(1)-process, i.e. from uncorrelated variables

$$e(t), t = \dots, -1, 0, 1, 2, \dots,$$

with  $E(e(t)) = 0$ ,  $V(e(t)) = 1$ , we construct

$$x(t) = e(t) + b_1 e(t-1).$$

If  $|b_1| < 1$ , the process can be inverted and  $e(t)$  simply retrieved from  $x(s), s \leq t$ :

$$\begin{aligned} e(t) &= x(t) - b_1 e(t-1) = x(t) - b_1(x(t-1) - b_1 e(t-2)) \\ &= \sum_{k=0}^n (-b_1)^k x(t-k) + (-b_1)^{n+1} e(t-n-1) = y_n(t) + z_n(t), \text{ say.} \end{aligned}$$

Here,

$$y_n(t) \in \mathcal{S}(x(s); s = t - n, \dots, t) \subseteq \mathcal{S}(x(s); s \leq t) = \mathcal{H}(x, t),$$

while

$$\|z_n(t)\| = |b_1|^{n+1} \rightarrow 0$$

as  $n \rightarrow \infty$ . Thus  $e(t) - y_n(t) \rightarrow 0$  and we have that

$$e(t) = \sum_{k=0}^{\infty} (-b_1)^k x(t-k) = \lim_{n \rightarrow \infty} \sum_{k=0}^n (-b_1)^k x(t-k) \in \mathcal{H}(x, t)$$

if  $|b_1| < 1$ . The representation of  $e(t)$  as a limit of finite linear combinations of  $x(t-k)$ -values is explicit and obvious.

For  $|b_1| = 1$  it is less obvious that  $e(t) \in \mathcal{H}(x, t)$ , but it is still possible to represent  $e(t)$  as a limit. For example, if  $b_1 = -1$ ,  $x(t) = e(t) - e(t-1)$ , and  $z_n(t) = e(t-n-1)$  does not converge to anything. But in any case,

$$e(t) = \sum_{k=0}^n x(t-k) + e(t-n-1),$$

and so, since the left hand side does not depend on  $n$ ,

$$\begin{aligned} e(t) &= \frac{1}{N} \sum_{n=1}^N e(t) = \frac{1}{N} \sum_{n=1}^N \sum_{k=0}^n x(t-k) + \frac{1}{N} \sum_{n=1}^N e(t-n-1) \\ &= \sum_{k=0}^N \left(1 - \frac{k}{N}\right) x(t-k) + \frac{1}{N} \sum_{n=1}^N e(t-n-1) = y_N(t) + z_N(t). \end{aligned}$$

Now,  $z_N(t) = \frac{1}{N} \sum_{n=1}^N e(t-n-1) = e(t) - y_N(t) \rightarrow 0$  by the law of large numbers, since all  $e(t)$  are uncorrelated with  $E(e(t)) = 0$  and  $V(e(t)) = 1$ . We have shown that  $e(t)$  is in fact the limit of a finite linear combination of  $x(s)$ -variables, i.e.  $e(t) \in \mathcal{H}(x, t)$ .

## Appendix D

# Spectral simulation of random processes

### D.1 The Fast Fourier Transform, FFT

A stationary process  $\{x(t), t \in \mathbb{R}\}$  with continuous spectrum  $f(\omega)$  can be efficiently simulated by Fourier methods from its spectral representation. One then has to discretize the continuous spectrum and use the approximation (4.29) from Section 4.3.3.

Fourier-simulation is most effectively performed with the help of the Fast Fourier Transform (FFT), or rather the inverse transform. This algorithm transforms a sequence of real or complex numbers  $Z(0), Z(1), \dots, Z(N-1)$  into its (inverse) discrete Fourier transform

$$z(n) = \sum_{k=0}^{N-1} Z(k) \exp(i2\pi kn/N), \quad (\text{D.1})$$

for  $n = 0, 1, \dots, N-1$ , where the integer  $N$  is a power of 2,  $N = 2^m$ . In the literature, there are as many ways to write the Fourier sum as there are combinatorial possibilities, with or without a factor  $N$  in the denominator and with or without a minus-sign in the exponential function. Almost every mathematical computer software toolbox contains efficient algorithms to perform the FFT according to (D.1).

The basis for the use of (D.1) to generate a sample sequence lies in the representation of a stationary process as an approximating sum of harmonic functions with random phase and amplitude; see (4.29) and the alternative form (4.30). The  $Z(k)$  will then be chosen as complex random variables with absolute value and argument equal to the desired amplitude and phase. When using the formula for simulation purposes, there are however a number of questions which have to be resolved, concerning the relation between the sampling interval and the frequency resolution, as well as the aliasing problem.

Before we describe the steps in the simulation we repeat the basic facts about processes with discrete spectrum, and the special problems that arise when sampling a continuous time process.

## D.2 Random phase and amplitude

To see how (D.1) can be used to generate a sample function we consider first the special stationary process (4.18) with discrete spectrum in Section 4.3.3, or the normalized form (4.30). Including the spectral jump at zero frequency it has the form,

$$x(t) = \rho_0 + \sum_{k=1}^{\infty} \rho_k \cos(\omega_k t + \phi_k). \quad (\text{D.2})$$

Here  $\rho_0$  is a random level shift, while  $\{\rho_k\}$  are the amplitudes and  $\{\phi_k\}$  the phases of the different harmonic components of  $x(t)$ . The frequencies  $\omega_k > 0$  can be any set of fixed positive frequencies.

If we define

$$\begin{aligned} Z(0) &= \rho_0, \\ Z(k) &= \rho_k \exp(i\phi_k), \quad \text{for } k = 1, 2, \dots \end{aligned}$$

it is easy to see that  $x(t)$  in (D.2) is the real part of a complex sum, so if we write  $y(t)$  for the imaginary part, then

$$x(t) + iy(t) = \sum_{k=0}^{\infty} Z(k) \exp(i\omega_k t). \quad (\text{D.3})$$

We repeat the fundamental properties of this representation.

If amplitudes and phases in (D.2) are independent and the phases  $\phi_k$  are uniformly distributed over  $[0, 2\pi)$ , then  $\{x(t), t \in \mathbb{R}\}$  is stationary and has a discrete spectral distribution with mass  $\sigma_k^2 = \frac{1}{2}E(\rho_k^2)$  and  $\sigma_0^2 = E(\rho_0^2)$  at the frequencies  $\omega_k > 0$ , and  $\omega_0 = 0$ , respectively. Further,  $Z(k) = \rho_k \exp(i\phi_k) = \sigma_k(U_k + iV_k)$  have the desired properties if the real and imaginary parts are independent standardized Gaussian random variables, with  $E(U_k) = E(V_k) = 0$  and variance  $V(U_k) = V(V_k) = 1$ .

It is possible to approximate every spectral distribution by a discrete spectrum. The corresponding process is then an approximation of the original process.

## D.3 Aliasing

If a stationary process  $\{x(t), t \in \mathbb{R}\}$  with continuous twosided spectral density  $f_x(\omega)$ , is sampled with a sampling interval  $d$ , the sequence  $\{x(nd), n =$

$0, \pm 1, \dots\}$  has a spectral density  $f_x^{(d)}(\omega)$  that can be restricted to any interval of length  $\pi/d$ , for example the interval  $(-\pi/d, \pi/d]$ . There it can be written as a folding of the original spectral density,

$$f_x^{(d)}(\omega) = \sum_{j=-\infty}^{\infty} f_x\left(\omega + \frac{2\pi j}{d}\right), \quad \text{for } -\pi/d < \omega \leq \pi/d.$$

The corresponding one-sided spectral density  $g_x^{(d)}(\omega)$  can then be defined on  $[0, \pi/d]$  as

$$f_x^{(d)}(\omega) + f_x^{(d)}(-\omega).$$

For reasons that will become clear later (Section D.6) we prefer to define it instead on  $[0, 2\pi/d)$  by

$$g_x^{(d)}(\omega) = \sum_{j=-\infty}^{\infty} f_x\left(\omega + \frac{2\pi j}{d}\right), \quad \text{for } 0 \leq \omega < 2\pi/d. \quad (\text{D.4})$$

## D.4 Simulation scheme

In view of (D.1) and (D.3) we would like to generate a finite part of the sum in (D.3) to get  $z(n)$  and then take the real part to get  $x(t)$  for  $t = nd$ ,  $n = 0, 1, \dots, N-1$ . To see the analogy clearly we repeat the expressions:

$$z(n) = \sum_{k=0}^{N-1} Z(k) \exp(i2\pi kn/N) \quad (\text{D.5})$$

$$x(t) = \Re \sum_{k=0}^{\infty} Z(k) \exp(i\omega_k t) \quad (\text{D.6})$$

Here is the scheme to follow.

**We have:** A real spectral density  $g_x(\omega)$  for  $\omega \geq 0$  for a stationary process  $\{x(t), t \in \mathbb{R}\}$ .

**We want:** A discrete time sample  $x(nd)$ ,  $n = 0, 1, \dots, N-1$  of  $\{x(t), t \in \mathbb{R}\}$  of size  $N = 2^m$  with sampling interval  $d$ , equally spaced over the time interval  $[0, T)$  with  $T = Nd$ ;

**Means:** Generate random variables  $Z(k) = \sigma_k(U_k + iV_k)$ ,  $k = 0, 1, \dots, N-1$ , with distribution described below, and take  $z(n) = \sum_{k=0}^{N-1} Z(k) \exp(i2\pi kn/N)$  according to (D.1). Then set

$$x(nd) = \Re z(n), \quad n = 0, 1, \dots, N-1.$$

This will give the desired realization.

## D.5 Difficulties and details

The Fourier simulation scheme rises a number of questions which have to be dealt with before it can be implemented. Here we shall comment on the important issues.

**Frequency spacing:** We have requested  $N$  time points regularly spaced in  $[0, T)$  in steps of  $d$ , and we want to use the special sum (D.1). This will impose a restriction both on the frequency spacing and on the maximum frequency accounted for. Comparing (D.5) and (D.6), bearing in mind that  $t = nd$ , we find that only frequencies that are of the form

$$\omega_k = \frac{2\pi k}{Nd} = \frac{2\pi k}{T} \quad \text{for } k = 0, 1, \dots, N-1,$$

appear in the simulation, and further that the highest frequency in the sum is  $\frac{2\pi(N-1)}{dN}$ , just barely below

$$\omega_{\max} = \frac{2\pi}{d}.$$

**Discretization of spectrum:** The continuous spectrum with density  $g_x(\omega)$  has to be replaced by a discrete spectrum with mass only at the frequencies  $\omega_k = \frac{2\pi k}{dN}$  which enter into the sum (D.5). The mass at  $\omega_k$  should be equal to

$$\sigma_k^2 = \frac{2\pi}{dN} g_x^{(d)}(\omega_k), \quad k = 0, 1, \dots, N-1. \quad (\text{D.7})$$

**Generation of the  $Z(\mathbf{n})$ :** Generate independent random variables

$$Z(k) = \sigma_k(U_k + iV_k)$$

with  $U_k$  and  $V_k$  from a normal distribution with mean zero and variance 1, for instance by the Box-Müller technique,

$$\begin{aligned} U_k &= \cos(2\pi R_1) \sqrt{-2 \ln R_2}, \\ V_k &= \sin(2\pi R_1) \sqrt{-2 \ln R_2}, \end{aligned}$$

where  $R_1$  and  $R_2$  are independent random numbers uniformly distributed in  $(0, 1]$ .

**Aliasing:** The restricted frequency range in (D.5), implies that the generated  $x(nd)$  will have variance  $\sum_{k=0}^{N-1} \sigma_k^2$ , where each  $\sigma_k^2$  is an infinite sum:

$$\sigma_k^2 = \frac{2\pi}{dN} \sum_{j=-\infty}^{\infty} f_x\left(\omega_k + \frac{2\pi j}{d}\right).$$

In practice one has to truncate the infinite series and use

$$\sigma_k^2 = \frac{2\pi}{dN} \sum_{j=-J}^J f_x\left(\omega_k + \frac{2\pi j}{d}\right), \quad k = 0, 1, \dots, N-1, \quad (\text{D.8})$$

where  $J$  is taken large enough. If  $f_x(\omega) \approx 0$  for  $\omega \geq \omega_{\max}$  one can take  $J = 0$ .

## D.6 Simulation of the envelope

The Fourier simulation will not only yield a realization of  $x(nd) = \Re z(n)$  but also of its Hilbert transform  $y(nd) = \Im z(n)$ . Therefore we can get the envelope as a byproduct,

$$\sqrt{x(nd)^2 + y(nd)^2}.$$

Thus generation of  $2N$  Gaussian random numbers  $U_k, V_k$ , for  $k = 0, 1, \dots, N-1$  will result in  $2N$  useful data points. If the aim is to generate only the  $x(nd)$ -series, one could restrict the sum (D.5) to only  $n = 0, 1, \dots, N/2 - 1$  and thus generate only  $N$  Gaussian variates.

## D.7 Summary

In order to simulate a sample sequence of a stationary process  $\{x(t), t \in \mathbb{R}\}$  with spectral density  $f_x(\omega)$  over a finite time interval one should do the following:

1. Choose the desired time interval  $[0, T)$ .
2. Choose a sampling interval  $d$  or the number of sample points  $N = 2^m$ . This will give a sequence of  $N$  process values  $x(nd)$ ,  $k = 0, 1, \dots, N-1$ .
3. Calculate and truncate the real discretized spectrum

$$\sigma_k^2 = \frac{2\pi}{dN} \sum_{j=-J}^J f_x\left(\omega_k + \frac{2\pi j}{d}\right), \quad k = 0, 1, \dots, N-1,$$

and take  $J$  so large that  $f_x(\omega) \approx 0$  for  $\omega > 2\pi(J+1)/d$ .

4. Generate independent standard normal variables

$$U_k, V_k, \quad \text{for } k = 0, 1, \dots, N-1$$

with mean zero and variance 1.

5. Set  $Z(k) = \sigma_k(U_k + iV_k)$  and calculate the (inverse) Fourier transform

$$z(n) = \sum_{k=0}^{N-1} Z(k) \exp(i2\pi kn/N).$$

6. Take the real part,

$$x(nd) = \Re z(n), \quad n = 0, 1, \dots, N-1;$$

this is the desired sequence.

7. To generate the envelope, take the imaginary part

$$y(nd) = \Im z(n), \quad n = 0, 1, \dots, N - 1;$$

the envelope is then

$$\sqrt{x(nd)^2 + y(nd)^2}.$$

# Literature

- [1] Adler, R. (1990): *An introduction to continuity, extrema, and related topics for general Gaussian processes*. IMS Lecture Notes-Monograph Series, Vol. 12.
- [2] Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004): *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC, Boca Raton. indexCarlin, B.P.
- [3] Belyaev, Yu.K. (1959): Analytic random processes. *Theory Probab. and its Applications*, English edition, **4**, 402.
- [4] Belyaev, Yu.K. (1961): Continuity and Hölder's conditions for sample functions of stationary Gaussian processes. *Proc. Fourth Berk. Symp. on Math. Stat. and Probability*, **2**, 22-33.
- [5] Breiman, L. (1968): *Probability*. Addison-Wesley, Reading. Reprinted 199? in SIAM series.
- [6] Cramér, H. (1942): On harmonic analysis in certain function spaces. *Arkiv Mat. Astron. Fysik*, 28B, no. 12.
- [7] Cramér, H. (1945): On the theory of stochastic processes. *Proc. Tenth Scand. Congr. of Math.*, Copenhagen, pp. 28-39.
- [8] Cramér, H. (1945): *Mathematical Methods of Statistics*. princeton University Press, 1945..
- [9] Cramér, H. and Leadbetter, M.R. (1967): *Stationary and related stochastic processes*. Wiley, New York. Reprinted by Dover Publications, 2004.
- [10] Dobrushin, R.L. (1960): Properties of the sample functions of a stationary Gaussian process. *Teoriya Veroyatnostei i ee Primeneniya*, **5**, 132-134.
- [11] Doob, J.L. (1953): *Stochastic processes*. Wiley, New York.
- [12] Durrett, R. (1996): *Probability: Theory and examples*. Doxbury Press.
- [13] Einstein, A. (1905): *Investigations on the theory of Brownian novement*. Reprinted by Dover Publications, 1956.

- 
- [14] Jordan, D.W. and Smith, P. (1999): *Nonlinear ordinary differential equations*. 3rd Ed. Oxford University Press.
- [15] Grenander, U. (1950): Stochastic processes and statistical inference. *Arkiv Mat.* **1**, 195-277.
- [16] Gut, A. (1995): *An intermediate course in probability*. Springer-Verlag.
- [17] Ibragimov, I.A. and Linnik, Yu.V. (1971): *Independent and stationary sequences of random variables*. Wolters-Noordhoff, Groningen.
- [18] Ibragimov, I.A. and Rozanov, Y.A. (1978): *Gaussian random processes*. Springer-Verlag, New York.
- [19] Kac, M. and Slepian, D. (1959): Large excursions of Gaussian processes. *Ann. Math. Statist.*, **30**, 1215–1228.
- [20] Kolmogorov, A. (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag, Berlin.
- [21] Lasota, A. and Mackey, M.C. (1994): *Chaos, fractals, and noise; stochastic aspects of dynamics*. Springer-Verlag, New York.
- [22] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983): *Extremes and related properties of random sequences and processes*. Springer-Verlag, New York.
- [23] Lindgren, G. (1975): Prediction from a random time point. *Annals of Probability*, **3**, 412–423.
- [24] Maruyama, G. (1949): The harmonic analysis of stationary stochastic processes. *Mem. Fac. Sci. Kyusyu Univ.* **A4**, 45-106.
- [25] Petersen, K. (1983): *Ergodic Theory*, Cambridge University Press, Cambridge.
- [26] von Plato, J. (1994): *Creating modern probability*. Cambridge University Press, Cambridge.
- [27] Rice, S.O. (1944, 1945): Mathematical analysis of random noise. *Bell Systems technical Journal*, **23**, 282-332, and **24**, pp 46-156. Reprinted in: Wax, N. (1954): *Selected papers on noise and stochastic processes*. Dover Publications, New York.
- [28] Rice, S.O. (1963): Noise in FM-receivers. In: *Time Series Analysis*, Ed: M. Rosenblatt, Chapter 25, pp. 395-422. Wiley, New York.
- [29] Royden, H.L. (1988): *Real Analysis*, 3rd Ed. Prentice Hall.

- 
- [30] Rychlik, I. (2000): On some reliability applications of Rice's formula for the intensity of level crossings. *Extremes*, **3**, 331–348.
- [31] Slepian, D. (1963): On the zeros of Gaussian noise. In: *Time Series Analysis*, Ed: M. Rosenblatt, pp. 104-115. Wiley, New York.
- [32] St. Denis, M. and Pierson, W.J. (1954): On the motion of ships in confused seas. *Transactions, Soc. Naval Architects and Marine Engineers*, Vol. 61, (1953) pp. 280-357.
- [33] van Trees, H. (1968): *Detection, Estimation, and Modulation Theory. Part I*, John Wiley & Sons.
- [34] WAFO, Wave Analysis for Fatigue and Oceanography. Available at <http://www.maths.lth.se/matstat/wafo/>.
- [35] Wong, E. and Hajek, B. (1985): *Stochastic processes in engineering systems*. Springer-Verlag, New York.
- [36] Williams, D. (1991): *Probability with Martingales*, Cambridge University Press, Cambridge.
- [37] Williams, D. (2001): *Weighting the Odds*. Cambridge University Press, Cambridge.
- [38] Yaglom, A.M. (1962): *An introduction to the theory of stationary random functions*. Prentice Hall, Englewood Cliffs.
- [39] B. Øksendal: *Stochastic differential equations, an introduction with applications*. Springer-Verlag, 5th Ed 2000.

# Index

- $\alpha$ 
  - spectral width parameter, 63, 77
- $\epsilon$ 
  - spectral width parameter, 77
- $\omega$ 
  - angular frequency, 2, 14
  - elementary outcome, 2
- $\sigma$ -algebra, 3, 175
- $\sigma$ -field, 3, 175
  - generated by events, 3
  - generated by random variables, 6
  - of invariant sets, 139
- Øksendal, B., 201
- additive
  - countably, 4, 175
  - finitely, 4
- Adler, R., 199
- alarm prediction, 64
- algebra, *see*  $\sigma$ -algebra
- aliasing, 88, 194, 196
- amplitude, 89, 94
  - for narrow band process, 118
- ARMA-process, 152
- auto-covariance function, 161
- auto-regressive process, 152
  
- Banach's theorem, 59
- band-limited white noise, 151
- Banerjee, S., 199
- Belyaev, Yu.K., 36, 153, 199
- Bennet, W.R., 97
- Bessel function, 167
  - modified, 169
- Birkhoff ergodic theorem, 142
- Bochner's theorem, 82
  
- Boltzmann, L., 133
- Boole's inequality, 32
- Borel
  - field in  $\mathbb{C}$ , 37
  - field in  $\mathbb{R}^\infty$ , 9
  - field in  $\mathbb{R}^T$ , 11
  - field in  $\mathbb{R}$ , 3
  - measurable function, 6
  - set, 3
- Borel-Cantelli lemma, 25, 32, 53, 184
- Box-Müller's method, 196
- Breiman, L., 199
- Brown, R., ix, 19
- Brownian motion, ix, 19, 112
- Bulinskaya, E.V., 43
- Bulinskaya, E.V., 56
  
- Cantor's theorem, 177, 179
- Carathéodory's extension theorem, 176
- Cauchy principal value, 86
- characteristic equation, 108
- characteristic function, 83
- Chebysjev's inequality, 34
- complete
  - basis, 123, 190
  - space, 92, 188
- completion
  - of probability measure, 5
  - of vector space, 188
- conditional covariance
  - normal distribution, 17, 69
- conditional expectation, 7, 64, 141
  - horizontal window, 66
  - normal distribution, 17, 69
  - vertical window, 66
- consistent family of distributions, 8

- continuity
  - absolute, 13
  - conditions for, 33
  - in  $L^2$ , 27, 45, 90
  - in quadratic mean, 27, 45, 90
  - modulus, 41
  - of Gaussian process, 35
  - of sample function, 30
  - of stationary process, 34, 35
  - probabilities on  $\mathbb{C}$ , 37
  - summary of conditions, 49
- convergence
  - almost surely, 27, 41, 181
  - Cauchy condition, 182
  - in distribution, 181
  - in probability, 27, 181
  - in quadratic mean, 27, 181, 190
  - Loève criterion, 184
  - uniform, of random functions, 183
- countable
  - additivity, 10
  - basis, 11
- covariance function, 13
- Cramér, H., 21, 24, 151, 199
- Cramér-Wold decomposition, 151, 190
- cross spectrum, 161
- cross-covariance function, 161
- cross-spectrum, 162
- crossings
  - expected number, 21, 57
  - finite number, 44
- derivative
  - as linear filter, 103
  - at upcrossing, 56
- deterministic process, 151, 153
- differentiability
  - conditions for, 37
  - in quadratic mean, 28, 46
  - of Gaussian process, 39
  - of sample function, 37, 46
  - summary of conditions, 49
- differential equation
  - linear, 108
- directional spectrum, 171
- dispersion relation, 23, 96, 170
- Dobrushin, R.L., 36, 199
- Doob, J.L., 199
- Durbin's formula, 74
- Durbin, J., 74
- Durrett, R., 199
- eigenfunction, 122
- Einstein, A., ix, 20, 97, 199
- ensemble average, 52, 139
- envelope, 63, 117
  - of narrow band process, 118
  - simulation, 197
- equivalent processes, 29
- ergodicity, 65, 133
  - almost surely, 52
  - and mixing, 156
  - definition, 140
  - in continuous time, 147
  - in quadratic mean, 52
  - Markov chain, 139
  - of Gaussian process, 148
  - of random sequence, 145
  - upcrossings, 67
- excursion
  - above a level, 73
- exponential smoothing, 106
- extension
  - of probability measure, 5, 176
- fast Fourier transform, 193
- FFT, 193
- field, *see*  $\sigma$ -field
  - homogeneous field, 166
  - isotropic random field, 15, 167
  - random, 165
- filter, 101, 106
  - amplification, 103
  - causal, 104
  - cross-correlation, 165
  - linear, 100
  - phase shift, 103
  - time-invariant, 101

- finite-dimensional
  - approximation by event, 12
  - distribution, 8
- FM-noise, 21
- folding, 88
- Fourier
  - integral, 13
  - inversion, 86
  - transform, 21, 193
- frequency, 89
  - mean frequency, 63
- frequency function for filter, 101
- Fubini's theorem, 60
  
- Garsia, A.M., 142
- Gaussian process, 16, 96
  - continuity, 35
  - differentiability, 39
  - ergodicity, 148
  - Slepian model, 68
  - unboundedness, 36
- Gelfand, A.E., 199
- generating function, 108
- geostatistics, 166
- Gram-Schmidt orthogonalization, 190
- Grenander, U., ix, 24, 127, 200
- Gut, A., 200
  
- Hölder condition, 40–42
- Hajek, B., 201
- Hamming, R.W., 22
- harmonic oscillator, 106
- Herglotz' lemma, 88
- Hermitian function, 80
- Hilbert space, 187, 190
  - generated by a process, 191
- Hilbert transform, 115, 197
- Hilbert's sixth problem, 178
- homogeneous field
  - spectral distribution, 15, 166
  - spectral representation, 15, 166
- horizontal window conditioning, 66
  
- Ibragimov, I.A., 200
- impulse response, 101, 104
  
- inner product, 188, 190
- integrability
  - conditions for, 49
  - in quadratic mean, 49
- interval
  - in  $\mathbb{R}^\infty$ , 9
- invariant
  - measure, 137
  - random variable, 138
  - set, 138
- inversion of spectrum
  - continuous time, 86
  - discrete time, 94
- isometry, 90
  
- JONSWAP spectrum, 14, 74, 77, 172
- Jordan, D.W., 200
- jump discontinuity, 40
  
- Kac and Slepian, horizontal window, 66
- Kac, M., 57, 58
- Kac, M., 200
- Karhunen-Loève expansion, 122
- Kolmogorov
  - existence theorem, 10, 80, 178
  - extension theorem, 178
  - Grundbegriffe, 9, 178
  - probability axiom, 3, 175
- Kolmogorov, A.N., 3, 200
  
- Langevin's equation, 20, 98
- Lasota, A., 200
- law of large numbers, 192
- Leadbetter, M.R., 199, 200
- level crossing, 57
- likelihood ratio test, 128
- Lindgren, G., 200
- linear
  - filter, 100, 101
  - interpolation, 100
  - oscillator, 14, 106
  - prediction, 18, 100
  - process, 104
  - regression, 101

- subspace, 188
- time-invariant filter, 102
- Linnik, Yu.V., 200
- Lipschitz condition, 41
- Loève criterion, 45, 50, 184
- Lord Rayleigh, 22, 97
  
- m-dependence, 157
- Mackey, M.C., 200
- Markov's inequality, 34
- Maruyama, G., 200
- Matérn, B., 169
- Mercer's theorem, 127
- metrically transitive, 140
- mixing
  - for Gaussian processes, 157
  - relation to ergodicity, 158
  - strong, 157
  - uniform, 157
  - weak, 157
- modulo game, 135
  - as ergodic sequence, 140
- Monotone convergence, 61
- Monte Carlo simulation, 135, 193
- moving average
  - discrete time, 191
  - infinite, 105
- multivariate normal distribution, 15
  
- narrow banded spectrum, 14, 63
- non-deterministic process, 151
- non-negative function, 80
- non-random walk
  - as ergodic sequence, 144
- norm, 188
- normal distribution
  - bivariate, 56
  - conditional, 16
  - inequalities, 18
  - multivariate, 15
- nugget effect, 169
  
- observable, 9, 11, 124
- Ornstein-Uhlenbeck process, 39, 98, 112, 156
  
- orthogonal
  - basis, 189
  - increments, 89
- oscillator
  - harmonic, 106
  - linear, 106
  
- Palm distribution, 66
- Perrin, J.B., 20
- Petersen, K., viii, 200
- phase, 89, 94
- Pierson, W.J., ix, 22, 201
- Pierson-Moskowitz spectrum, 172
- point process, 66
- Poisson process, 40
- prediction, 100
  - after upcrossing, 68
  - linear, 18
  - quadratic mean, 7
- primordial
  - randomness, 151
  - soup, 151
- principal components, 121
- probability measure, 175
- projection
  - in Hilbert space, 100, 189
  - uncorrelated residuals, 101, 152
  - uniqueness, 101
- pseudo randomness, 134
- Pythagorean theorem, 188
  
- quadratic mean
  - continuity, 45
  - convergence, 45
    - Loève criterion, 45
  - differentiability, 46
  - integrability, 49
  - optimal prediction, 7
  
- random
  - amplitude, 89, 94, 194
  - moving surface, 23, 169
  - phase, 89, 94, 194
  - waves, 23, 170
- Rayleigh distribution, 117

- for slope at upcrossing, 70
- RC-filter, 106
- reconstruction, 18
- rectangle
  - $n$ -dimensional, 4, 25
  - generalized, 4, 9
- regression approximation, 74
- regular process, 150, 151
  - as moving average, 154
- Rice's formula, 21, 57
  - alternative proof, 60
  - for Gaussian process, 62
  - for non-stationary process, 61
- Rice, S., ix
- Rice, S.O., 21, 22, 57, 58, 79, 97, 200
- Riemann-Lebesgue's lemma, 149
- Rootzén, H., 200
- Royden, H.L., viii, 200
- Rozanov, Y.A., 200
- Rychlik, I., 201
  
- sample function average, 52
- sampling theorem, 119
- scalar product, 188
- Schwarz' inequality, 188
- second order stationary process, 79
- separable
  - process, 29
  - space, 189
- shift transformation, 147
- signal detection, 127
- simulation
  - by Fourier method, 193
  - of envelope, 197
- singular
  - process, 150, 151
  - spectral conditions, 154
  - spectrum, 153
- Slepian model, 57, 68
- Slepian, D., 57, 200, 201
- Smith, P., 200
- spectral
  - density, 13, 86
  - distribution, 13
    - for process, 82
    - for random field, 166
    - for sequence, 88
    - for vector process, 162
  - inversion, 86
  - moment, 13, 49
  - representation
    - of process, 21, 89
    - of random field, 166
    - of sequence, 100
    - of vector process, 163
  - width, 63, 77
- spectrum
  - absolutely continuous, 94, 153
  - amplitude spectrum, 165
  - coherence spectrum, 165
  - directional, 171
  - discrete, 93, 153
  - encountered, 23
  - JONSWAP, 14, 74, 77, 172
  - one-sided, 87, 95
  - phase spectrum, 165
  - Pierson-Moskowitz, 172
  - singular, 153
  - Whittle-Matérn, 169
- spreading function, 23, 96, 171
- stationary
  - intrinsically, 166
- stationary process, 13
  - as curve in Hilbert space, 191
  - complex, 82
  - second order, 13
  - strictly, 79
  - weakly, 13, 79
- StDenis, M., ix, 22, 201
- stochastic calculus, 113
- stochastic differential equation, 98, 108
- Stone-Weierstrass theorem, 92
- strong Markov property, 64
- subspace
  - generated, 191
  - spanned, 191

- tangents, non-existence of, 43  
time average, 52, 138  
time-invariant filter, 101  
total variation, 59  
transfer function, 101  
transformation  
    generated by stationary process, 137  
    generating stochastic process, 136  
    measurable, 136  
    measure preserving, 136  
    quadratic, 135  
    shift, 137  
triangle inequality, 188  
Tukey, J.W., 22
- upcrossings  
    counting, 60  
    expected number, 57  
    intensity, 58  
    slope at, 56
- variogram, 166  
vertical window conditioning, 66  
von Plato, J., 133, 200
- WAFO, 72, 74, 201  
wave  
    average length, 171  
    average period, 171  
    characteristics, 75  
    frequency, 23, 169, 170  
    number, 23, 169, 170  
    Pierson-Moskowitz spectrum, 119  
wave direction, 171  
wave frequency spectrum, 170  
wave number spectrum, 171  
Weyl's equidistribution theorem, 144  
white noise, 14, 94, 98, 112  
Whittle, P., 169  
Whittle-Matérn spectrum, 169  
Wiener process, 19, 112  
    as spectral process, 97  
    continuity, 34  
    derivative of, 20, 99, 112  
    Hölder condition, 42  
    infinite variation, 55  
    integral, 51  
    Karhunen-Loève expansion, 124  
Wiener, N., 22  
Williams, D., viii, 201  
Wold, H., 151  
Wong, E., 201  
Yaglom, A.M., 153, 201





October 2006  
Mathematical Statistics  
Centre for Mathematical Sciences  
Lund University  
Box 118, SE-221 00 Lund, Sweden  
<http://www.maths.lth.se/>