Statistics for Industry, Technology, and Engineering

Wojbor A. Woyczyński

A First Course in Statistics for Signal Analysis

Third Edition

Birkhäuser

Statistics for Industry, Technology, and Engineering

More information about this series at http://www.springer.com/series/15677

Wojbor A. Woyczyński

A First Course in Statistics for Signal Analysis

Third Edition

Birkhäuser

Wojbor A. Woyczyński
Cleveland, OH, USA

ISSN 2662-5555 ISSN 2662-5563 (electronic)
Statistics for Industry, Technology, and Engineering
ISBN 978-3-030-20907-0 ISBN 978-3-030-20908-7 (eBook)
https://doi.org/10.1007/978-3-030-20908-7

Mathematics Subject Classification: 62-01, 60G12, 60-01, 60G10, 60G15, 60G35, 62M10, 62M15, 62M20

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to my children:
Martin Wojbor, Gregory Holbrook, and
Lauren Pike.
They make it all worth it.

Foreword to the Third Edition

The third edition contains two additional chapters. Chapter 3 is devoted to wavelets and the uncertainty principle, and Chap. 11 discusses the forecasting problems for stationary time series. In Chap. 3, we discuss the transition from the windowed Fourier transform to the continuous wavelet transforms and then move on to Haar wavelets and mutiresolution analysis. In Chap. 11, the forecasting (prediction) problems are discussed in the context of the Wold decomposition theorem, and a solution to the optimal predictor problem is found in the context of the spectral representation of stationary time series.

Both topics are essential for a deeper understanding of statistical analysis of random signals and make the book more complete. Some misprints in the previous edition have also been corrected.

https://sites.google.com/a/case.edu/waw
Cleveland, OH, USA
February 2019

Wojbor A. Woyczyński

Foreword to the Second Edition

The basic structure of the Second Edition remains the same but many changes have been introduced responding to several years' worth of comments of students and other users of the First Edition. Most of the figures have been redrawn to better show the scale of the quantities represented in them, some notation and terminology has been adjusted to better reflect the concepts under discussion, and several sections have been considerably expanded by addition of new examples, illustrations, and commentary. Thus the original conciseness has been somewhat softened. A typical example here would be an addition of Remark 4.1.2 which explains how one can see the Bernoulli white noise in continuous time as a scaling limit of switching signals with exponential inter-switching times. There are also new, more applied exercises as well, such as Problem 9.7.9 on simulating signals produced by spectra generated by incandescent and luminescent lamps.

Still the book remains more mathematical than many other signal processing books. So, at Case Western Reserve University the course (required, for most of the Electrical Engineering, and some Biomedical Engineering juniors/seniors) based on this book runs in parallel with a signal processing course that is entirely devoted to practical applications and software implementation. This one-two punch approach has been working well, and the engineers seem to appreciate the fact that all probability/statistics/Fourier analysis foundations are developed within the book; adding extra mathematical courses to a tight undergraduate engineering curriculum is almost impossible. A gaggle of graduate students in applied mathematics, statistics and assorted engineering areas also regularly enrolls. They are often asked to make in-class presentations of special topics included in the book but not required of the general undergraduate audience.

Finally, by popular demand, there is now a large appendix which contains solutions of selected problems from each of the nine chapters. Here, most of the credit goes to my former graduate students who served as TAs for my courses: Aleksandra Piryatinska (now, at San Francisco State University), Sreenivas Konda (now, at Temple University), Dexter Cahoy (now, at Louisiana Tech), and Peipei Shi (now, at Eli Lilly, Inc.). In preparing the second edition the author took into account

useful comments that appeared in several reviews of the original book; the review published in September 2009 in the Journal of American Statistical Association by Charles Boncelet was particularly thorough and insightful.

http://stat.case.edu/~Wojbor Wojbor A. Woyczyński
Cleveland, OH, USA
May 2010

Introduction

This book was designed as a text for a first, one-semester course in statistical signal analysis for students in engineering and physical sciences. It has been developed over the last few years as lecture notes used by the author in classes mainly populated by electrical, systems, computer, and biomedical engineering juniors/seniors and graduate students in sciences and engineering who have not been previously exposed to this material. It was also used for industrial audiences as educational and training materials and for an introductory time series analysis class.

The only prerequisite for this course is a basic two- to three-semester calculus sequence; no probability or statistics background is assumed except the usual high school elementary introduction. The emphasis is on a crisp and concise, but fairly rigorous presentation of fundamental concepts in the statistical theory of stationary random signals and relationships between them. The author's goal was to write a compact but readable book of less than 200 pages countering the recent trend toward fatter and fatter textbooks. Since Fourier series and transforms are of fundamental importance in random signal analysis and processing, this material is developed from scratch in Chap. 2, emphasizing the time-domain vs. frequency-domain duality. Our experience showed that although harmonic analysis is normally included in the calculus syllabi, students' practical understanding of its concepts is often hazy. Chapter 3 introduces the alternative harmonic analysis concepts based on the wavelet transforms, while Chap. 4 explains the basic concepts of probability theory, the law of large numbers, the stability of fluctuations law, and the statistical parametric inference procedures based on the latter.

In Chap. 5, the fundamental concept of a stationary random signal and its autocorrelation structure is introduced. This time-domain analysis is then expanded to frequency domain by discussion in Chap. 6 of power spectra of stationary signals. How stationary signals are affected by their transmission through linear systems is the subject of Chap. 7. This transmission analysis permits a preliminary study of the issues of designing filters with the optimal signal-to-noise ratio; this is done in Chap. 8. Chapter 9 concentrates on Gaussian signals where the autocorrelation structure completely determines all the statistical properties of the signal. The text concludes, in Chap. 10, with description of algorithms for computer simulations

of stationary random signals with given power spectrum density. The routines are based on the general spectral representation theorem for such signals which is also derived in this chapter. Finally, Chap. 11 discusses the prediction problem for stationary time series.

The book is essentially self-contained, assuming the indispensable calculus background mentioned above. A complementary bibliography, for readers who would like to pursue the study of random signals in greater depth, is described at the end of this volume.

Some general advice to students using this book: The material is deliberately written in a compact, economical style. To achieve the understanding needed for independent solving of the problems listed at the end of each chapter in the Problems and Exercises sections, it is not sufficient to read through the text in the manner you would read through a newspaper or a novel. It is necessary to look at every single statement with a "magnifying glass" and to decode it in your own technical language so that you can use it *operationally* and not just be able to talk about it. The only practical way to accomplish this goal is to go through each section with pencil and paper, explicitly completing, if necessary, routine analytic intermediate steps that were omitted in the exposition for the sake of the clarity of the presentation of the bigger picture. It is the latter that the author wants you to keep at the end of the day; there is no danger in forgetting all the little details if you know that you can recover them by yourself when you need them.

Finally, the author would like to thank Professors Mike Branicky and Ken Loparo of the Department of Electrical and Computer Engineering and Professor Robert Edwards of the Department of Chemical Engineering of Case Western Reserve University for their kind interest and help in the development of this course and comments on the original version of this book. My graduate students Alexey Usoltsev and Alexandra Piryatinska also contributed to the editing process, and I appreciate the time they spent on this task. Partial support for this writing project from the Columbus Instruments International Corporation of Columbus, Ohio, Dr. Jan Czekajewski, President, is here also gratefully acknowledged.

Four anonymous referees spent considerable time and effort trying to improve the original manuscript. Their comments are appreciated, and, almost without exception, their sage advice was incorporated in the final version of the book. I thank them for their help.

Contents

Part I

1 Description of Signals 3
1.1 Types of Random Signals 3
1.2 Characteristics of Signals 11
1.3 Time Domain and Frequency Domain Descriptions of Periodic Signals 12
1.4 Building a Better Mousetrap: Complex Exponentials 19
1.5 Problems and Exercises 22

2 Spectral Representation of Deterministic Signals: Fourier Series and Transforms 25
2.1 Complex Fourier Series for Periodic Signals 25
2.2 Approximation of Periodic Signals by Finite Fourier Sums 35
2.3 Aperiodic Signals and Fourier Transforms 41
2.4 Basic Properties of Fourier Transform 44
2.5 Fourier Transforms of Some Non-integrable Signals: Dirac's Delta-Impulse 47
2.6 Discrete and Fast Fourier Transforms 51
2.7 Problems and Exercises 53

3 Uncertainty Principle and Wavelet Transforms 57
3.1 Time–Frequency Localization and the Uncertainty Principle 57
3.2 Windowed Fourier Transform 60
3.3 Continuous Wavelet Transforms 65
3.4 Haar Wavelets and Multiresolution Analysis 78
3.5 Continuous Daubechies' Wavelets 84
3.6 Exercises 89

4 Random Quantities and Random Vectors 91
4.1 Discrete, Continuous, and Singular Random Quantities 91
4.2 Expectations and Moments of Random Quantities 113

4.3 Random Vectors, Conditional Probabilities, Statistical Independence, and Correlations 117
4.4 The Least Squares Fit, Linear Regression 129
4.5 The Law of Large Numbers and the Stability of Fluctuations Law 133
4.6 Estimators of Parameters and Their Accuracy: Confidence Intervals 136
4.7 Problems and Exercises 144

Part II

5 Stationary Signals 151
5.1 Stationarity and Autocovariance Functions 151
5.2 Estimating the Mean and the Autocovariance Function, Ergodic Signals 166
5.3 Problems and Exercises 170

6 Power Spectra of Stationary Signals 175
6.1 Mean Power of a Stationary Signal 175
6.2 Power Spectrum and Autocovariance Function 177
6.3 Power Spectra of Interpolated Digital Signals 184
6.4 Problems and Exercises 189

7 Transmission of Stationary Signals Through Linear Systems 193
7.1 Time Domain Analysis 193
7.2 Frequency Domain Analysis and System's Bandwidth 202
7.3 Digital Signal, Discrete Time Sampling 206
7.4 Problems and Exercises 212

Part III

8 Optimization of Signal-to-Noise Ratio in Linear Systems 217
8.1 Parametric Optimization for a Fixed Filter Structure 217
8.2 Filter Structure Matched to Input Signal 221
8.3 The Wiener Filter 224
8.4 Problems and Exercises 227

9 Gaussian Signals, Covariance Matrices, and Sample Path Properties 229
9.1 Linear Transformations of Random Vectors 229
9.2 Gaussian Random Vectors 232
9.3 Gaussian Stationary Signals 237
9.4 Sample Path Properties of General and Gaussian Stationary Signals 239
9.5 Problems and Exercises 245

10 Spectral Representation of Discrete-Time Stationary Signals and Their Computer Simulations 247
10.1 Spectral Representation 247
10.2 Autocovariance as a Positive-Definite Sequence 249
10.3 Cumulative Power Spectrum of Discrete-Time Stationary Signal 251
10.4 Stochastic Integration with Respect to Signals with Uncorrelated Increments 254
10.5 Spectral Representation of Stationary Signals 259
10.6 Computer Algorithms: Complex-Valued Case 263
10.7 Computer Algorithms: Real-Valued Case 269
10.8 Problems and Exercises 275

11 Prediction Theory for Stationary Random Signals 279
11.1 The Wold Decomposition Theorem and Optimal Predictors 279
11.2 Application of the Spectral Representation to the Solution of the Prediction Problem 282
11.3 Examples of Linear Prediction for Stationary Time Series 287
11.4 Problems and Exercises 289

Solutions to Selected Problems and Exercises 291

Bibliographical Comments 323

Index 327

Notation

To be used only as a guide and not as a set of formal definitions

$\mathbf{AV}_x$	Time average of signal $x(t)$
BW_n	Equivalent-noise bandwidth of the system
$BW_{1/2}$	Half-power bandwidth of the system
$\mathbf{C}$	The set of all complex numbers
$\mathrm{Cov}(X, Y)$	$\mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)]$, covariance of X and Y
δ_{mn}	Kronecker's delta, = 0 if $m \neq n$, and =1 if $m = n$
$\delta(x)$	Dirac delta "function"
$\mathbf{EN}_x$	Energy of signal $x(t)$
$\mathbf{E}(X)$	Expected value (mean) of random quantity X
$F_X(x)$	Cumulative distribution function (c.d.f.) of random quantity X
$f_X(x)$	Probability density function (p.d.f.) of random quantity X
$\gamma_X(\tau)$	$\mathbf{E}(X(t)-\mu_X)(X(t+\tau)-\mu_X)$ autocovariance function of a stationary signal $X(t)$
$h(t)$	Impulse response function of a linear system
$H(f)$	Transfer function of a linear system, Fourier transform of $h(t)$
$\lvert H(f)\rvert^2$	Power transfer function of a linear system
$L_0^2(\mathbf{P})$	Space of all zero-mean random quantities with finite variance
$m^\alpha(X)$	$\mathbf{E}\lvert X\rvert^\alpha$—$\alpha$-th absolute moment of random quantity X
$\mu^k(X)$	$\mathbf{E}(X^k)$—k-th moment of random quantity X
$N(\mu, \sigma^2)$	Gaussian (normal) probability distribution with mean μ and variance σ^2
P	Period of a periodic signal
$\mathbf{P}(A)$	Probability of event A
$\mathbf{PW}_x$	Power of signal $x(t)$
$Q_X(\alpha)$	$F_X^{-1}(\alpha)$, α's quantile of random quantity X
R	Resolution
$\mathbf{R}$	The set of all real numbers
$\rho_{X,Y}$	$\mathrm{Cov}(X, Y)/(\sigma_X\sigma_Y)$, correlation coefficient of X and Y
$\mathrm{Std}\,(X)$	$\sigma_X = \sqrt{\mathrm{Var}(X)}$—the standard deviation of random quantity X

$S_X(f)$	Power spectral density of stationary signal $X(t)$
$\mathcal{S}_X(f)$	Cumulative power spectrum of stationary signal $X(t)$]
T	Sampling period
$u(t)$	Heaviside unit step function, $u(t) = 0$, for $t < 0$, and $= 1$, for $t \geq 0$
Var (X)	$\mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2$—the variance of random quantity X
$W(n)$	Discrete-time white noise
$\mathcal{W}(n)$	Cumulative discrete-time white noise
$W(t)$	Continuous-time white noise
$\mathcal{W}(t)$	The Wiener process
$x(t)$, $y(t)$, $etc.$	Deterministic signals
$\boldsymbol{X}$	$(X_1, X_2, \ldots, X_d)$—a random vector in dimension d
$x(t) * y(t)$	Convolution of signals $x(t)$ and $y(t)$
$X(f)$, $Y(f)$	Fourier transforms of signals $x(t)$, and $y(t)$, respectively
X, Y, Z	Random quantities (random variables)
z^*	Complex conjugate of complex number z, i.e., if $z = \alpha + j\beta$ then $z^* = \alpha - j\beta$
$\lfloor a \rfloor$	"Floor" function, the largest integer not exceeding number a
$\langle \, . \, , \, . \, \rangle$	Inner (dot, scalar) product of vectors or signals
$\Leftrightarrow$	If, and only if
:=	Is defined as

Part I

Chapter 1
Description of Signals

Abstract Signals are everywhere. Literally. The universe is bathed in the background radiation, the remnant of the original Big Bang and, as your eyes scan this page, a signal is being transmitted to your brain where different sets of neurons analyze it and process it. All human activities are based on processing and analysis of sensory signals but the goal of this book is somewhat narrower. The signals we will be mainly interested in can be described as *data* resulting from quantitative measurements of some physical phenomena and our emphasis will be on data that display *randomness* that may be due to different causes, be it errors of measurements, the algorithmic complexity, or the chaotic behavior of the underlying physical system itself.

1.1 Types of Random Signals

For the purpose of this book, signals will be functions of real variable t interpreted as time. To describe and analyze signals we will adopt the functional notation: $x(t)$ will denote the value of a nonrandom signal at time t. The values themselves can be real or complex numbers, in which case we will symbolically write $x(t) \in \mathbf{R}$, or, respectively, $x(t) \in \mathbf{C}$. In certain situations it is necessary to consider vector-valued signals with $x(t) \in \mathbf{R}^d$, where d stands for the dimension of the vector $x(t)$ with d real components.

Signals can be classified into different categories depending on their features. For example:

- *Analog signals* are functions of continuous time and their values form a continuum. *Digital signals* are functions of discrete time dictated by the computer's clock and their values are also discrete and dictated by the resolution of the

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_1

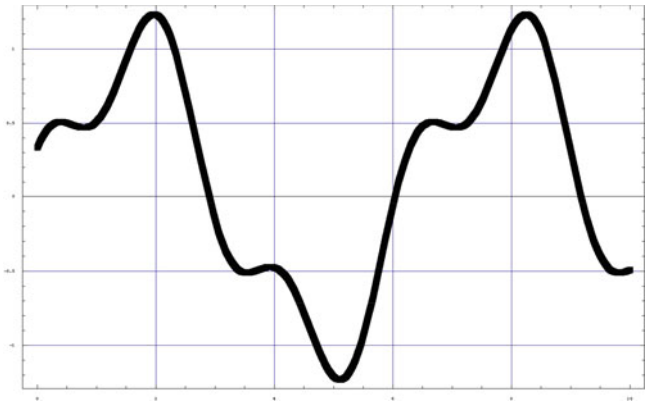


Fig. 1.1 Signal $x(t) = \sin(t) + \frac{1}{3}\cos(3t)$ [V] is analog and periodic with period $P = 2\pi$ [s]. It is also deterministic

system. Of course, one can also encounter mixed type signals which are sampled at discrete times but whose values are not restricted to any discrete set of numbers.

- *Periodic signals* are functions whose values are periodically repeated. In other words, for a certain number $P > 0$, we have $x(t + P) = x(t)$, for any t. Number P is called the *period of the signal*. *Aperiodic signals* are signals that are not periodic.
- *Deterministic signals* are signals not affected by random noise; there is no uncertainty about their values. *Random signals*, often also called *stochastic processes* , include an element of uncertainty; their analysis requires use of statistical tools and providing such tools is the principal goal of this book.

For example, signal $x(t) = \sin(t) + \frac{1}{3}\cos(3t)$ [V] shown in Fig. 1.1 is deterministic, analog, and periodic with period $P = 2\pi$ [s]. The same signal, digitally sampled during the first 5 s at time intervals equal to 0.5 s, with resolution 0.01 V, gives tabulated values:

t	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
$x(t)$	0.50	0.51	0.93	1.23	0.71	−0.16	0.51	−0.48	−0.78	−1.21

This sampling process is called the *analog-to-digital conversion*: given the *sampling period* T and the *resolution* R, the digitized signal $x_d(t)$ is of the form

$$x_d(t) = R\left\lfloor \frac{x(t)}{R} \right\rfloor, \quad \text{for} \quad t = T, 2T, \ldots, \tag{1.1.1}$$

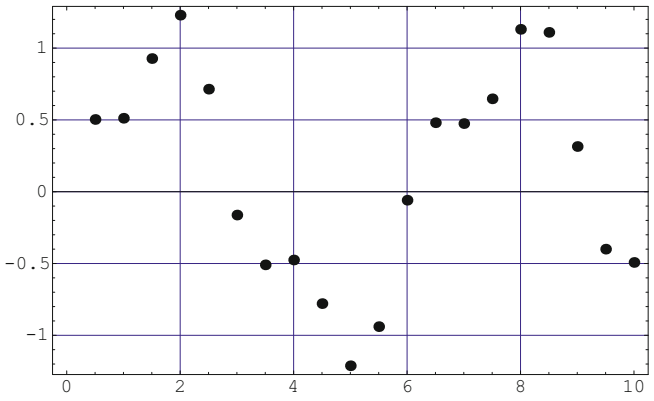


Fig. 1.2 Signal $x(t) = \sin(t) + \frac{1}{3}\cos(3t)$ [V] digitally sampled at time intervals equal to 0.5 s with resolution 0.01 V

where the, convenient to introduce here, "floor" function $\lfloor a \rfloor$ is defined as the largest integer not exceeding real number a. For example, $\lfloor 5.7 \rfloor = 5$, but $\lfloor 5.0 \rfloor = 5$, as well.

Note the role the resolution R plays in the above formula. Take, for example, $R = 0.01$. If the signal $x(t)$ takes all the continuous values between $m = \min_t x(t)$ and $M = \max_t x(t)$, then $x(t)/0.01$ takes all the continuous values between $100 \cdot m$ and $100 \cdot M$, but $\lfloor x(t)/0.01 \rfloor$ takes only integer values between $100 \cdot m$ and $100 \cdot M$. Finally, $0.01\lfloor x(t)/0.01 \rfloor$ takes as its values only all the discrete numbers between m and M that are 0.01 apart (Fig. 1.2).

Randomness of signals can have different origin, be it quantum *uncertainty principle*, *computational complexity* of algorithms, *chaotic behavior* in dynamical systems, or random fluctuations and errors in measurement of outcomes of independently repeated experiments.[1] The usual way to study them is via their aggregated statistical properties. The main purpose of this book is to introduce some of the basic mathematical and statistical tools useful in analysis of random signals that are produced under *stationary conditions*, that is, in situations where the measured signal may be stochastic and contain random fluctuations, but the basic underlying random mechanism producing it does not change over time; think here about outcomes of independently repeated experiments, each consisting of tossing a single coin (Fig. 1.3).

At this point, to help the reader visualize the great variety of random signals appearing in the physical sciences and engineering, it is worthwhile to review a gallery of pictures of random signals, both experimental and simulated, presented in Figs. 1.4, 1.5, 1.6, 1.7, and 1.8. The captions explain the context in each case.

[1] See, e.g., M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity, and Chaotic Behavior in Engineering and Science,* Birkhäuser-Boston, 1998.

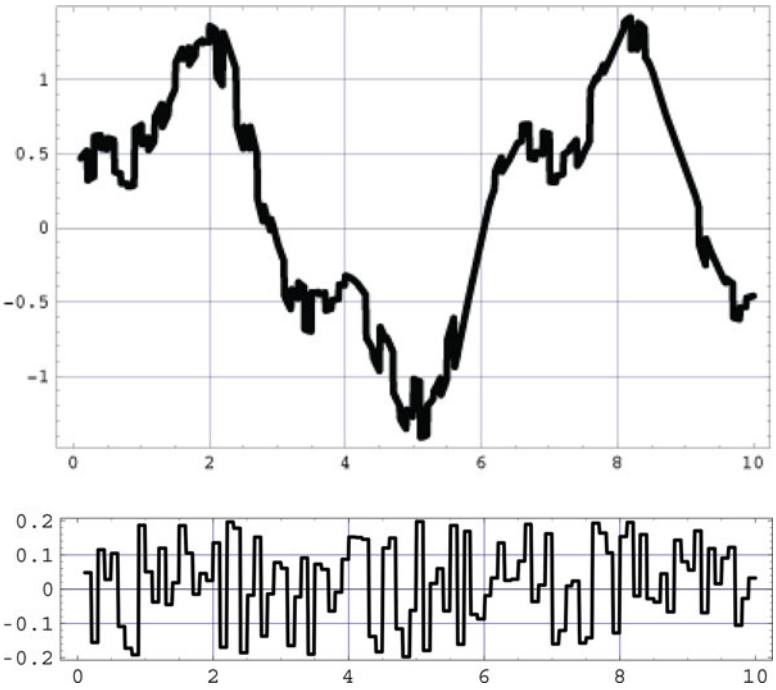


Fig. 1.3 Signal $x(t) = \sin(t) + \frac{1}{3}\cos(3t)$ [V] in presence of additive random noise with average amplitude of 0.2 V. The magnified noise component itself is pictured underneath the graph of the signal

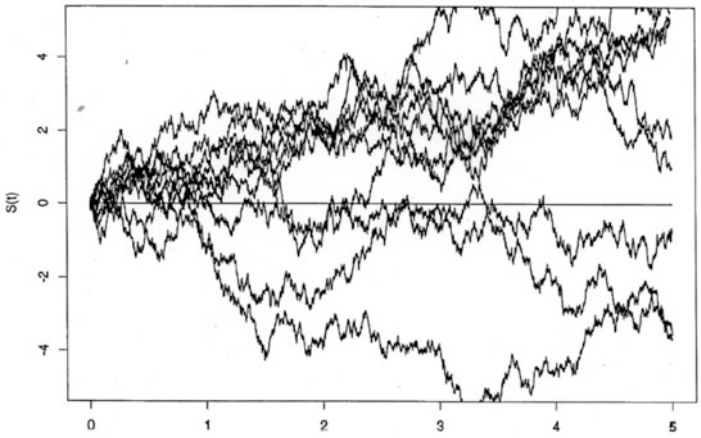


Fig. 1.4 Several, computer-generated trajectories (sample paths) of a random signal called the *Brownian motion* stochastic process or the *Wiener stochastic process*. Its trajectories, although very rough, are continuous. It is often used as a simple model of *diffusion*. The random mechanism that created different trajectories was the same. Its importance for our subject matter will become clear in Chap. 9

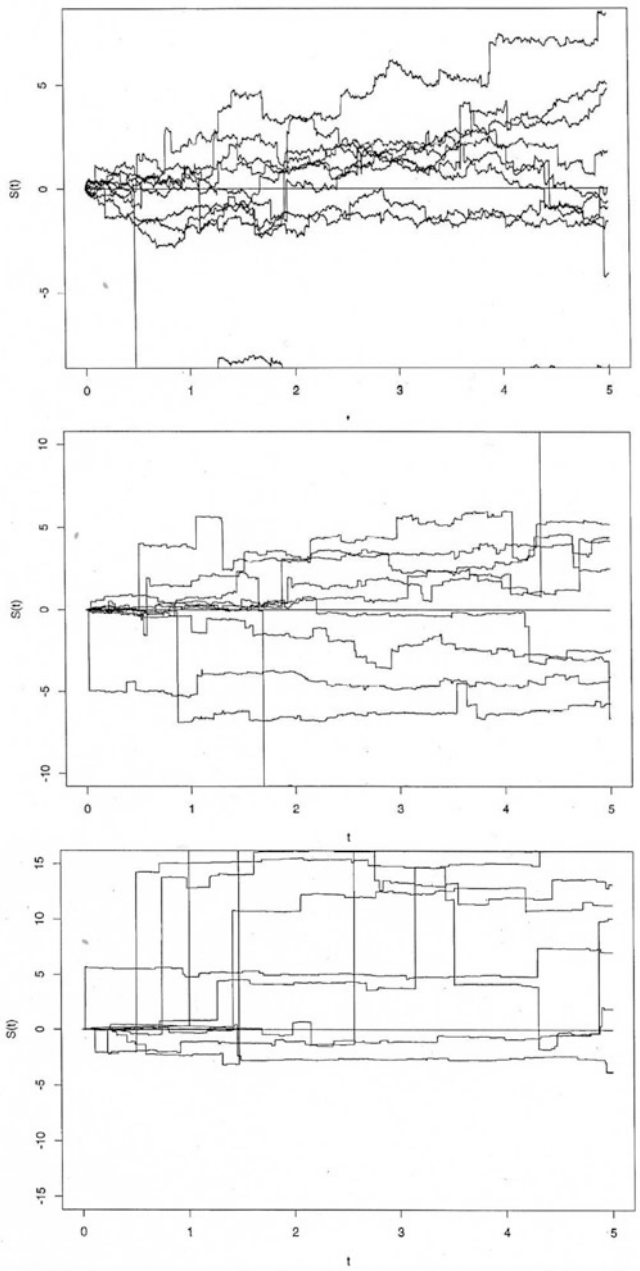


Fig. 1.5 Several, computer-generated trajectories (sample paths) of random signals called *Lévy stochastic processes* with parameter $\alpha = 1.5, 1$, and 0.75, respectively (from top to bottom). They are often used to model anomalous diffusion processes wherein diffusing particles are also permitted to change their position by jumping. Parameter α indicates intensity of jumps of different sizes. Parameter value $\alpha = 2$ corresponds to the Wiener process (shown in Fig. 1.4) which has trajectories that have no jumps. In each figure, the random mechanism that created different trajectories was the same. However, different random mechanisms led to trajectories presented in different figures

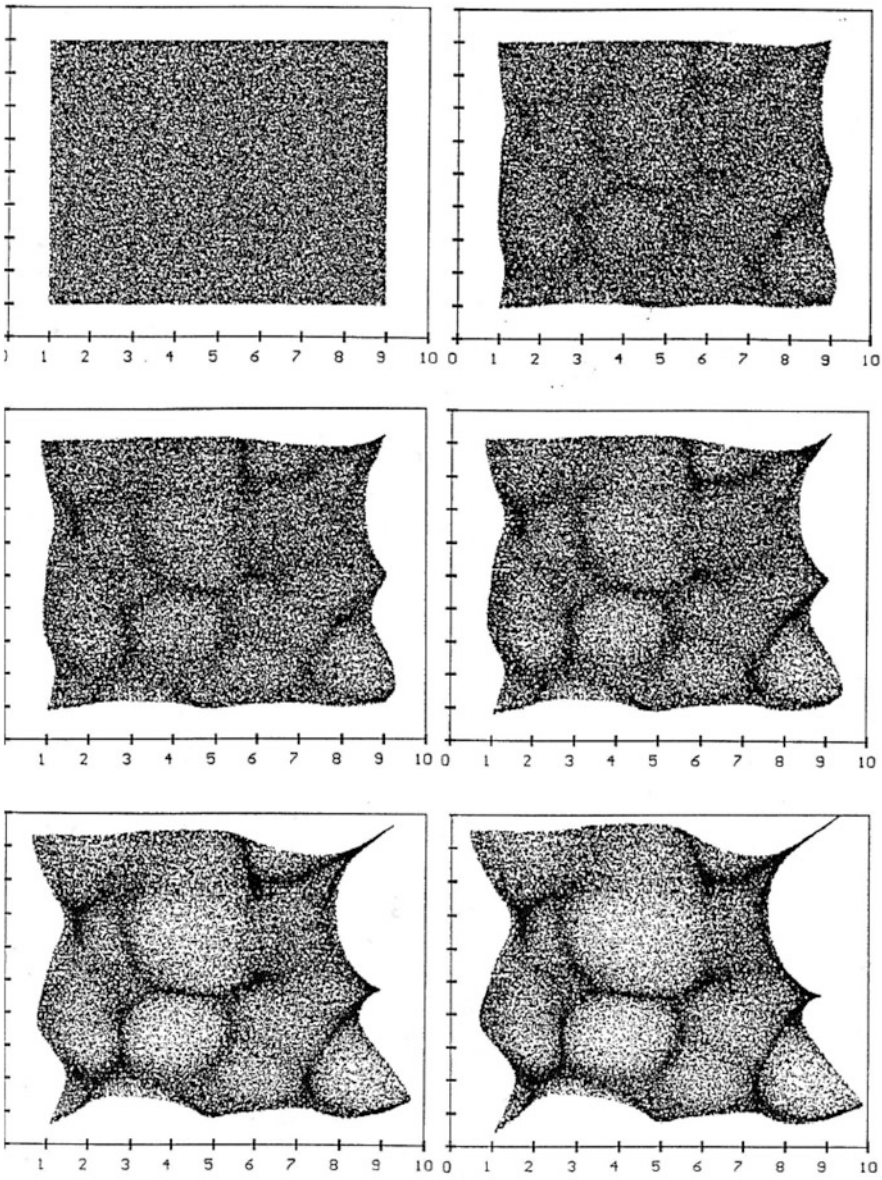


Fig. 1.6 Computer simulation of the evolution of passive tracer density in a turbulent velocity field with random initial distribution and random "shot-noise" initial velocity data. The simulation was performed for 100,000 particles. The consecutive frames show the location of passive tracer particles at times $t = 0.0, 0.3, 0.6, 1.0, 2.0, 3.0$ s

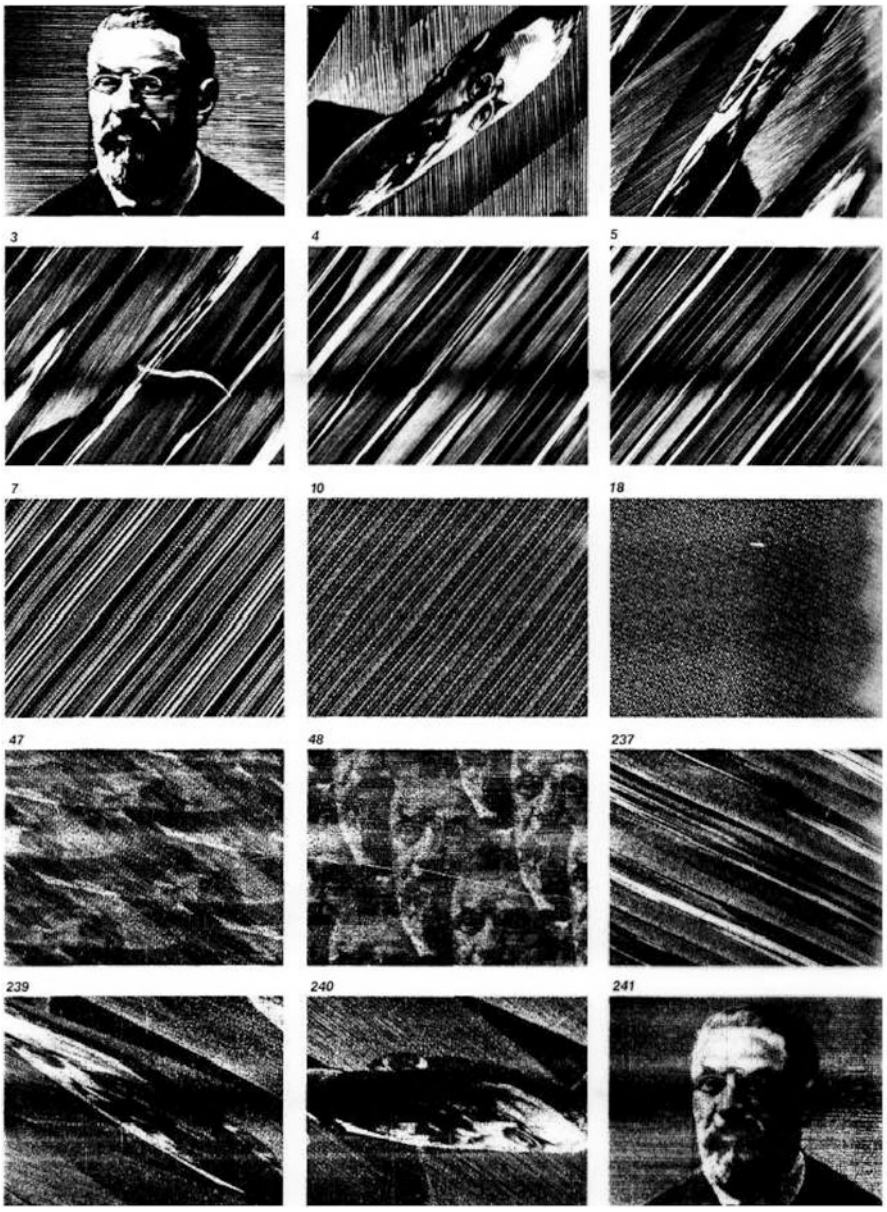


Fig. 1.7 Some deterministic signals (in this case, the images) transformed by deterministic systems can appear random. The above picture shows a series of iterated transformations of the original image via a fixed linear 2D mapping (matrix). The number of iterations applied is indicated in the top left corner of each image. The curious behavior of iterations, the original image first dissolving into seeming randomness only to return later to an almost original condition, is related to the so-called *ergodic* behavior. Thus irreverently transformed is Professor Henri Poincaré (1854–1912) of the University of Paris, the pioneer of ergodic theory of stationary phenomena (From *Scientific American*, reproduced with permission. Copyright 1986, James P. Crutchfield)

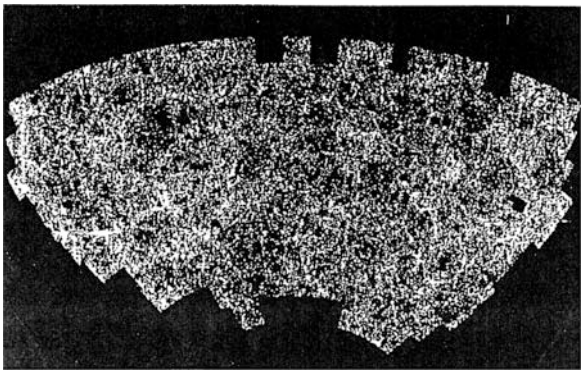


Fig. 1.8 A signal (again, an image) representing the large-scale and apparently random distribution of mass in the universe. The data come from the APM galaxy survey and shows more than two million galaxies in a section of sky centered on the South Galactic pole. The so-called *adhesion model* of the large scale mass distribution in the Universe uses Burgers equation to model the relevant velocity fields

The signals shown in Figs. 1.4 and 1.5 are, obviously, not stationary and have a diffusive character. However, their increments (differentials) are stationary and, in Chap. 9, they will play an important role in construction of the spectral representation of stationary signals themselves. The signal shown in Fig. 1.4 can be interpreted as a *trajectory*, or *sample path*, of a *random walker* moving, in discrete time steps, up or down a certain distance with equal probabilities 1/2 and 1/2. However, in the picture these trajectories are viewed from far away, and in accelerated time, so that both time and space appear continuous.

In certain situations the randomness of the signal is due to uncertainty about initial conditions of the underlying phenomenon which otherwise can be described by perfectly deterministic models such as partial differential equations. A sequence of pictures in Fig. 1.6 shows evolution of the system of particles with an initially random (and homogeneous in space) spatial distribution. The particles are then driven by the velocity field $\vec{v}(t,\vec{x}) \in \mathbf{R}^2$ governed by the so-called *2D Burgers equation*[2]

$$\frac{\partial \vec{v}(t,\vec{x})}{\partial t} + \Big(\nabla \cdot \vec{v}(t,\vec{x})\Big)\vec{v}(t,\vec{x}) = D\left(\frac{\partial^2 \vec{v}(t,\vec{x})}{\partial x_1^2} + \frac{\partial^2 \vec{v}(t,\vec{x})}{\partial x_2^2}\right), \tag{1.1.2}$$

where $\vec{x} = (x_1, x_2)$, the *nabla* operator $\nabla = \partial/\partial x_1 + \partial/\partial x_2$, and the positive constant D is the coefficient of diffusivity. The initial velocity field is also assumed to be random.

[2] See, e.g., W.A. Woyczyński, *Burgers-KPZ Turbulence–Göttingen Lectures*, Springer-Verlag 1998.

1.2 Characteristics of Signals

Several physical characteristics of signals are of primary interest.

- *The time average of the signal:* For analog, continuous-time signals the time average is defined by the formula

$$\mathbf{AV}_x = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)\,dt, \tag{1.2.1}$$

and for digital, discrete-time signals which are defined only for the time instants $t = nT$, $n = 0, 1, 2, \ldots, N-1$, it is defined by the formula

$$\mathbf{AV}_x = \frac{1}{N} \sum_{n=0}^{N-1} x(nT). \tag{1.2.2}$$

For periodic signals, it follows from (1.2.1) that

$$\mathbf{AV}_x = \frac{1}{P} \int_0^P x(t)\,dt, \tag{1.2.3}$$

so that, for the signal $x(t) = \sin t + (1/3)\cos(3t)$ pictured in Fig. 1.1, the time average is 0 as both $\sin t$ and $\cos(3t)$ integrate out to zero over the period $P = 2\pi$.

- *Energy of the signal:* For an analog signal $x(t)$, the total energy

$$\mathbf{EN}_x = \int_0^\infty |x(t)|^2\,dt, \tag{1.2.4}$$

and for digital signals

$$\mathbf{EN}_x = \sum_{n=0}^{\infty} |x(nT)|^2 \cdot T. \tag{1.2.5}$$

Observe that the energy of a periodic signal, such as the one from Fig. 1.1, is necessarily infinite if considered over the whole positive time line. Also note that, since in what follows it will be convenient to consider complex-valued signals, the above formulas include notation for the square of the modulus of a complex number: $|z|^2 = (\mathrm{Re}\, z)^2 + (\mathrm{Im}\, z)^2 = z \cdot z^*$; more about it in the next section.

- *Power of the signal:* Again, for an analog signal, the (average) power

$$\mathbf{PW}_x = \lim_{T\to\infty} \frac{1}{T} \int_0^T |x(t)|^2\,dt \tag{1.2.6}$$

and for a digital signal

$$\mathbf{PW}_x = \lim_{N\to\infty} \frac{1}{NT} \sum_{n=0}^{N-1} |x(nT)|^2 \cdot T = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |x(nT)|^2. \tag{1.2.7}$$

As a consequence, for an analog periodic signal with period P,

$$\mathbf{PW}_x = \frac{1}{P} \int_0^P |x(t)|^2 \, dt. \tag{1.2.8}$$

For example, for the signal in Fig. 1.1,

$$\mathbf{PW}_x = \frac{1}{2\pi} \int_0^{2\pi} \Big(\sin t + (1/3)\cos(3t) \Big)^2 \, dt \tag{1.2.9}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left(\sin^2 t + \frac{2}{3} \sin t \cos(3t) + \frac{1}{9} \cos^2(3t) \right) dt$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{1}{2}(1 - \cos(2t)) + \frac{2}{3}\frac{1}{2}(\sin(4t) - \sin(2t)) + \frac{1}{9}\frac{1}{2}(1 + \cos(6t)) \right) dt$$

$$= \frac{1}{2\pi} \left(\frac{1}{2} 2\pi + \frac{1}{9}\frac{1}{2} 2\pi \right) = \frac{5}{9}.$$

The above routine calculation, deliberately carried out here in detail, was somewhat tedious because of the need for various trigonometric identities. To simplify such manipulations and make the whole theory more elegant, we will introduce in the next section a complex number representation of the trigonometric functions via the so-called de Moivre formulas.

Remark 1.2.1 (Timeline Infinite in Both Direction) Sometimes it is convenient to consider signals defined for all time instants t, $-\infty < t < +\infty$, rather than just for positive t. In such cases all of the above definitions have to be adjusted in obvious ways, replacing the one-sided integrals and sums by two-sides integrals and sums, and adjusting the averaging constants correspondingly.

1.3 Time Domain and Frequency Domain Descriptions of Periodic Signals

The Time Domain Description The trigonometric functions

$$x(t) = \cos(2\pi f_0 t), \qquad \text{and} \qquad y(t) = \sin(2\pi f_0 t),$$

represent a harmonically oscillating signal with period $P = 1/f_0$ (measured, say, in seconds [s]), and the frequency f_0 (measured, say, in cycles per second, or Hertz [Hz]), and so do the trigonometric functions

$$x(t) = \cos(2\pi f_0(t+\theta)), \qquad \text{and} \qquad y(t) = \sin(2\pi f_0(t+\theta))$$

shifted by the phase-shift θ. The powers

$$\mathbf{PW}_x = \frac{1}{P}\int_0^P \cos^2(2\pi f_0 t)\,dt = \frac{1}{P}\int_0^P \frac{1}{2}(1+\cos(4\pi f_0 t))\,dt = \frac{1}{2}, \tag{1.3.1}$$

$$\mathbf{PW}_y = \frac{1}{P}\int_0^P \sin^2(2\pi f_0 t)\,dt = \frac{1}{P}\int_0^P \frac{1}{2}(1-\cos(4\pi f_0 t))\,dt = \frac{1}{2}, \tag{1.3.2}$$

using the trigonometric formulas from Tables 1.1 and 1.2. The phase shifts, obviously do not change the power of the above harmonic signals.

Table 1.1 Trigonometric formulas

$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \sin\beta\cos\alpha;$
$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta;$
$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2};$
$\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2};$
$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2};$
$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2};$
$\sin^2\alpha - \sin^2\beta = \cos^2\beta - \cos^2\alpha = \sin(\alpha+\beta)\sin(\alpha-\beta);$
$\cos^2\alpha - \sin^2\beta = \cos^2\beta - \sin^2\alpha = \cos(\alpha+\beta)\cos(\alpha-\beta);$
$\sin\alpha\cos\beta = \frac{1}{2}\Big[\sin(\alpha+\beta) + \sin(\alpha-\beta)\Big];$
$\cos\alpha\cos\beta = \frac{1}{2}\Big[\cos(\alpha+\beta) + \cos(\alpha-\beta)\Big];$
$\sin\alpha\sin\beta = \frac{1}{2}\Big[\cos(\alpha-\beta) - \cos(\alpha+\beta)\Big];$

Table 1.2 Complex numbers and De Moivre formulas

(i) By definition,

$$j = \sqrt{-1}.$$

(ii) Hence, for any integer m,

$$j^{4m} = 1, \quad j^{4m+1} = j, \quad j^{4m+2} = -1, \quad j^{4m+3} = -j.$$

(iii) Cartesian representation of the complex number:

$$z = a + jb, \quad a = \mathrm{Re}\, z, \quad b = \mathrm{Im}\, z,$$

where both a and b are real numbers and are called, respectively, the real and imaginary components of z. The complex number,

$$z^* = a - jb,$$

is called the complex conjugate of z.

(iv) The polar representation of the complex number (it is a good idea to think about complex numbers as representing points, or vectors, in the two-dimensional plane spanned by the two basic unit vectors, 1 and j):

$$z = |z|(\cos\theta + j\sin\theta) = |z| \cdot e^{j\theta},$$

and

$$z^* = |z|(\cos\theta - j\sin\theta) = |z| \cdot e^{-j\theta},$$

where

$$|z| = \sqrt{a^2 + b^2} = \sqrt{z \cdot z^*}, \quad \text{and} \quad \theta = \mathrm{Arg}\, z = \arctan \frac{\mathrm{Im} z}{\mathrm{Re} z},$$

is called, respectively, the *modulus* of z, and the *argument* of z. Alternatively,

$$\mathrm{Re}\, z = \frac{z + z*}{2} = |z|\cos\theta, \quad \mathrm{Im}\, z = \frac{z - z*}{2j} = |z|\sin\theta.$$

(v) For any complex number $w = \beta + j\alpha$,

$$e^w = e^{\beta + j\alpha} = e^{\beta}(\cos\alpha + j\sin\alpha).$$

(vi) For any complex number $z = a + jb = |z|e^{j\theta}$, and any integer n,

$$z^n = |z|^n e^{jn\theta} = (a^2 + b^2)^{n/2}(\cos n\theta + j\sin n\theta).$$

Taking their linear combination (like the one in Fig. 1.1), with amplitudes A and B, respectively,

$$z(t) = Ax(t) + By(t) = A\cos(2\pi f_0(t+\theta)) + B\sin(2\pi f_0(t+\theta)), \tag{1.3.3}$$

also yields a periodic signal with frequency f_0. For a signal written in this form we no longer need to include the phase shift explicitly since

$$\cos(2\pi f_0(t+\theta)) = \cos(2\pi f_0 t)\cos(2\pi f_0\theta) - \sin(2\pi f_0 t)\sin(2\pi f_0\theta),$$

and

$$\sin(2\pi f_0(t+\theta)) = \sin(2\pi f_0 t)\cos(2\pi f_0\theta) + \cos(2\pi f_0 t)\sin(2\pi f_0\theta),$$

so that

$$z(t) = a\cos(2\pi f_0 t) + b\sin(2\pi f_0 t), \tag{1.3.4}$$

with the new amplitudes

$$a = A\cos(2\pi f_0\theta) + B\sin(2\pi f_0\theta), \qquad \text{and} \qquad b = B\cos(2\pi f_0\theta) - A\sin(2\pi f_0\theta).$$

The power of the signal $z(t)$, in view of (1.3.1) and (1.3.2), is given by the Pythagorean-like formula

$$\mathbf{PW}_z = \frac{1}{P}\int_0^P z^2(t)\,dt = \frac{1}{P}\int_0^P (a\cos(2\pi f_0 t) + b\sin(2\pi f_0 t))^2\,dt$$

$$= a^2\cdot\mathbf{PW}_x + b^2\cdot\mathbf{PW}_y + 2ab\frac{1}{P}\int_0^P \cos(2\pi f_0 t)\sin(2\pi f_0 t)\,dt = \frac{1}{2}(a^2+b^2), \tag{1.3.5}$$

because (see Tables 1.1 and 1.2, again)

$$\frac{1}{P}\int_0^P \cos(2\pi f_0 t)\sin(2\pi f_0 t)\,dt = \frac{1}{P}\int_0^P \frac{1}{2}\sin(4\pi f_0 t)\,dt = 0. \tag{1.3.6}$$

The above property (1.3.6), called *orthogonality* of the sine and cosine signals, will play a fundamental role in this book.

The next observation is that signals

$$z(t) = a\cos(2\pi(mf_0)t) + b\sin(2\pi(mf_0)t), \qquad m = 0, 1, 2, \ldots,$$

have the frequency equal to the multiplicity m of the *fundamental* frequency f_0, and as such have, in particular, period P (but also period P/m). Their power is

also equal to $(a^2 + b^2)/2$. So, if we superpose M of them, with possibly different amplitudes a_m and b_m, for different $m = 0, 1, 2, \ldots, M$, the result is a periodic signal

$$x(t) = \sum_{m=0}^{M} \Big(a_m \cos(2\pi (m f_0) t) + b_m \sin(2\pi (m f_0) t) \Big)$$

$$= a_0 + \sum_{m=1}^{M} \Big(a_m \cos(2\pi (m f_0) t) + b_m \sin(2\pi (m f_0) t) \Big) \tag{1.3.7}$$

with period P, and the fundamental frequency $f_0 = 1/P$, which has the mean and power

$$\mathbf{AV}_x = a_0, \qquad \text{and} \qquad \mathbf{PW}_x = a_0^2 + \frac{1}{2} \sum_{m=1}^{M} (a_m^2 + b_m^2). \tag{1.3.8}$$

The above result follows from the fact that not only sine and cosine signals (of arbitrary frequencies) are orthogonal to each other (see, (1.3.6)) but also cosines of different frequencies are *orthogonal* to each other, and so are sines. Indeed, if $m \neq n$, that is, $m - n \neq 0$, then

$$\frac{1}{P} \int_0^P \cos(2\pi m f_0 t) \cos(2\pi n f_0 t)\, dt \tag{1.3.9}$$

$$= \frac{1}{P} \int_0^P \frac{1}{2} \Big(\cos(2\pi (m-n) f_0 t) + \cos(2\pi (m+n) f_0 t) \Big)\, dt = 0,$$

and

$$\frac{1}{P} \int_0^P \sin(2\pi m f_0 t) \sin(2\pi n f_0 t)\, dt \tag{1.3.10}$$

$$= \frac{1}{P} \int_0^P \frac{1}{2} \Big(\cos(2\pi (m-n) f_0 t) - \cos(2\pi (m+n) f_0 t) \Big)\, dt = 0.$$

Example 1.3.1 (Superposition of Simple Cosine Oscillations) Consider the signal

$$x(t) = \sum_{m=1}^{12} \frac{1}{m^2} \cos(2\pi m t). \tag{1.3.11}$$

Its fundamental frequency is $f_0 = 1$, its average $\mathbf{AV}_x = 0$, and its power (see, (1.3.8))

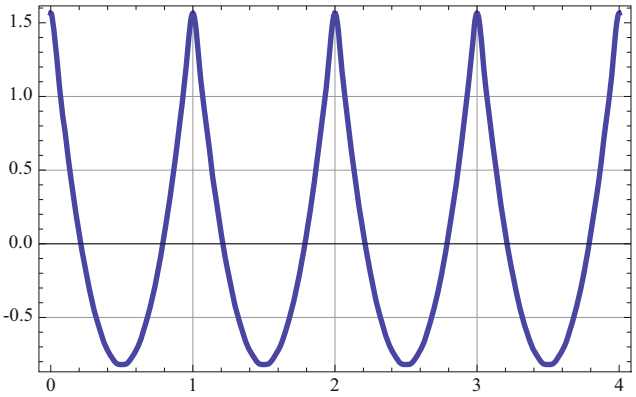


Fig. 1.9 Signal $x(t) = \sum_{m=1}^{12} m^{-2} \cos(2\pi mt)$ in its time-domain representation

$$\mathbf{PW}_x = \frac{1}{2} \sum_{m=1}^{12} \left(\frac{1}{m^2} \right)^2 \approx 0.541.$$

With its sharp cusps, the shape of the above signal is unlike that of any simple harmonic oscillation and one could start wondering what kind of other periodic signals can be well represented (approximated) by superpositions of harmonic oscillations of the form (1.3.7). The answer, discussed at length in Chap. 2, is that almost all of them can, as long as their power is finite (Fig. 1.9).

The Frequency Domain Description The signal $x(t)$ in Example 1.3.1 would be completely specified if, instead of writing the whole formula (1.3.11), we just listed the frequencies present in the signal and the corresponding amplitudes, that is, considered the list

$$(1, 1/1^2), (2, 1/2^2), (3, 1/3^3), \ldots, (12, 1/12^2).$$

Similarly, in the case of the general superposition (1.3.7), it would be sufficient to list the cosine and sine frequencies and associated amplitudes, that is, compile the lists

$$(0, a_0), (1f_0, a_1), (2f_0, a_2), \ldots, (Mf_0, a_M), \tag{1.3.12}$$

and

$$(1f_0, b_1), (2f_0, b_2), \ldots, (Mf_0, b_M). \tag{1.3.13}$$

The lists (sequences) ((1.3.12) and (1.3.13)) are called the *frequency domain (spectral) representation* of the signal (1.3.7).

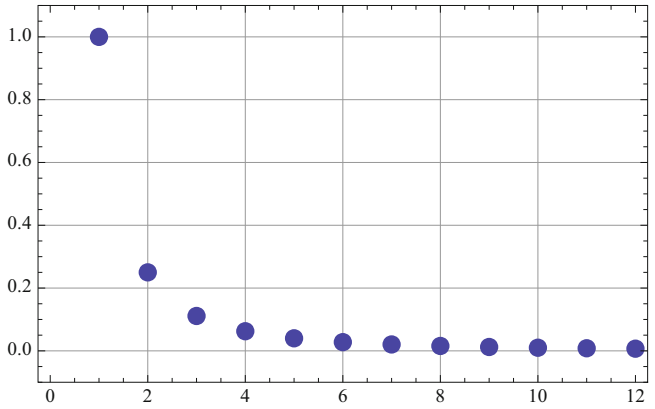


Fig. 1.10 Signal $x(t) = \sum_{m=1}^{12} m^{-2} \cos(2\pi mt)$ in its frequency domain representation. Only the amplitudes of frequencies $m = 1, 2, \ldots, 12$, are shown since all the phase shifts are zero

Remark 1.3.1 (Amplitude-Phase Form of the Spectral Representation) Alternatively, if the signal $x(t)$ in (1.3.7) is rewritten in the amplitude-phase form,

$$x(t) = \sum_{m=0}^{M} c_n \cos(2\pi(mf_0)(t + \theta_m)),$$

then the frequency domain representation must list the frequencies present in the signal, mf_0, $m = 0, 1, \ldots, M$, and the corresponding amplitudes $c_m\, m = 0, 1, \ldots, M$, and phases θ_m, $m = 0, 1, \ldots, M$.

For the signal from Example 1.3.1, such a representation is graphically pictured in Fig. 1.10. We will see in Chap. 2 that, for any periodic signal, the spectrum is always concentrated on a discrete set of frequencies, namely, the multiplicities of the fundamental frequency.

Finally, the formula (1.3.8) shows how the total power of signal $x(t)$ is distributed over different frequencies. Such a distribution, provided by the list

$$(0, a_0^2), (1f_0, (a_1^2 + b_1^2)/2), (2f_0, (a_2^2 + b_2^2)/2), \ldots, (Mf_0, (a_M^2 + b_M^2)/2), \tag{1.3.14}$$

is called the *power spectrum* of the periodic signal (1.3.7).

Observe that, in general, knowledge of the power spectrum is not sufficient for the reconstruction of the signal $x(t)$ itself, while knowledge of the whole representation in the frequency domain is.

To complete our elementary study of periodic signals note that if an arbitrary signal is studied only in a finite time interval $[0, P]$, then it can always be treated as a periodic signal with period P since one can extend its definition periodically to the whole time line by copying its waveform from the interval $[0, P]$ to intervals $[P, 2P]$, $[2P, 3P]$, and so on.

1.4 Building a Better Mousetrap: Complex Exponentials

Catching the structure of periodic signals via their decomposition into a superposition of basic trigonometric functions leads to some cumbersome calculations employing various trigonometric identities (as we have seen in Sect. 1.3). A greatly simplified and also more elegant approach to the same problem employs a representation of trigonometric functions in terms of exponential functions of the imaginary variable. The cost of moving into the complex domain is not high as we will rely, essentially, on a single relationship

$$e^{j\alpha} = \cos\alpha + j\sin\alpha, \quad \text{where} \quad j = \sqrt{-1}, \tag{1.4.1}$$

which is known as *de Moivre formula*,[3] and which immediately yields two identities

$$\cos\alpha = \frac{1}{2}(e^{j\alpha} + e^{-j\alpha}), \qquad \text{and} \qquad \sin\alpha = \frac{1}{2j}(e^{j\alpha} - e^{-j\alpha}). \tag{1.4.2}$$

In what follows, we are going to routinely utilize the complex number techniques. Thus, for the benefit of the reader, the basic notation and facts about them are summarized in Table 1.2.

Since de Moivre formula is so crucial for us, it is important to understand where it is coming from. The proof is straightforward and relies on the power series expansion of the exponential function,

$$e^{j\alpha} = \sum_{k=0}^{\infty} \frac{j^k \alpha^k}{k!}. \tag{1.4.3}$$

However, the powers of the imaginary unit j can be expressed via a simple formula

$$j^k = \begin{cases} 1, & \text{if} k = 4m; \\ j, & \text{if} k = 4m+1; \\ -1, & \text{if} k = 4m+2; \\ -j, & \text{if} k = 4m+3, \end{cases}$$

so the whole series (1.4.3) splits neatly into the real part, corresponding to even indices of the form $k = 2n, n = 0, 1, 2, \ldots,$, and the imaginary part, corresponding to the odd indices of the form $k = 2n+1, n = 0, 1, 2, \ldots$:

[3] Throughout this book we denote the imaginary unit $\sqrt{-1}$ by the letter j, which is a standard usage in the electrical engineering signal processing literature as the, usual in the mathematical literature, letter i is reserved for electrical current.

$$\sum_{k=0}^{\infty} \frac{j^k \alpha^k}{k!} = \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n}}{(2n)!} + j \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n+1}}{(2n+1)!}.$$

Now, it suffices to recognize in the above formula the familiar power series expansions for trigonometric functions,

$$\cos\alpha = \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n}}{(2n)!}, \qquad \sin\alpha = \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n+1}}{(2n+1)!}$$

to obtain de Moivre formula.

Given de Moivre formulas which provides a representation of sine and cosine functions via the complex exponentials, we can now rewrite the general superposition of harmonic oscillation

$$x(t) = a_0 + \sum_{m=1}^{M} a_m \cos(2\pi m f_0 t) + \sum_{m=1}^{M} b_m \sin(2\pi m f_0 t), \tag{1.4.4}$$

in terms of the complex exponentials

$$x(t) = \sum_{m=-M}^{M} z_m e^{j2\pi m f_0 t}, \tag{1.4.5}$$

with the real amplitudes, a_m and b_m, in representations (1.4.4), and the complex amplitudes z_m in the representation (1.4.5), connected by the formulas

$$a_0 = z_0, \qquad a_m = z_m + z_{-m}, \qquad b_m = j(z_m - z_{-m}), \qquad m = 1, 2, \ldots,$$

or, equivalently,

$$z_0 = a_0 \qquad z_m = \frac{a_m - j b_m}{2}, \qquad z_{-m} = \frac{a_m + j b_m}{2}, \qquad m = 1, 2, \ldots$$

The above relationships show that for the signal of the form (1.4.5) to represent a real-valued signal $x(t)$ it is necessary and sufficient that the paired amplitudes for symmetric frequencies, mf_0 and $-mf_0$, be complex conjugates of each other:

$$z_{-m} = z_m^*, \qquad m = 1, 2, \ldots \tag{1.4.6}$$

However, in the future it will be convenient to consider general complex-valued signals of the form (1.4.5) without the restriction (1.4.6) on its complex amplitudes.

At the first sight, the above introduction of complex numbers and functions of complex-valued variables may seem as an unnecessary complication in the analysis of signals. But let us calculate the power of the signal $x(t)$ given by (1.4.5). The

need for unpleasant trigonometric formulas disappears as now we need to integrate only exponential functions. Indeed, remembering the $|z|^2 = z \cdot z^*$ now stands for the square of the modulus of a complex number, we have

$$\mathbf{PW}_x = \frac{1}{P}\int_{t=0}^{P} |x(t)|^2\,dt = \frac{1}{P}\int_{t=0}^{P} \Big|\sum_{m=-M}^{M} z_m e^{j2\pi m f_0 t}\Big|^2 dt$$

$$= \frac{1}{P}\int_{t=0}^{P} \Big(\sum_{m=-M}^{M} z_m e^{j2\pi m f_0 t} \cdot \sum_{k=-M}^{M} z_k^* e^{-j2\pi k f_0 t}\Big)\,dt$$

$$= \frac{1}{P}\sum_{m=-M}^{M}\sum_{k=-M}^{M} z_m z_k^* \int_{t=0}^{P} e^{j2\pi(m-k)f_0 t}\,dt = \sum_{m=-M}^{M} |z_m|^2, \tag{1.4.7}$$

because, for $m - k \neq 0$,

$$\frac{1}{P}\int_{t=0}^{P} e^{j2\pi(m-k)f_0 t}\,dt = \frac{1}{j2\pi(m-k)f_0} e^{j2\pi(m-k)f_0 t}\Big|_{t=0}^{P} = 0, \tag{1.4.8}$$

as the function $e^{j2\pi(m-k)f_0 t} = \cos(2\pi(m-k)f_0 t) + j\sin(2\pi(m-k)f_0 t)$ is periodic with period P, and for $m - k = 0$,

$$\frac{1}{P}\int_{t=0}^{P} e^{j2\pi(m-k)f_0 t}\,dt = 1. \tag{1.4.9}$$

Thus all the off-diagonal terms in the double sum in (1.4.7) disappear. The formulas (1.4.8) and (1.4.9) express mutual orthogonality and normalization of the complex exponential signals,

$$e^{j2\pi m f_0 t}, \qquad m = 0, \pm 1, \pm 2, \ldots, \pm M.$$

In view of (1.4.7), the distribution of the power of the signal (1.4.5) over different multiplicities of the fundamental frequency f_0 can be written as a list with simple structure,

$$(mf_0, |z_m|^2), \qquad m = 0, \pm 1, \pm 2, \ldots, \pm M. \tag{1.4.10}$$

Remark 1.4.1 (Aperiodic Signals) Nonperiodic signals can also be analyzed in terms of their frequency domains but their spectra are not discrete. We will study them later on.

1.5 Problems and Exercises

1* Find the real and imaginary parts of $(j+3)/(j-3)$; $(1+j\sqrt{2})^3$; $1/(2-j)$; $(2-3j)/(3j+2)$.[4]

2* Find the moduli $|z|$ and arguments θ of complex numbers $z=5$; $z=-2j$; $z=-1+j$; $z=3+4j$.

3* Find the real and imaginary components of complex numbers $z=5\,e^{j\pi/4}$; $z=-2\,e^{j(8\pi+1.27)}$; $z=-1\,e^{j}$; $z=3\,e^{je}$.

4* Show that

$$\frac{5}{(1-j)(2-j)(3-j)}=\frac{j}{2},\quad \text{and}\quad (1-j)^4=-4.$$

5* Sketch sets of points in complex plane (x,y), $z=x+jy$, such that $|z-1+j|=1$; $|z+j|\le 3$; $\mathrm{Re}\,(z^*-j)=2$; $|2z-j|=4$; $z^2+(z^*)^2=2$.

6* Using de Moivre's formulas find $(-2j)^{1/2}$ and $\mathrm{Re}\,(1-j\sqrt{3})^{77}$. Are these complex numbers uniquely defined?

7 Write the signal $x(t)=\sin t+\cos(3t)/3$ from Fig. 1.1 as a sum of phase-shifted cosines.

8 Using de Moivre's formulas write the signal $x(t)=\sin t+\cos(3t)/3$ from Fig. 1.1 as a sum of complex exponentials.

9 Find the time average and power of the signal $x(t)=-2e^{-j2\pi 4t}+3e^{-j2\pi t}+1-2e^{j2\pi 3t}$. What is the fundamental frequency of this signal? Plot the distribution of power of $x(t)$ over different frequencies. Write this (complex) signal in terms of cosines and sines. Find and plot its real and imaginary parts.

10* Using de Moivre's formula derive the complex exponential representation (1.4.5) of the signal $x(t)$ given by the cosine series representation $x(t)=\sum_{m=1}^{M}c_m\cos(2\pi m f_0(t+\theta_m))$.

11 Find the time average and power of the signal $x(t)$ from Fig. 1.9. Use a symbolic manipulation language such as *Mathematica* or *Matlab* if you like.

12* Using a computing platform such as *Mathematica*, *Maple*, or *Matlab* produces plots of the signals

$$x_M(t)=\frac{\pi}{4}+\sum_{m=1}^{M}\left[\frac{(-1)^m-1}{\pi m^2}\cos mt-\frac{(-1)^m}{m}\sin mt\right],$$

[4] Solutions of the problems marked by the asterisk can be found at the end of the book in the chapter *Solutions to Selected Problems and Exercises.*

for $M = 0, 1, 2, 3, \ldots, 9$ and $-2\pi < t < 2\pi$. Then produce their plots in the frequency-domain representation. Calculate their power (again, using *Mathematica*, *Maple*, or *Matlab*, if you wish). Produce plots showing how power is distributed over different frequencies for each of them. Write down your observations. What is likely to happen with the plots of these signals as we take more and more terms of the above series, that is, as $M \to \infty$? Is there a limit signal $x_\infty(t) = \lim_{M\to\infty} x_M(t)$? What could it be?

13* Use the analog-to-digital conversion formula (1.1.1) to digitize signals from Problem 13 for a variety of sampling periods and resolutions. Plot the results.

14* Use your computing platform to produce a discrete-time signal consisting of a string of random numbers uniformly distributed on the interval [0,1]. For example, in *Mathematica*, the command

```
Table[Random[], {20}]
```

will produce the following string of 20 random numbers between 0 and 1:

```
{0.175245, 0.552172, 0.471142, 0.910891, 0.219577,
0.198173, 0.667358, 0.226071, 0.151935, 0.42048,
0.264864, 0.330096, 0.346093, 0.673217, 0.409135,
0.265374, 0.732021, 0.887106, 0.697428, 0.7723}
```

Use the "random numbers" string as additive noise to produce random versions of the digitized signals from Problem 14. Follow the example described in Fig. 1.3. Experiment with different string length and various noise amplitudes. Then center the noise around zero and repeat your experiments.

Chapter 2
Spectral Representation of Deterministic Signals: Fourier Series and Transforms

Abstract In this chapter we will take a closer look at the spectral or frequency-domain representation of deterministic (nonrandom) signals which was already mentioned in Chap. 1. The tools introduced below, usually called *Fourier* or *harmonic analysis*, will play a fundamental role later on in our study of random signals. Almost all of the calculations will be conducted in the complex form. Compared with working in the real domain, manipulation of formulas written in the complex form turns out to be simpler and all the tedium of remembering various trigonometric formulas is avoided. All of the results written in the complex form can be translated quickly into results for real trigonometric series expressed in terms of sines and cosines via the familiar from Chap. 1 de Moivre's formula, $e^{jt} = \cos t + j \sin t$.

2.1 Complex Fourier Series for Periodic Signals

Any finite-power, complex-valued signal $x(t)$, periodic with period P (say, seconds), can[1] be written in the form of an *infinite* complex Fourier series, meant as a limit (in a sense to be made more precise later), for $M \to \infty$, of finite superposition of complex harmonic exponentials discussed in Sect. 1.4 (see (1.4.5)):

$$x(t) = \sum_{m=-\infty}^{\infty} z_m e^{j2\pi m f_0 t} = \sum_{m=-\infty}^{\infty} z_m e^{jm\omega_0 t}, \tag{2.1.1}$$

where $f_0 = \frac{1}{P}$ is the *fundamental frequency* of the signal (measured in Hz=1/s), and $\omega_0 = 2\pi f_0$ is called the fundamental *angular velocity* (measured in radians/s). The complex number z_m, where m can take values $\ldots, -2, -1, 0, 1, 2, \ldots$, is called

[1]For mathematical issues related to the feasibility of such a representation, see the discussion in the subsection of this section devoted to the analogy between the orthonormal basis in a 3D space and complex exponentials.

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_2

the m-th Fourier coefficient of signal $x(t)$. Think about it as the amplitude of the harmonic component, with the frequency mf_0, of the signal $x(t)$.

In this text we will carry out our calculations exclusively in terms of the fundamental frequency f_0, although one can find in the printed and software signal processing literature sources where all the work is done in terms of ω_0. It is an arbitrary choice, but some formulas are simpler if written in the frequency domain; transition from one system to the other is easily accomplished by adjusting various constants appearing in the formulas.

The infinite Fourier series representation (2.1.1) is unique in the sense that two different signals[2] will have two different sequences of Fourier coefficients. The uniqueness is a result of the fundamental property of complex exponentials

$$e_m(t) := e^{j2\pi m f_0 t}, \qquad m = \ldots, -2, -1, 0, 1, 2, \ldots \tag{2.1.2}$$

called *orthonormality*:

The scalar product (sometimes also called inner, or dot, product) of two complex exponentials e_n and e_m is 0 if the exponentials are different, and it is 1 if they are the same. Indeed,

$$\langle e_n, e_m \rangle := \frac{1}{P} \int_0^P e_n(t) e_m^*(t)\, dt \tag{2.1.3}$$

$$= \frac{1}{P} \int_0^P e^{j2\pi(n-m) f_0 t}\, dt = \begin{cases} 0, & if\ n \neq m; \\ 1, & if\ n = m. \end{cases}$$

Recall that, for a complex number $z = a + jb = |z|e^{j\theta}$ with real component a and imaginary component b, the complex conjugate $z^* = a - jb = |z|e^{-j\theta}$. Sometimes it is convenient to describe the orthonormality using the so-called *Kronecker delta* notation:

$$\delta(n) = \begin{cases} 0, & \text{if} n \neq 0; \\ 1, & \text{if} n = 0. \end{cases}$$

Then, simply,

$$\langle e_m, e_n \rangle = \delta(n - m).$$

Using the orthonormality property we can directly evaluate the coefficients z_m in the Fourier series (2.1.1) of a given signal $x(t)$ by formally calculating the scalar product of $x(t)$ and $e_m(t)$:

[2] Meaning that their difference has positive power.

$$\langle x, e_m\rangle = \frac{1}{P}\int_0^P \left(\sum_{n=-\infty}^{\infty} z_n e_n(t)\right)\cdot e_m^*(t)\,dt \tag{2.1.4}$$

$$= \sum_{n=-\infty}^{\infty} z_n \frac{1}{P}\int_0^P e_n(t)e_m^*(t)\,dt = z_m,$$

so that we get an explicit formula for the Fourier coefficient of signal $x(t)$,

$$z_m = \langle x, e_m\rangle = \frac{1}{P}\int_0^P x(t)e^{-j2\pi m f_0 t}\,dt. \tag{2.1.5}$$

Thus the basic Fourier expansion (2.1.1) can now be rewritten in the form of a formal identity

$$x(t) = \sum_{n=-\infty}^{\infty} \langle x, e_n\rangle e_n(t). \tag{2.1.6}$$

It is worthwhile to recognize that the above calculations on infinite series and interchanges of the order of integration and infinite summations were purely formal, that is, the soundness of the limit procedures was not rigorously established. The missing steps can be found in the mathematical literature devoted to Fourier analysis.[3] For our purposes suffice it to say that if a periodic signal $x(t)$ has finite power

$$\mathbf{PW}_x = \|x\|_{L^2}^2 := \frac{1}{P}\int_0^P |x(t)|^2\,dt < \infty, \tag{2.1.7}$$

and the concept of convergence of the functional infinite series (2.1.1) is defined in the right way, then all of the above formal manipulations can be rigorously justified. We will return to this issue at the end of this section. In what follows we will usually consider signals with finite power.

Real-Valued Signals Signal $x(t)$ is real-valued if and only if the coefficients z_m satisfy the algebraic condition,

$$z_{-m} = z_m^*, \tag{2.1.8}$$

in which case cancellation of the imaginary parts in the Fourier series (2.1.1) occurs. Indeed, under assumption (2.1.8),

$$z_m = |z_m|e^{j\theta_m}, \quad \theta_{-m} = -\theta_m, \tag{2.1.9}$$

[3] See, e.g., A. Zygmund, *Trigonometric Series*, Cambridge University Press.

and, since

$$\frac{e^{j\alpha} + e^{-j\alpha}}{2} = \cos\alpha,$$

we get

$$x(t) = c_0 + \sum_{m=1}^{\infty} c_m \cos(2\pi m f_0 t + \theta_m), \tag{2.1.10}$$

where

$$c_0 = z_0, \quad \text{and} \quad c_m = 2|z_m|, \quad m = 1, 2, \ldots \tag{2.1.11}$$

The power $\mathbf{PW}_x$ of a periodic signal $x(t)$ given by its Fourier series (2.1.1) can also be directly calculated from its Fourier coefficient z_m. Indeed, again calculating formally, we obtain that

$$\mathbf{PW}_x = \frac{1}{P}\int_0^P |x(t)|^2\,dt = \frac{1}{P}\int_0^P x(t)x^*(t)\,dt$$

$$= \frac{1}{P}\int_0^P \left(\sum_{k=-\infty}^{\infty} z_k e_k(t)\right)\cdot\left(\sum_{m=-\infty}^{\infty} z_m e_m(t)\right)^* dt$$

$$= \sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} z_k z_m^* \frac{1}{P}\int_0^P e_k(t)e_m^*(t)\,dt = \sum_{m=-\infty}^{\infty} z_m z_m^*,$$

in view of the orthonormality (2.1.3) of the complex exponentials. The multiplication of the two infinite series was carried out term-by-term. The resulting relationship

$$\mathbf{PW}_x = \frac{1}{P}\int_0^P |x(t)|^2\,dt = \sum_{m=-\infty}^{\infty} |z_m|^2 \tag{2.1.12}$$

is known as the *Parseval formula*. A similar calculation for the scalar product $(1/P)\int_0^P x(t)y^*(t)$ of two different periodic signals, $x(t)$ and $y(t)$, gives an *extended Parseval formula* listed in Table 2.1.

Remark 2.1.1 (Distribution of Power Over Frequencies in a Periodic Signal) Parseval's formula describes how the power $\mathbf{PW}_x$ of the signal $x(t)$ is distributed over different frequencies. The sequence (or its plot)

$$(mf_0, |z_m|^2), \quad m = 0, \pm 1, \pm 2, \ldots \tag{2.1.13}$$

Table 2.1 Analogy between orthogonal expansions in 3D and in the space of periodic signals with finite power

Objects	
3D vectors	*Signals with finite power*
$\vec{x} = (x_1, x_2, x_3)$	$x(t) = \sum_{m=-\infty}^{\infty} z_m e_m(t),$
$\vec{y} = (y_1, y_2, y_3)$	$y(t) = \sum_{m=-\infty}^{\infty} w_m e_m(t),$
Bases	
Unit coordinate vectors	*Complex exponentials*
	$\vdots$
$\vec{e_1} = (1, 0, 0)$	$e_1(t) = e^{j2\pi f_0 t}$
$\vec{e_2} = (0, 1, 0)$	$e_2(t) = e^{j2\pi(2f_0)t}$
$\vec{e_3} = (0, 0, 1)$	$e_3(t) = e^{j2\pi(3f_0)t}$
	$\vdots$
Scalar products	
$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{3} x_i y_i$	$\langle x(t), y(t) \rangle = \frac{1}{P} \int_0^P x(t) y^*(t) dt$
Orthonormality	
$\langle \vec{e_m}, \vec{e_n} \rangle = \delta(n-m)$	$\langle e_m(t), e_n(t) \rangle = \delta(n-m)$
Expansions	
Basis	*Fourier*
$\vec{x} = \sum_{m=1}^{3} \langle \vec{x}, \vec{e_m} \rangle \vec{e_m}$	$x(t) = \sum_{m=-\infty}^{\infty} \langle x, e_m \rangle e_m(t)$
Formulas	
Pythagoras'	*Parseval's*
$\|\|\vec{x}\|\|^2 = \sum_{m=1}^{3} x_m^2$	$\mathbf{PW}_x = \frac{1}{P} \int_0^P \|x(t)\|^2 dt = \sum_{m=-\infty}^{\infty} \|z_m\|^2$
Scalar product	*Extended Parseval's*
$\langle \vec{x}, \vec{y} \rangle = \sum_{m=1}^{3} x_m y_m$	$\frac{1}{P} \int_0^P x(t) y^*(t)\, dt = \sum_{m=-\infty}^{\infty} z_m w_m^*$

is called the *power spectrum* of the signal $x(t)$. Simply stated, it says that the harmonic component of $x(t)$, with frequency mf_0, has power $|z_m|^2$ (always a nonnegative number!).

Analogy between the orthonormal basis of vectors in the 3D space $\mathbf{R}^3$ *and the complex exponentials; the completeness theorem:* It is useful to think about the complex exponentials $e_m(t) = e^{2\pi j m f_0 t}, m = \ldots, -1, 0, 1, \ldots,$ as an infinite-dimensional version of the orthonormal basic vectors in $\mathbf{R}^3$. In this mental picture the periodic signal $x(t)$ is now thought of as an infinite-dimensional "vector" uniquely expandable into an infinite linear combination of the complex exponentials in the same way a 3D vector is uniquely expandable into a finite linear combination of the three unit coordinate vectors. Table 2.1 describes this analogy more fully. Note that the Parseval formula can now be seen just as an infinite-dimensional extension of the familiar Pythagorean theorem.

So far, the delicate issue of the very feasibility of the Fourier expansion (2.1.1) for any periodic signal with finite power has been left out. Note that in the 3D case, the fact that any vector $\vec{x}$ is representable in the form $x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3$,

where $\vec{e}_1, \vec{e}_2, \vec{e}_3$, are the unit coordinate vector, is due to the fact that $\vec{e}_1, \vec{e}_2, \vec{e}_3$, is a "maximal" system of orthogonal vectors in 3D; it cannot be further expanded. In other words, if a vector $\vec{e}$ is orthogonal to $\vec{e}_1, \vec{e}_2, \vec{e}_3$, then it must be zero. A similar situation arises if one considers the system of *all* basic harmonic complex exponentials,[4] $e_m(t) = e^{2\pi jmf_0 t}, m = \ldots, -1, 0, 1, \ldots$, in the space of finite power periodic complex signals with period $P = 1/f_0$. If $x(t)$ is such a signal and $\langle x(t), e_m(t)\rangle = 0$, for all $m = \ldots, -1, 0, 1, \ldots$, then necessarily $x(t) = 0$. This fact is known as the *Completeness Theorem* for complex exponentials and one can find its proof in any mathematical textbook on harmonic or functional analysis. Removing even one of the complex exponentials from the above system creates an incomplete orthonormal system.

Examples Recall that a signal is called *even* if it is symmetric under the change of the direction of time, i.e., if $x(t) = x(-t)$; it is called *odd* if it is antisymmetric under the change of the direction of time, i.e., if $x(t) = -x(-t)$. The real Fourier expansion of a real-valued signal $x(t)$ the periodic extension thereof to the whole real line is even, i.e., $x(t) = x(-t)$, for all $t \in \mathbf{R}$, will contain only cosine functions (which are even) and, similarly, the real Fourier expansion of an odd real-valued signal $x(t) = -x(-t)$ will contain only sine functions (which are odd). This phenomenon will be illustrated in the following examples.

Example 2.1.1 (Pure Cosine Expansion of an Even Rectangular Waveform) Consider a rectangular waveform with period P, and amplitude $a > 0$, defined by the formula

$$x(t) = \begin{cases} a, & \text{for} 0 \le t < P/4; \\ 0, & \text{for} P/4 \le t < 3P/4; \\ a, & \text{for} 3P/4 \le t < P. \end{cases}$$

The signal is pictured below, for particular values $P = 1$, and $a = 1$ (Fig. 2.1).

Calculation of coefficients z_m in the expansion of the signal $x(t)$ into a complex Fourier series is here straightforward: For $m = 0$,

$$z_0 = \frac{1}{P}\int_0^P x(t)e^{-j2\pi 0t/P}dt = \frac{a}{P}\left(\frac{P}{4} - 0 + P - \frac{3P}{4}\right) = \frac{a}{2}.$$

In the case $m \neq 0$,

$$z_m = \frac{1}{P}\int_0^P x(t)e^{-j2\pi mt/P}dt$$

[4] Note that the sequence also includes the constant $e_0(t) \equiv 1$.

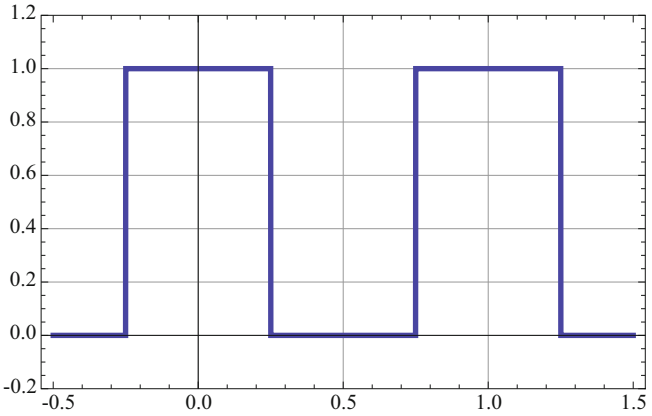


Fig. 2.1 An even rectangular waveform signal from Example 2.1.1. The period $P = 1$ and the amplitude $a = 1$

$$= \frac{a}{P}\left(\int_0^{P/4} e^{-j2\pi mt/P}dt + \int_{3P/4}^{P} e^{-j2\pi mt/P}dt\right)$$

$$= \frac{a}{P}\left(\frac{P}{-j2\pi m}e^{-j2\pi mt/P}\Big|_0^{P/4} + \frac{P}{-j2\pi m}e^{-j2\pi mt/P}\Big|_{3P/4}^{P}\right)$$

$$= \frac{a}{-j2\pi m}\left(e^{-j(\pi/2)m} - 1 - e^{-j(3\pi/2)m} + 1\right)$$

$$= -\frac{a}{\pi m}e^{-j(2\pi/2)m}\left(\frac{e^{j(\pi/2)m} - e^{-j(\pi/2)m}}{2j}\right)$$

$$= -\frac{a}{\pi m}\cos\pi m \ \sin\frac{\pi}{2}m = -\frac{a}{\pi m}(-1)^m \sin\frac{\pi}{2}m.$$

If $m = 2k$, then $\sin((\pi/2)m) = 0$, and if $m = 2k + 1$, $k = 0, \pm1, \pm2, \ldots$, then $\sin(\pi/2)m = (-1)^k$, which gives, for $k = \pm1, \pm2, \ldots$,

$$z_{2k} = 0,$$

and

$$z_{2k+1} = \frac{-a}{\pi(2k+1)}(-1)^{2k+1}(-1)^k = \frac{(-1)^k a}{\pi(2k+1)}.$$

Thus the complex Fourier expansion of the signal $x(t)$ is

$$x(t) = \frac{a}{2} + \frac{a}{\pi}\sum_{k=-\infty}^{\infty} \frac{(-1)^k}{2k+1} e^{j2\pi(2k+1)t/P}.$$

Observe that for any $m = \ldots, -1, 0, 1, \ldots$, we have $z_m = z_{-m}$. Pairing up complex exponentials with the exponents of opposite signs, and using de Moivre's formula, we arrive at the real Fourier expansion that contains only cosine functions:

$$x(t) = \frac{a}{2} + \frac{a}{\pi}\left(2\cos(2\pi t/P) - \frac{2}{3}\cos(2\pi 3t/P) + \ldots\right).$$

Example 2.1.2 (Pure Sine Expansion of an Odd Rectangular Waveform) Consider a periodic rectangular waveform of period P which is defined by the formula

$$x(t) = \begin{cases} a, & \text{for} 0 \le t < P/4; \\ 0, & \text{for} P/4 \le t < 3P/4; \\ -a, & \text{for} 3P/4 \le t < P. \end{cases}$$

The signal is pictured below for particular values $P = 1$ and $a = 1$ (Fig. 2.2).

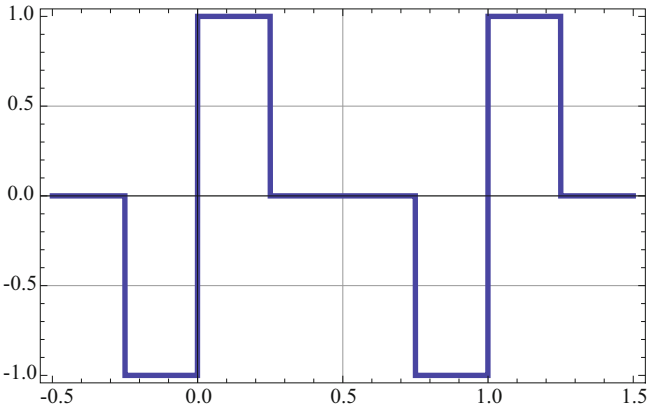


Fig. 2.2 An odd rectangular waveform signal from Example 2.1.2. The period $P = 1$ and the amplitude $a = 1$

For $m = 0$,

$$z_0 = \frac{1}{P}\int_0^P x(t)dt = 0,$$

and, for $m \neq 0$,

$$z_m = \frac{a}{P}\left(\int_0^{P/4} e^{-j2\pi mt/P}dt - \int_{3P/4}^{P} e^{-j2\pi mt/P}dt\right)$$

$$= \frac{-a}{j2\pi m}\left(e^{-j(\pi/2)m} - 1 - 1 + e^{-j(3\pi/2)m}\right)$$

$$= -\frac{a}{j2\pi m}\left[e^{-j(2\pi/2)m}\left(e^{j(\pi/2)m} + e^{-j(\pi/2)m}\right) - 2\right]$$

$$= -\frac{a}{j\pi m}\left((-1)^m \cdot \cos\frac{\pi}{2}m - 1\right),$$

since, by de Moivre's formula, $e^{-j\pi m} = \cos\pi m - j\sin\pi m$, and $\cos\pi m = (-1)^m$, and $\sin\pi m = 0$, for any integer m. On the other hand, $\cos(\pi/2)m = 0$ if m is odd, and $= (-1)^k$ when $m = 2k$ is even, so we get that

$$z_m = \begin{cases} a/(j\pi(2k+1)), & \text{for odd } m = 2k+1; \\ a[1-(-1)^k]/(j\pi 2k), & \text{for even } m = 2k. \end{cases}$$

Thus the complex Fourier series of the signal $x(t)$ is of the form

$$x(t) = \frac{a}{\pi}\sum_{k=-\infty}^{\infty}\left[\frac{1}{j(2k+1)}e^{j2\pi(2k+1)t/P} + \frac{[1-(-1)^k]}{j2k}e^{j2\pi(2k)t/P}\right].$$

Observe that in this case, for any $m = \ldots, -1, 0, 1, \ldots$, we have $z_m = -z_{-m}$, so pairing-up the exponentials with opposite signs in the exponents, and using de Moivre's formula, we get a real Fourier series expansion for $x(t)$ that contains only sine functions:

$$x(t) = \frac{2a}{\pi}\left[\sin 2\pi(1)t/P + \sin(2\pi(2)t/P) + \frac{1}{3}\sin(2\pi(3)t/P)\right.$$

$$\left.+0\cdot\sin(2\pi(4)t/P) + \frac{1}{5}\sin(2\pi(5)t/P) + \frac{1}{6}\sin(2\pi(6)t/P) + \ldots\right]$$

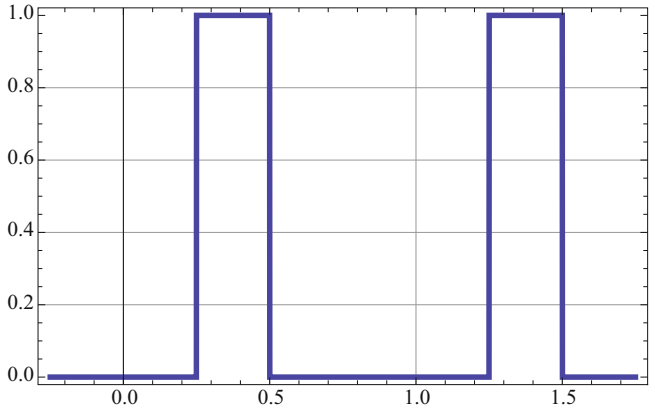


Fig. 2.3 A neither odd nor even rectangular waveform signal from Example 2.1.3. The period $P = 1$ and the amplitude $a = 1$

The purpose of going through the above example was to show that, for irregular periodic signals, the calculation of Fourier coefficients can get quite messy although the final result may display a pleasing symmetry.

Example 2.1.3 (A General Expansion for a Rectangular Waveform Which Is Neither Odd Nor Even) Consider a periodic rectangular waveform of period P which is defined by the formula

$$x(t) = \begin{cases} 0, & \text{for } 0 \le t < P/4; \\ a, & \text{for } P/4 \le t < P/2; \\ 0, & \text{for } P/2 \le t < P. \end{cases}$$

The signal is pictured below for parameter values $P = 1$ and $a = 1$ and, for simplicity's sake, we will carry out our calculations only in that case (Fig. 2.3).
For $m = 0$,

$$z_0 = \int_{1/4}^{1/2} 1\, dt = \frac{1}{4}.$$

For $m \neq 0$,

$$z_m = |z_m| e^{i\theta_m} = \int_{1/4}^{1/2} e^{-j2\pi mt} dt = \frac{1}{-j2\pi m}\left[e^{-j2\pi m/2} - e^{-j2\pi m/4}\right]$$

$$= \frac{1}{\pi m} e^{-j3\pi m/4} \left(\frac{e^{j\pi m/4} - e^{-j\pi m/4}}{2j}\right) = \frac{1}{\pi m} \sin(\frac{\pi}{4} m) e^{-j3\pi m/4}.$$

Thus

$$|z_m| = \frac{1}{\pi m}\sin(\pi m/4), \quad \text{and} \quad \theta_m = -3\pi m/4,$$

and the complex Fourier series for $x(t)$ is

$$x(t) = \frac{1}{4} + \sum_{m=-\infty,\, m\neq 0}^{\infty} \frac{1}{\pi m}\sin(\pi m/4)e^{-j3\pi m/4}e^{j2\pi mt}.$$

Again, pairing-up the complex exponentials with opposite signs in the exponents we obtain the real expansion in terms of the cosines, but this time with phase shifts that depend on m:

$$x(t) = \frac{1}{4} + \sum_{m=1}^{\infty} \frac{2}{\pi m}\sin(\pi m/4)\cos\left(2\pi mt - 3\pi m/4\right),$$

which, using the trigonometric formula $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$, also can be written as a general real Fourier series,

$$x(t) = a_0 + \sum_{m=1}^{\infty} a_m\cos(2\pi mt) + b_m\sin(2\pi mt),$$

with

$$a_0 = \frac{1}{4}, \qquad a_m = \frac{2}{\pi m}\sin\frac{\pi m}{4}\cos\frac{3\pi m}{4}, \qquad b_m = \frac{2}{\pi m}\sin\frac{\pi m}{4}\sin\frac{3\pi m}{4}.$$

2.2 Approximation of Periodic Signals by Finite Fourier Sums

Up to this point the equality in the Fourier series representation

$$x(t) = \sum_{m=-\infty}^{\infty} \langle x, e_m\rangle e_m(t),$$

for periodic signals, or its real version in terms of sine and/or cosine functions, was understood only formally. But, of course, usefulness of such an expansion will depend on whether we can show that the signal $x(t)$ can be well approximated by a finite cut-off of the infinite Fourier series, that is, on whether we can prove that

$$x(t) \approx s_M(t) := \sum_{m=-M}^{M} \langle x, e_m \rangle e_m(t) \tag{2.2.1}$$

for M large enough, with the error in the above approximate equality $\approx$ rigorously estimated.

One can pursue here several options:

Approximation in Power—Mean-Square Error If the error of approximation is measured as the power of the difference between the signal $x(t)$ and the finite Fourier sum $s_M(t)$ in (2.2.1), then the calculation is relatively simple and the error is often called the *mean-square error*. Indeed, using the Parseval formula,

$$\mathbf{PW}_{x-s_M} = \|x - s_M\|_{L^2}^2 := \frac{1}{P} \int_0^P |x(t) - s_M(t)|^2 \, dt$$

$$= \frac{1}{P} \int_0^P | \sum_{m=-\infty}^{\infty} \langle x, e_m \rangle e_m(t) - s_M(t)|^2 \, dt$$

$$= \frac{1}{P} \int_0^P | \sum_{|m|>M} \langle x, e_m \rangle e_m(t)|^2 \, dt = \sum_{|m|>M} |\langle x, e_m \rangle|^2,$$

which converges to 0, as $M \to \infty$, because we assumed that the power of the signal is finite:

$$\mathbf{PW}_x = \sum_{m=-\infty}^{\infty} |\langle x, e_m \rangle|^2 < \infty.$$

Note that the unspoken assumption here is that the orthonormal system $e_n(t), n = 0, \pm 1, \pm 2, \ldots,$ is rich enough to make the Fourier representation possible for any finite power signal. This assumption, often called *completeness* of the above orthonormal system, can actually be rigorously proven (see the footnote and other sources cited in the *Bibliographical Comments* at the end of this volume).

Approximation at Each Time Instant t Separately This type of approximation is often called the *point-wise approximation* and the goal is to verify that, for each time instant t,

$$\lim_{M \to \infty} s_M(t) = x(t). \tag{2.2.2}$$

Here the situation is delicate, as examples at the end of this section will show, and the assumption that signal $x(t)$ has finite power is not sufficient to guarantee the point-wise approximation. Neither is a stronger assumption that the signal is continuous. However,

If the signal is continuous, except, possibly, at a finite number of points, and has a bounded continuous derivative, except, possibly, at a finite number of points, then the point-wise approximation (2.2.2) holds true at all points of continuity of the signal.

Uniform Approximation in Time t If one wants to control the error of approximation simultaneously (uniformly) for all times t, then more stringent assumptions on the signal are necessary. Namely, we have the following theorem[5]:

If the signal is continuous everywhere and has a bounded continuous derivative except at a finite number of points then

$$\max_{0\le t\le P} \left| x(t) - s_M(t) \right| \to 0 \quad \text{as} \quad M \to \infty. \tag{2.2.3}$$

Note that the above statements do not resolve the question of what happens with the finite Fourier sums at discontinuity points of a signal, like those encountered in the rectangular waveforms in Examples 2.1.1–2.1.3. It turns out that under the assumptions of the above-quoted theorems, the points of discontinuity of the signal $x(t)$ are necessarily jumps, that is the left and right limits

$$x(t_-) = \lim_{s\uparrow t} x(s), \quad \text{and} \quad x(t_+) = \lim_{s\downarrow t} x(s) \tag{2.2.4}$$

exist, and the finite Fourier sums $s_M(x)$ of $x(t)$ converge, as $M \to \infty$, to the average value of the signal at the jump:

$$\lim_{M\to\infty} s_M(t) = \frac{x(t_-) + x(t_+)}{2}. \tag{2.2.5}$$

Example 2.2.1 (Approximation of a Rectangular Signal by Finite Fourier Sums) For the signal $x(t)$ in Example 2.1.1, the first three nonzero terms of its cosine expansion were

$$x(t) = \frac{a}{2} + \frac{a}{\pi}\left(2\cos(2\pi\frac{t}{P}) - \frac{2}{3}\cos(2\pi\frac{3t}{P}) + \dots\right).$$

Hence, in the case of period $P = 1$ and amplitude $a = 1$, the first four approximating sums are as follows:

$$s_0(t) = \frac{1}{2}, \qquad s_1(t) = \frac{1}{2} + \frac{2}{\pi}\cos 2\pi t,$$

$$s_2(t) = \frac{1}{2} + \frac{2}{\pi}\cos 2\pi t, \qquad s_3(t) = \frac{1}{2} + \frac{2}{\pi}\cos 2\pi t - \frac{2}{3\pi}\cos 6\pi t.$$

[5]Proofs of these two mathematical theorems and other results quoted in this section can be found in, e.g., T.W. Körner, *Fourier Analysis*, Cambridge University Press, 1988.

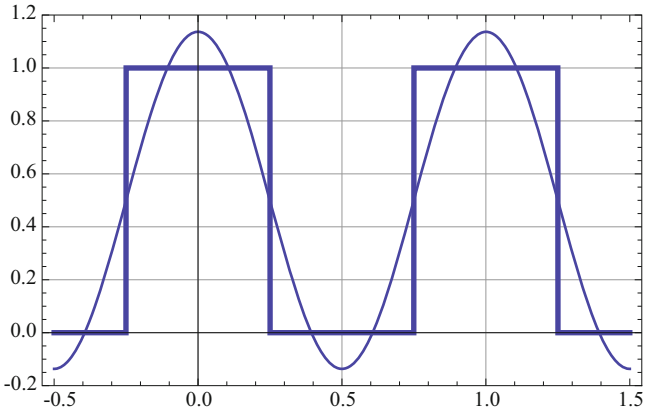


Fig. 2.4 Graph of the Fourier sum $s_1(t)$ for the rectangular waveform signal $x(t)$ from Example 2.1.1, plotted against the original signal $x(t)$

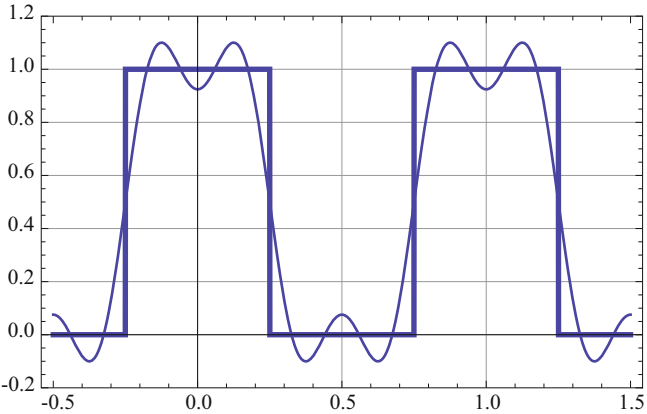


Fig. 2.5 Graph of the Fourier sum $s_3(t)$ for the rectangular waveform signal $x(t)$ from Example 2.1.1, plotted against the original signal $x(t)$. Note the behavior of the Fourier sum $s_3(t)$ at the signal's discontinuities, where it matches the average value of the signal at both sides of the jump, reflecting the asymptotics of formula (2.2.5)

The graphs of $s_1(t)$ and $s_3(t)$ are compared with the original signal $x(t)$ in Figs. 2.4 and 2.5. Note the behavior of the Fourier sums at the signal's discontinuities where the Fourier sums converge to the average value of the signal on both sides of the jump according to formula (2.2.5).

Remark 2.2.1 (Irregular Behavior of Fourier Sums) A word of warning is appropriate here. Abandoning the assumptions in the above two theorems leads very quickly to difficulties with approximation of the signal by its Fourier series. For example, there are continuous signals which, at some time instants, have finite Fourier sums

diverging to infinity. However, even for them, one can guarantee that the averages of consecutive Fourier sums converge to the signal for each t:

$$\frac{s_0(t) + s_1(t) + \ldots + s_M(t)}{M+1} \to x(t), \quad \text{as} \quad M \to \infty.$$

The expression on the left-hand side of the above formula is called the M-th *Césaro average* of the Fourier series. If one only assumes that the signal $x(t)$ is integrable, that is $\int_0^P |x(t)|\, dt < \infty$, which is the minimum assumption assuring that the Fourier coefficients $z_m = \langle x, e_m \rangle$ make sense, then one can find signals whose Fourier sums diverge to infinity, for all time instants t.

The Gibbs Phenomenon Another observation is that the finite Fourier sums of a signal satisfying the assumptions of the above quoted statements, despite being convergent to the signal, may have shapes that are very unlike the signal itself.

Example 2.2.2 (Behavior of Fourier Sums at Signal's Discontinuities) Consider the signal $x(t)$, with period $P = 1$, defined by the formula

$$x(t) = t, \quad \text{for} \quad -1/2 \le t < 1/2.$$

Clearly it is an odd signal, so $z_0 = 0$. For $m \neq 0$, integrating by parts,

$$z_m = \int_{-1/2}^{1/2} t e^{-j2\pi mt}\, dt = t \frac{-1}{j2\pi m} e^{-j2\pi mt} \Big|_{-1/2}^{1/2} - \frac{-1}{j2\pi m} \int_{-1/2}^{1/2} e^{-j2\pi mt}\, dt$$

$$= \frac{-1}{j2\pi m}(-1)^m$$

because the last integral is zero. The complex Fourier expansion of $x(t)$ is

$$x(t) = \sum_{m=-\infty,\, m\neq 0}^{\infty} \frac{-1}{j2\pi m}(-1)^m e^{j2\pi mt},$$

which yields a pure sine real Fourier expansion

$$x(t) = \sum_{m=1}^{\infty} \left(\frac{-1}{j2\pi m}(-1)^m e^{j2\pi mt} + \frac{-1}{j2\pi(-m)}(-1)^{-m} e^{j2\pi(-m)t} \right)$$

$$= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{\pi m} \sin(2\pi mt).$$

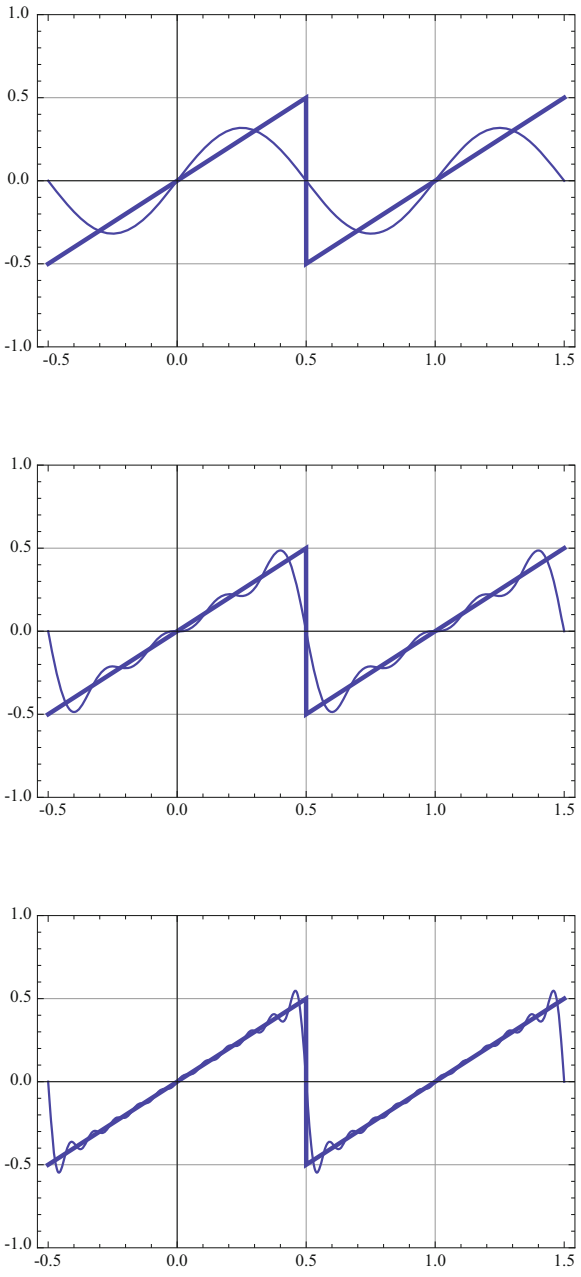


Fig. 2.6 Approximation of the periodic signal $x(t)$ from Example 2.2.2 by Fourier sums $s_1(t)$, $s_4(t)$, and $s_{10}(t)$ (top to bottom). Visible is the Gibbs phenomenon demonstrating that the shape of the Fourier sum near a point of discontinuity of the signal does not necessarily resemble the shape of the signal itself

Figure 2.6 shows approximation of the periodic signal $x(t)$ from Example 2.2.2 by Fourier sums $s_1(t)$, $s_4(t)$, and $s_{10}(t)$. Visible is the so-called *Gibbs phenomenon* demonstrating that the shape of the Fourier sum near a point of discontinuity of the signal does not necessarily resemble the shape of the signal itself. Yet, as the order

M of the approximation increases, the oscillations move closer to the jump so that the mean-square convergence of finite Fourier sums to the signal $x(t)$ still obtains.

2.3 Aperiodic Signals and Fourier Transforms

Periodic Signals with Increasing Period: From Fourier Series to Fourier Transform Consider a signal $x_P(t)$ of period P and fundamental frequency $f_0 = 1/P$. We already know that such signals can be represented by its Fourier series

$$x_P(t) = \sum_{m=-\infty}^{\infty} \left[\frac{1}{P} \int_{-P/2}^{P/2} x(s) e^{-j2\pi m f_0 s}\, ds \right] \cdot e^{j2\pi m f_0 t}. \tag{2.3.1}$$

Notice that, for the purposes of this section, we have written the formula for the Fourier coefficients of $x_P(t)$ as an integral over a symmetric interval $(-P/2, P/2]$ rather than the usual interval of periodicity $(0, P]$. Since both the signal $x_P(t)$ and complex exponentials

$$\exp(-j2\pi m f_0 s) = \cos(2\pi m f_0 s) + j \sin(2\pi m f_0 s)$$

are periodic with period P, any interval of length P will do.

Instead of considering aperiodic signals right off the bat we will make a gradual transition from analysis of periodic to aperiodic signals by considering what happens with the Fourier series if in the above representation (2.3.1) period P increases to ∞; the limit case of infinite period $P = \infty$ would then correspond to the case of an aperiodic signal.

To see the limit behavior of the Fourier series (2.3.1) we shall introduce the following notation:

1. The multiplicities of the fundamental frequency will become a running discrete variable f_m:

$$f_m = m \cdot f_0;$$

2. The increments of the new running variable will be denoted by

$$\Delta f_m = f_m - f_{m-1} = f_0 = \frac{1}{P}.$$

In this notation the Fourier expansion (2.3.1) can be rewritten in the form

$$x_P(t) = \sum_{m=-\infty}^{\infty} \left[\int_{-P/2}^{P/2} x(s) e^{-j2\pi f_m s}\, ds \right] e^{j2\pi f_m t}\, \Delta f_m \tag{2.3.2}$$

because $\Delta f_m = f_0 = 1/P$. Now, if the period $P \to \infty$, which is the same as assuming that the fundamental frequency $f_0 = \Delta f_m \to 0$, the sum on the right-hand side of the formula (2.3.2) converges to the integral so that our Fourier representation (2.3.2) of a periodic signal $x_P(t)$ becomes the following integral identity for the aperiodic signal:

$$x_\infty(t) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x_\infty(s) e^{-j2\pi f s}\, ds \right] e^{j2\pi f t} df. \tag{2.3.3}$$

The inner transformation,

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t}\, dt, \tag{2.3.4}$$

is called the *Fourier transform* of signal $x(t)$, and the outer transform,

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi f t} df, \tag{2.3.5}$$

is called the *inverse Fourier transform* of (complex in general) function $X(f)$. The variable in the Fourier transform is the frequency f.

Note that since $|e^{-j2\pi f t}| = 1$, the necessary condition for the existence of the Fourier transform in the usual sense is the absolute integrability of the signal:

$$\int_{-\infty}^{\infty} |x(t)|\, dt < \infty. \tag{2.3.6}$$

Later on we will try to extend its definition to some important nonintegrable signals.

Example 2.3.1 (Fourier Transform of a Double Exponential Signal) Let us trace the above limit procedure in the case of an aperiodic signal $x_\infty(t) = e^{-|t|}$. If this signal is approximated by periodic signals with period P obtained by truncating $x(t)$ to the interval $[-P/2, P/2)$ and extending it periodically, i.e.,

$$x_P(t) = e^{-|t|}, \quad \text{for} \quad -P/2 \le t < P/2,$$

then the Fourier coefficients of the latter are, remembering that $P = 1/f_0$,

$$z_{m,P} = \frac{1}{P} \int_{-P/2}^{P/2} e^{-|t|} e^{-j2\pi m t/P}\, dt$$

$$= \frac{2f_0}{1 + (2\pi m f_0)^2} \left(1 - e^{-1/(2f_0)} \Big(\cos(2\pi m f_0) + 2\pi m f_0 \sin(2\pi m f_0) \Big) \right).$$

Since the original periodic signal $x_P(t)$ was even, the Fourier coefficients $z_m = z_{-m}$ so that the discrete spectrum of $x_P(t)$ is symmetric. Now, as $P \to \infty$, that is $f_0 = 1/P \to 0$, the exponential term $e^{-1/(2f_0)} \to 0$, and with $f_0 = \Delta f$, $m f_0 = f$, we get that

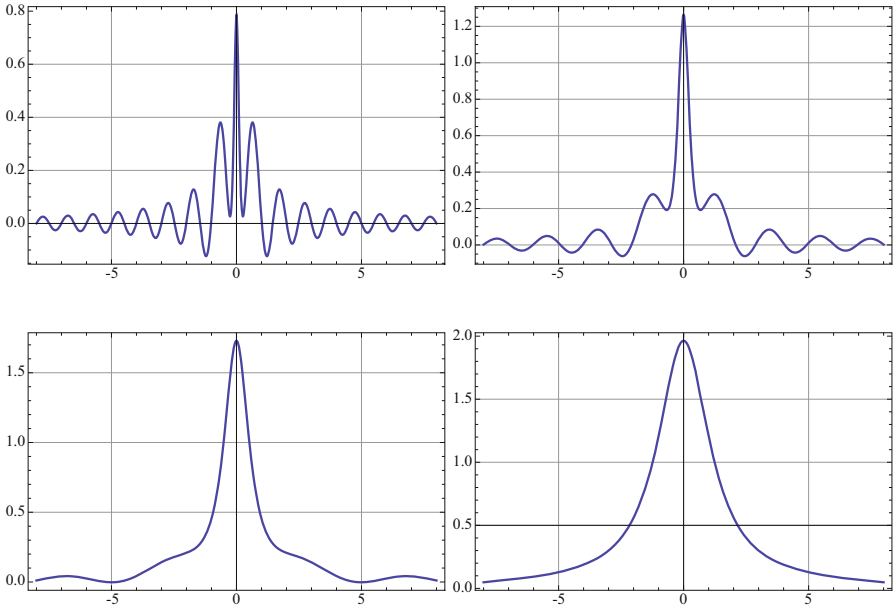


Fig. 2.7 Adjusted Fourier coefficients $Z_{m,P} \cdot P$ of a truncated, end periodically extended double exponential signal from Example 2.3.1 (shown above, for graphical convenience, as functions of continuous parameter m) approach the Fourier transform $X_\infty(f)$ of the aperiodic signal $x_\infty(t) = e^{-|t|}$. The values of P, from top left to bottom right, are 1, 2, 4, 8

$$z_{m,P} \to \frac{2}{1+(2\pi f)^2}\, df, \qquad \text{as} \qquad P \to \infty.$$

Thus, the Fourier transform of the aperiodic signal $x_\infty(t)$ is

$$X_\infty(f) = \frac{2}{1+(2\pi f)^2}.$$

Taking the inverse Fourier transform we verify[6] that

$$\int_{-\infty}^{\infty} \frac{2}{1+(2\pi f)^2} e^{j2\pi f t}\, df = e^{-|t|}.$$

Figure 2.7 illustrates the convergence, as period P increases, of Fourier coefficients $z_{m,P}$ to the Fourier transform $X_\infty(f)$.

[6]When faced with integrals of this sort the reader is advised to consult a book of integrals, or a computer package such as *Mathematica* or *Maple*.

2.4 Basic Properties of Fourier Transform

The property that makes the Fourier transform of signals so useful is its *linearity*, that is, the Fourier transform of a linear composition $\alpha x(t) + \beta y(t)$ of signals $x(t)$ and $y(t)$ is the same linear composition $\alpha X(f) + \beta Y(f)$ of their Fourier transforms. To facilitate notation we will often denote the fact that $X(f)$ is the Fourier transform of signal $x(t)$ by writing $x(t) \mapsto X(f)$ (read, $x(t)$ maps into $X(f)$). So,

$$\alpha x(t) + \beta y(t) \longmapsto \alpha X(f) + \beta Y(f). \tag{2.4.1}$$

The proof is instantaneous using linearity of the integral.

The familiar Parseval formula for periodic signals carries over in the form

$$\mathbf{EN}_x = \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |X(f)|^2 \, df. \tag{2.4.2}$$

It shows how the total energy of the signal is distributed of the *continuous* range of frequencies from minus to plus infinity. The nonnegative function $|X(f)|^2$ is called the *energy spectrum* of the aperiodic signal $x(t)$. The energy of the signal contained in the frequency band $[f_1, f_2]$ can then be calculated as the integral of the square of the modulus of its Fourier transform over that frequency interval:

$$\mathbf{EN}_x[f_1, f_2] = \int_{f_1}^{f_2} |X(f)|^2 \, df. \tag{2.4.2a}$$

An observant reader will see immediately that integrability of the signal necessary to define the Fourier transform is not sufficient for the validity of the Parseval formula (2.4.2) as the finiteness of the integral $\int_{-\infty}^{\infty} |x(t)| \, dt$ does not imply that the signal has finite energy $\mathbf{EN}_x$ (and, vice versa, finiteness of $\mathbf{EN}_x$ does not imply the absolute integrability of the signal, see Problem 2.7.11).

Parseval's formula also has the following useful extension

$$\int_{-\infty}^{\infty} x(t) \cdot y(t) \, dt = \int_{-\infty}^{\infty} X(f) \cdot Y^*(f) \, df. \tag{2.4.3}$$

In the context of transmission of signals through linear systems the critical property of the Fourier transform is that the *convolution* $[x * y](t)$ of signals $x(t)$ and $y(t)$,

$$[x * y](t) = \int_{-\infty}^{\infty} x(s) y(t - s) \, ds \tag{2.4.4}$$

a fairly complex, nonlocal operation, has the Fourier transform that is simply the point-wise product of the corresponding Fourier transforms

$$[x * y](t) \longmapsto X(f) \cdot Y(f). \tag{2.4.5}$$

Indeed,

$$\int_{-\infty}^{\infty} [x * y](t) e^{-j2\pi ft}\, dt = \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} x(s) y(t-s)\, ds \Big] e^{-j2\pi ft} dt =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y(t-s) e^{-j2\pi f(t-s)} x(s) e^{-j2\pi fs}\, ds\, dt$$

$$= \int_{-\infty}^{\infty} y(u) e^{-j2\pi fu}\, du \cdot \int_{-\infty}^{\infty} x(s) e^{-j2\pi fs}\, ds = X(f) \cdot Y(f),$$

where the penultimate equality resulted from the substitution $t - s = u$.

Since many electrical circuits are described by linear differential equations the behavior of the Fourier transform under differentiation of the signal is another important issue. Here the calculation is also direct:

$$\int_{-\infty}^{\infty} x'(t) e^{-j2\pi ft} dt = x(t) e^{-j2\pi ft}|_{-\infty}^{\infty} + j2\pi f \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt =$$

$$= 0 + j2\pi f X_Z(f).$$

The first term is 0 because the signal's absolute integrability (remember, we have to assume it to guarantee the existence of the Fourier transform) implies that $x(\infty) = x(-\infty) = 0$. Thus we have a rule

$$x'(t) \longmapsto (j2\pi f) \cdot X(f). \tag{2.4.6}$$

Similarly, one can employ the Fourier Transform technique to study linear partial differential equations which describe temporal evolution of physical phenomena in continuous media, see Problem 2.7.18.

The above and other, simple-to-derive operational rules for Fourier transforms are summarized in Table 2.2.

Example 2.4.1 (Deterministic Gaussian Signal and Its Fourier Transform Have the Same Functional Shape) Consider the curious example of a signal of the form $x(t) = e^{-\pi t^2}$ which has the familiar bell shape. Its Fourier transform is

$$X(f) = \int_{-\infty}^{\infty} e^{-\pi t^2 - j2\pi ft}\, dt = \int_{-\infty}^{\infty} e^{-\pi(t+jf)^2} e^{-\pi f^2}\, dt = e^{-\pi f^2},$$

Table 2.2 Properties of the Fourier Transform

Signal		Fourier transform
	Linearity	
$\alpha x(t) + \beta y(t)$	$\longmapsto$	$\alpha X(f) + \beta Y(f)$
	Convolution	
$[x * y](t)$	$\longmapsto$	$X(f) \cdot Y(f)$
	Differentiation	
$x^{(n)}(t)$	$\longmapsto$	$(j2\pi f)^n X(f)$
	Time reversal	
$x(-t)$	$\longmapsto$	$X(-f)$
	Time delay	
$x(t - t_0)$	$\longmapsto$	$X(f) \cdot e^{-j2\pi t_0 f}$
	Frequency translation	
$x(t) \cdot e^{j2\pi f_0 t}$	$\longmapsto$	$X(f - f_0)$
	Frequency differentiation	
$(-j)^n t^n x(t)$	$\longmapsto$	$(2\pi)^{-1} X^{(n)}(f)$
	Frequency convolution	
$x(t)y(t)$	$\longmapsto$	$[X * Y](f)$

because

$$\int_{-\infty}^{\infty} e^{-\pi(t+jf)^2}\, dt = \int_{-\infty}^{\infty} e^{-\pi t^2}\, dt = 1.$$

Indeed, changing to polar coordinates r, θ, we can evaluate easily that

$$\Big(\int_{-\infty}^{\infty} e^{-\pi t^2}\, dt\Big)^2 = \int_{-\infty}^{\infty} e^{-\pi t^2}\, dt \cdot \int_{-\infty}^{\infty} e^{-\pi s^2}\, ds$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\pi(t^2+s^2)}\, dt\, ds = \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\pi r^2} r\, dr = 1.$$

Thus signal $x(t) = e^{-\pi t^2}$ has the remarkable property of having the Fourier transform of exactly the same functional shape. This fact has profound consequences in Fourier analysis, mathematical physics, quantum mechanics, and the theory of partial differential equations.

2.5 Fourier Transforms of Some Non-integrable Signals: Dirac-Delta Impulse

There exist important nonintegrable signals, such as $x(t) = const$, or $x(t) = \cos t$ that are not absolutely integrable over the whole time-line and, as a result, their Fourier Transforms are not well defined in the context of the classical calculus. Nevertheless, to cover these and other important cases, it is possible to extend the standard calculus by introduction of the so-called *Dirac-delta* "function" $\delta(f)$ which, loosely speaking, is an infinitely high but infinitely narrow spike located at $f = 0$ which, very importantly, has the "area," that is the "integral," equal to 1. Of course, one can similarly introduce the time-domain Dirac-delta $\delta(t)$ in which case it is often called the *Dirac-delta impulse*.

Heuristically (but one can also make this approach rigorous) the best way to think about the Dirac-delta is as a limit,

$$\delta(f) = \lim_{\epsilon \to 0} r_\epsilon(f), \tag{2.5.1}$$

where

$$r_\epsilon(f) = \begin{cases} 1/(2\epsilon), & \text{for } -\epsilon \le f \le +\epsilon; \\ 0, & \text{elsewhere,} \end{cases}$$

is a family of rectangular functions of width 2ϵ, which have the area 1 underneath; see Fig. 2.8.

Obviously, the choice of the rectangular functions is not unique here. Any sequence of nonnegative functions which integrate to 1 over the whole real line and converge to zero pointwise at every point different from the origin would do.

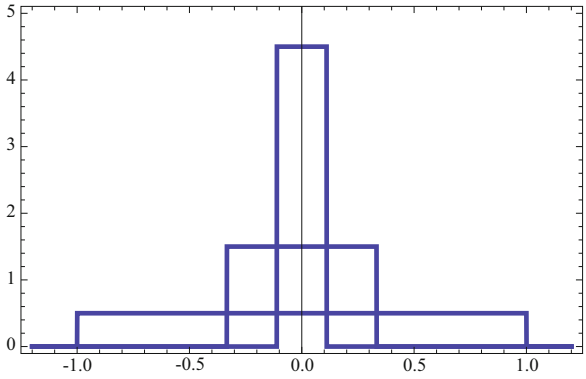


Fig. 2.8 Approximation of the Dirac-delta $\delta(f)$ by rectangular functions $r_\epsilon(f)$ for $\epsilon = 1, 1/3$, and 1/9

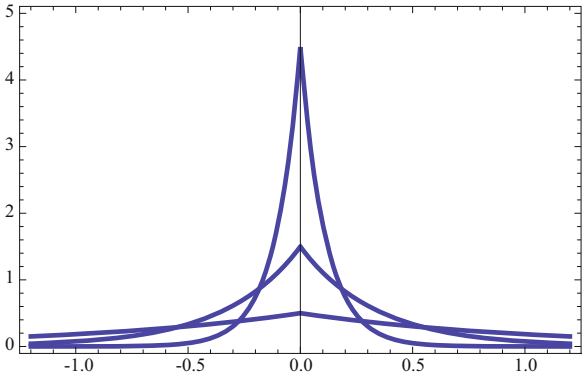


Fig. 2.9 Approximation of the Dirac delta $\delta(f)$ by two-sided exponential functions $(1/(2a))\exp(-|f|/a)$ for $a = 1, 1/3$, and $1/9$

For example, as approximants to the Dirac-delta we can also take the family of double-sided exponential functions of variable x,

$$\frac{1}{2a}\exp\left(\frac{|f|}{a}\right),$$

indexed by parameter $a \to 0+$. Three functions of this family, for parameter values $a = 1,\ 1/3,\ 1/9$, are pictured in Fig. 2.9.

More formally, the Dirac-delta will be defined here as a "functional" characterized by its "probing property" describing its scalar products with other, regular functions:

$$\langle \delta, X \rangle := \int_{-\infty}^{\infty} \delta(f) X(f)\, df = X(0). \tag{2.5.2}$$

In other words, integration of a function $X(f)$ against the Dirac-delta produces the value of that function at $f = 0$. This property permits us to use the Dirac-delta operationally whenever it appears inside integrals.

The "probing" formula (2.5.2) can be justified remembering our intuitive definition (2.5.1). Indeed, if function $X(f)$ is regular enough, then

$$\int_{-\infty}^{\infty} \delta(f) X(f)\, df = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} r_\epsilon(f) X(f)\, df$$

$$= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} X(f)\, df = X(0)$$

in view of the fundamental theorem of calculus.

Other properties of the Direc-delta follow immediately, For the Dirac-delta shifted to $f = f_0$,

$$\int_{-\infty}^{\infty} \delta(f - f_0) X(f)\, df = X(f_0). \tag{2.5.3}$$

Also,

$$\int_{-\epsilon}^{\epsilon} \delta(f)\, df = 1, \tag{2.5.4}$$

and

$$\int_{-\infty}^{\infty} \delta(f) X(f) df = 0, \quad \text{if} \quad X(0) = 0, \tag{2.5.5}$$

The last property is often intuitively stated as follows:

$$\delta(f) = 0, \quad \text{for} \quad f \neq 0. \tag{2.5.6}$$

Equipped with the Dirac-delta technique we can immediately obtain the Fourier transform of some nonintegrable signals.

Example 2.5.1 (Fourier Transforms of Complex Exponentials) Finding the Fourier transform of the harmonic oscillation signal $x(t) = e^{j2\pi f_0 t}$ is impossible by direct integration as

$$\int_{-\infty}^{\infty} e^{j2\pi f_0 t} e^{-j2\pi f t} dt = \frac{1}{j2\pi(f_0 - f)} \Big(\cos 2\pi(f_0 - f)t + j \sin 2\pi(f_0 - f)t\Big)\Big|_{t=-\infty}^{\infty},$$

and the limits

$$\lim_{t \to \pm\infty} \cos 2\pi(f_0 - f)t, \quad \text{and} \quad \lim_{t \to \pm\infty} \sin 2\pi(f_0 - f)t,$$

do not exist. But one immediately notices that, in view of (2.5.2), the inverse transform of the shifted Dirac-delta is,

$$\int_{-\infty}^{\infty} \delta(f - f_0) e^{j2\pi f t} df = e^{j2\pi f_0 t}.$$

Thus the Fourier transform of $x(t) = e^{j2\pi f_0 t}$ is $\delta(f - f_0)$. In particular, the Fourier transform of a constant 1 is $\delta(f)$ itself.

Example 2.5.2 (Fourier Transforms of Real Harmonic Oscillations) The Fourier transform of the signal $x(t) = \cos 2\pi t$ has to be found in a similar fashion as direct integration of $\int_{-\infty}^{\infty} \cos(2\pi t)\, e^{-j2\pi f t} dt$ is again impossible. But one observes that the inverse Fourier transform

$$\int_{-\infty}^{\infty} \frac{1}{2}\Big(\delta(f - 1) + \delta(f + 1)\Big) e^{j2\pi f t}\, df = \frac{e^{j2\pi t} + e^{-j2\pi t}}{2} = \cos 2\pi t,$$

so the Fourier transform of $\cos 2\pi t$ is $(\delta(f - 1) + \delta(f + 1))/2$.

Table 2.3 Common Fourier transforms

Signal		Fourier transform
$e^{-a\|t\|}$	$\longmapsto$	$\dfrac{2a}{a^2+(2\pi f)^2}, \quad a>0$
$e^{-\pi t^2}$	$\longmapsto$	$e^{-\pi f^2}$
$\begin{cases} 1, & \text{for} \|t\| \le 1/2; \\ 0, & \text{for} \|t\| > 1/2. \end{cases}$	$\longmapsto$	$\dfrac{\sin \pi f}{\pi f}$
$\begin{cases} 1-\|t\|, & \text{for} \|t\| \le 1; \\ 0, & \text{for} t\| > 1. \end{cases}$	$\longmapsto$	$\dfrac{\sin^2 \pi f}{\pi^2 f^2}$
$e^{j2\pi f_0 t}$	$\longmapsto$	$\delta(f-f_0)$
$\delta(t)$	$\longmapsto$	1
$\cos 2\pi f_0 t$	$\longmapsto$	$\dfrac{\delta(f+f_0)+\delta(f-f_0)}{2}$
$\sin 2\pi f_0 t$	$\longmapsto$	$j\dfrac{\delta(f+f_0)-\delta(f-f_0)}{2}$
$e^{-at}\cdot u(t)$	$\longmapsto$	$\dfrac{1}{a+j2\pi f}, \quad a>0$

Table 2.3 lists Fourier transforms of some common signals. Here, and thereafter, $u(t)$ denotes Heaviside's Unit Step Function equal to 0, for $t<0$, and 1, for $t \ge 0$.

Calculus of Dirac-Delta "Functions": Theory of Schwartzian Distributions There exists a large theory of Dirac-delta "functions," and of similar mathematical objects called distributions (in the sense of Schwartz),[7] which develops tools that help carry out operations such as distributional differentiation. To give the reader a little taste of it let us start here with the classical integration-by-parts formula which, for usual, vanishing at $f=\pm\infty$ functions $X(f)$, and $Y(f)$, states that

$$\langle X, Y'\rangle = \int_{-\infty}^{\infty} X(f)\cdot Y'(f)\,df = -\int_{-\infty}^{\infty} X'(f)\cdot Y(f)\,df = -\langle X', Y\rangle. \tag{2.5.7}$$

This identity, applied formally, can be used as the *definition* of the derivative $\delta'(f)$ of the Dirac-delta by assigning to it the following probing property:

$$\langle X, \delta'\rangle = \int_{-\infty}^{\infty} X(f)\cdot \delta'(f)\,df = -\int_{-\infty}^{\infty} X'(f)\cdot \delta(f)\,df = -X'(0). \tag{2.5.8}$$

Symbolically, we can write

$$X(f)\cdot\delta'(f) = -X'(f)\cdot\delta(f).$$

[7] For a more complete exposition of the theory and applications of the Dirac delta and related "distributions," see A.I. Saichev and W.A. Woyczyński, *Distributions in the Physical and Engineering Sciences, Volume 1: Distributional Calculus, Integral Transforms and Wavelets*, Birkhäuser-Boston, 1998. Also, see Bibliographical Comments at the end of this volume.

In the particular case $X(f) = f$ (here, the function has to be thought of as a limit of functions vanishing at $\pm\infty$) we get

$$f \cdot \delta'(f) = -\delta(f),$$

a useful computational formula.

2.6 Discrete and Fast Fourier Transforms

In practice, for many signals, we only sample the value of the signal at discrete times, although in reality the signal continues between these sampling times. In such cases we can approximate the integrals involved in calculation of the Fourier transforms in the same way as one does in numerical integration in calculus, using left-handed rectangles, trapezoids, Simpson's rule, etc. We use the simplest approximation, which is equivalent to assuming that the signal is constant between the sampling times (and rectangles' areas approximate the area under the function).

So, suppose that the sampling period is T_s, with the sampling frequency $f_s = 1/T_s$, so that the signal's sample is given in the form of a finite sequence,

$$x_k = x(kT_s),\ k = 0, 1, 2, \ldots, N-1, \tag{2.6.1}$$

so that we can interpret it as a periodic signal with period

$$P = \frac{1}{f_0} = NT_s = \frac{N}{f_s}. \tag{2.6.2}$$

The integral in formula (2.3.1) approximating the Fourier transform of the signal $x(t)$ at discrete frequencies $mf_0, m = 0, 1, 2, \ldots, N-1$, can now be, in turn, approximated by the sum:

$$X_m = X(mf_0) = \frac{1}{P}\sum_{k=0}^{N-1} x(kT_s)e^{-j2\pi mf_0 kT_s} \cdot T_s$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} x_k e^{-j2\pi mk/N}, \tag{2.6.3}$$

in view of relationships (2.6.2). The sequence,

$$X_m,\ m = 0, 1, 2, \ldots, N-1, \tag{2.6.4}$$

is traditionally called the *Discrete Fourier Transform (DFT)* of the signal sample $x_k,\ k = 0, 1, 2, \ldots, N-1$, described in (2.6.1).

Note that the calculation of the DFT via formula (2.6.3) calls for N^2 multiplications,

$$x_k \cdot e^{-j2\pi mk/N},\ m, k = 0, 1, 2, \ldots, N-1.$$

One often says that the formula's *computational (algorithmic) complexity* is of the order N^2. This computational complexity, however, can be dramatically reduced by cleverly grouping terms in the sum (2.6.3). The technique, which usually is called the *Fast Fourier Transform (FFT)*, was known to Carl Friedrich Gauss at the beginning of the nineteenth century, but was rediscovered and popularized by Cooley and Tukey in 1965.[8] We will explain it in the special case when the signal's sample size N is a power of 2. So assume that $N = 2^n$, and let $\omega_N = e^{-j2\pi/N}$. The complex number ω_N is called a complex N-th root of unity because $\omega_N^N = 1$. Obviously, for $M = N/2$, we have

$$\omega_{2M}^{(2k)m} = \omega_M^{km}, \quad \omega_M^{M+m} = \omega_M^m, \quad \text{and} \quad \omega_{2M}^{M+m} = -\omega_{2M}^m. \tag{2.6.5}$$

The crucial observation is to recognize that the sum (2.6.3) can be split into two pieces

$$X_m = \frac{1}{2}\Big(X_m^{\text{even}} + X_m^{\text{odd}} \cdot \omega_{2M}^m\Big), \tag{2.6.6}$$

where

$$X_m^{\text{even}} = \frac{1}{M}\sum_{k=0}^{M-1} x_{2k}\omega_M^{km}, \quad \text{and} \quad X_m^{\text{odd}} = \frac{1}{M}\sum_{k=0}^{M-1} x_{2k+1}\omega_M^{km}, \tag{2.6.7}$$

and that, in view of (2.6.5),

$$X_{m+M} = \frac{1}{2}\Big(X_m^{\text{even}} - X_m^{\text{odd}} \cdot \omega_{2M}^m\Big). \tag{2.6.8}$$

As a result, only values $X_m, m = 0, 1, 2, \ldots, M-1 = N/2-1$, have to be calculated by computationally laborious multiplications. The values $X_m, m = M, M+1, \ldots, 2M-1 = N-1$, are simply obtained by formula (2.6.8). The above trick is then repeated at levels $N/2^2, N/2^3, \ldots, 2$. If we denote by $CC(n)$ the *computations complexity* of the above scheme, that is the number of multiplications required, we see that

$$CC(n) = 2CC(n-1) + 2^{n-1},$$

[8] Cooley, J.W. and Tukey, O.W. "An Algorithm for the Machine Calculation of Complex Fourier Series." Math. Comput. 19, 297–301, 1965.

with the first term on the right being the result of halving the size of the sample at each step, and the second term resulting from multiplications of X_m^{odd} by ω_{2M}^m in (2.6.6) and (2.6.8). Iterating the above recursive relation one obtains that

$$CC(n) = 2^{n-1}\log_2 2^n = \frac{1}{2}N\log_2 N, \tag{2.6.9}$$

a major improvement over the N^2-order of the computational complexity of the straightforward calculation of DFT.

2.7 Problems and Exercises

1* Prove that the system of real harmonic oscillations

$$\sin(2\pi m f_0 t), \quad \cos(2\pi m f_0 t), \quad m = 1, 2, \ldots$$

forms an orthogonal system. Is the system normalized? Is the system complete? Use the above information to derive formulas for coefficients in the Fourier expansions in terms of sines and cosines. Model this derivation on calculations in Sect. 2.1.

2* Using the results from Problem 1 find formulas for amplitudes c_m and phases θ_m in the expansion of a periodic signal $x(t)$ in terms of only cosines, $x(t) = \sum_{m=0}^{\infty} c_m \cos(2\pi m f_0 t + \theta_m)$.

3 Find a general formula for the coefficients in the cosine Fourier expansion for the even rectangular waveform $x(t)$ from Example 2.1.1.

4 Find a general formula for the coefficients b_m in the sine Fourier expansion for the odd rectangular waveform $x(t)$ from Example 2.1.2.

5 Carry out calculations of Example 2.1.3 in the case of arbitrary period P and amplitude a.

6 Find three consecutive approximations by finite Fourier sums of the signal $x(t)$ from Example 2.1.3. Graph them and compare the graphs with the graph of the original signal.

7 Find the complex and real Fourier series for the periodic signal with period P defined by the formula

$$x(t) = \begin{cases} a, & \text{for} 0 \le t < P/2; \\ -a, & \text{for} P/2 \le t < P. \end{cases}$$

In the case $P = \pi$ and $a = 2.5$ produce graphs comparing the signal $x(t)$ and its finite Fourier sums of order 1, 3, and 6.

8 Find the complex and real Fourier series for the periodic signal with period $P = 1$ defined by the formula

$$x(t) = \begin{cases} 1 - t/2, & \text{for} 0 \le t < 1/2; \\ 0, & \text{for} 1/2 \le t < 1. \end{cases}$$

Produce graphs comparing the signal $x(t)$ and its finite Fourier sums of order 1,3, and 6.

9* Find the complex and real Fourier series for the periodic signal $x(t) = |\sin t|$. Produce graphs comparing the signal $x(t)$ and its finite Fourier sums of order 1, 3, and 6. In electrical engineering, signal $|\sin t|$ is produced by running the sine signal through a rectifier.

10 Find the complex and real Fourier series for the periodic signal with period $P = \pi$ defined by the formula

$$x(t) = e^t, \quad \text{for} \quad -\pi/2 < t \le \pi/2.$$

Produce graphs comparing the signal $x(t)$ and its finite Fourier sums of order 1,3, and 6.

11 Find an example of a signal $x(t)$ that is absolutely integrable, i.e., $\int_{-\infty}^{\infty} |x(t)|\,dt < \infty$ but has infinite energy $\mathbf{EN}_x = \int_{-\infty}^{\infty} |x(t)|^2\,dt$, and vice versa, find an example of a signal which has finite energy but is not absolutely integrable.

12 Provide a detailed verification of Fourier transform properties listed in Table 2.2. Provide a detailed verification of the Fourier transforms Table 2.3.

13*

(a) The nonperiodic signal $x(t)$ is defined as equal to 1/2 on the interval $[-1, +1]$, and 0 elsewhere. Plot it and calculate its Fourier transform $X(f)$. Plot the latter.
(b) The nonperiodic signal $y(t)$ is defined as equal to (t+2)/4 on the interval $[-2, 0]$, $(-t+2)/4$ on the interval $[0, 2]$, and 0 elsewhere. Plot it and calculate its Fourier transform $Y(f)$. Plot the latter.
(c) Compare the Fourier transforms $X(f)$ and $Y(f)$. What conclusion do you draw about the relationship of the original signals $x(t)$ and $y(t)$?

14 Find the Fourier transform of the periodic signal $x(t) = \sum_{m=-\infty}^{\infty} z_m e^{j2\pi m f_0 t}$.

15 Find the Fourier Transform of the solution $x(t)$ of the differential equation $x''(t) + x(t) = \cos t$.

16 Find the Fourier transform of the signals given below. Graph both the signal and its Fourier transform (real and imaginary parts separately, if necessary):

(a)

$$x(t) = \frac{1}{1+t^2}, \qquad -\infty < t < \infty,$$

(b)

$$e^{-t^2/2}, \qquad -\infty < t < \infty,$$

(c)

$$x(t) = \begin{cases} \sin t \cdot e^{-t}, & \text{for} t \geq 0; \\ 0, & \text{for} t < 0, \end{cases}$$

(d)

$$x(t) = \sin t \cdot e^{-|t|},$$

(e)

$$x(t) = y * z(t), \quad y(t) = u(t) - u(t-1), \quad z(t) = e^{-|t|},$$

where $u(t)$ is the unit step signal =0 for negative t and =1 for $t \geq 0$.

17 Find the convolution $(x * x)(t)$ if $x(t) = u(t) - u(t-1)$, where $u(t)$ is the Unit Step Function. First, use the original definition of the convolution and then verify your result using the Fourier transform method.

18* Utilize the Fourier transform (in the space variable z) to find a solution of the diffusion (heat) partial differential equation

$$\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial z^2},$$

for a function $u(t, z)$ satisfying the initial condition $u(0, z) = \delta(z)$. The solution of the above equation is often used to describe the temporal evolution of the density of a diffusing substance.[9]

[9]It was the search for solutions to this problem that induced Jean-Baptiste Fourier (born March 21, 1768, in Auxerre, France, died May 16, 1830, in Paris) to introduce in his treatise *Théorie analytique de la chaleur* (1822; *The Analytical Theory of Heat*) the tools of infinite functional series and integral transforms now known under the names of Fourier series and transforms. During the Napoleonic era Fourier was also known as an Egyptologist and administrator. The modern young author of research papers, impatient with delays in publication of his/her work, should find solace in the fact that appearance of Fourier's great memoir was held up by the referees for 15 years; it was first presented to the Institut de France on December 21, 1807.

19 Assuming the validity of the Parseval formula $\int_{-\infty}^{\infty} |x(t)|^2\,dt = \int_{-\infty}^{\infty} |X(f)|^2\,df$, prove its extended version $\int_{-\infty}^{\infty} x(t)\cdot y^*(t)\,dt = \int_{-\infty}^{\infty} X(f)\cdot Y^*(f)\,df$. *Hint:* In the case of real-valued $x(t)$, $y(t)$, $X(f)$, and $Y(f)$, it suffices to utilize the obvious identity $4xy = (x+y)^2 - (x-y)^2$, but in the general, complex case, first verify, and then apply the following *polarization identity*:

$$4xy^* = |x+y|^2 - |x-y|^2 + j(|x+jy|^2 - |x-jy|^2).$$

Remember that the modulus square $|z|^2 = zz^*$.

20 Consider the triangular signal, $x(t) = 1+t$, for $t \in [-1, 0]$, $= 1-t$, for $t \in [0, 1]$, and equal to zero elsewhere. Find its Fourier transform, and the Fourier transform of its second derivative $x''(t)$. Calculate the second derivative first.

Chapter 3
Uncertainty Principle and Wavelet Transforms

Abstract The method of *wavelet transforms*, which provides a decomposition of functions in terms of a fixed orthogonal family of functions of constant shape but varying scales and locations, recently acquired broad significance in the analysis of signals and of experimental data from various physical phenomena. Its value for the whole spectrum of problems in many areas of science and engineering, including the study of electromagnetic and turbulent hydrodynamic fields, image reconstruction algorithms, prediction of earthquakes and tsunami waves, and statistical analysis of economic data, is by now quite obvious.

Although the systematic ideas of wavelet transforms have been developed only since the early 1980s, to get the proper intuitions about sources of their effectiveness it is necessary to become familiar with a few more traditional ideas, tools, and methods. One of those is the celebrated *uncertainty principle* for the Fourier transforms which will be given special attention in this chapter. A close relative of the wavelet transform—the *windowed Fourier transform*—will also be studied in this context.

3.1 Time–Frequency Localization and the Uncertainty Principle

Consider a (perhaps complex-valued) signal $x(t)$ such that

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = 1. \tag{3.1.1}$$

The quantity $|x(t)|^2$ can be thought of as the signal's "energy" density and describes its distribution in time. If the signal $x(t)$ is square integrable but the

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_3

condition (3.1.1) is not satisfied, then one can always normalize it by considering $x(t)/(\int_{-\infty}^{\infty} |x(t)|^2 dt)^{1/2}$. In this context, the quantity

$$\int t|x(t)|^2 dt$$

can be interpreted as the location in time of the signal's "center of gravity," or its mean location. For the purposes of this section, and without loss of generality, we will assume that its mean location is at 0 or, in other words, that $\int t|x(t)|^2 dt = 0$. In this case, the quantity

$$\sigma^2[x] = \int_{-\infty}^{\infty} t^2 |x(t)|^2 dt \tag{3.1.2}$$

measures the average square deviation from the mean time location, or the degree of *localization* of the signal around its mean in the time domain.

On the other hand, the Fourier transform,

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t} dt,$$

displays no direct information about the signal's time localization, but has explicit information about its frequency localization. The square of its modulus $|X(f)|^2$ is the frequency domain counterpart of the time density $|x(t)|^2$. Note that, by Parseval's formula (2.4.2),

$$\int_{-\infty}^{\infty} |X(f)|^2 df = \int_{-\infty}^{\infty} |x(t)|^2 dt,$$

so that $|X(f)|^2$ can be viewed as the signal's normalized density in the frequency domain. Assume (again, without loss of generality) that the mean frequency

$$\int_{-\infty}^{\infty} f|X(f)|^2 df = 0.$$

Then the quantity

$$\sigma^2[X] = \int_{-\infty}^{\infty} f^2 X(f)|^2 df \tag{3.1.3}$$

measures the mean square deviation from the mean frequency location, or the degree of localization of the signal in the frequency domain.

The *uncertainty principle* asserts that there exists a lower bound on the simultaneous localization of the signal in time and frequency domains. More precisely, it states that

$$\sigma^2[x]\sigma^2[X] \geq 1/4, \tag{3.1.4}$$

whenever the variances $\sigma^2[x]$ and $\sigma^2[X]$ are well defined. Note the universal constant 1/4.

To see why the uncertainty principle holds true, consider the integral

$$I(a) = \int_{-\infty}^{\infty} |atx(t) + x'(t)|^2 dt \geq 0, \tag{3.1.5}$$

where a is a real parameter. Then, since

$$|atx(t) + x'(t)|^2 = (atx + x')(atx^* + (x')^*),$$

we get that

$$I(a) = a^2 \int_{-\infty}^{\infty} t^2|x|^2 dt + a \int_{-\infty}^{\infty} t(x(x')^* + x'x^*)\, dt + \int_{-\infty}^{\infty} |x'|^2\, dt. \tag{3.1.6}$$

The first integral in (3.1.6) is just $\sigma^2[x]$ (by definition (3.1.2)). The second integral is equal to

$$\int_{-\infty}^{\infty} t(xx^*)' dt = t|x(t)|^2\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} |x|^2 dt = -1,$$

since $t|x(t)|^2$ decays to zero at $\pm\infty$ in view of the assumption $\sigma^2[x] < \infty$. Finally, the third integral is equal to

$$\int_{-\infty}^{\infty} f^2|X(f)|^2 df = \sigma^2[X]$$

because of Parseval's formula and the fact that the Fourier transform of x' is equal to $jfX(f)$. As a result, the integral

$$I(a) = a^2\sigma^2(x) - a + \sigma^2(X). \tag{3.1.7}$$

This is a quadratic polynomial in variable a and, in view of (3.1.5), it is nonnegative for all values of x. As such, it has a nonpositive discriminant

$$1 - 4\sigma^2(x)\sigma^2(X) \leq 0,$$

which immediately yields the uncertainty principle (3.1.4).

Remark 3.1.1 (The Heisenberg Uncertainty Principle in Quantum Mechanics) The (3D version of the) above uncertainty principle concerning time-frequency localization has a celebrated interpretation in quantum mechanics, where the principle

asserts that the position and the momentum of a particle cannot be simultaneously measured with arbitrary accuracy. Indeed, in quantum mechanics the particle is represented by a complex wave function, and its modulus square is the probability density of its position in space. The observables are represented by operators on wave functions. The *position observable* is represented by a multiplication by variable (vector) and the *momentum observable* is represented by the operation of differentiation. However, via the Fourier transform, the latter also becomes an operation of multiplication but by an independent variable (vector) in the frequency domain. Thus the uncertainty principle (3.1.4) gives the universal lower bound for the product of variances of the probability distributions of the position and of the momentum. In the three-dimensional space, and in the physical units, the lower bound 1/4 in (3.1.4) has to be replaced by a different mathematical constant multiplied by a universal physical constant called the Planck constant.

Remark 3.1.2 One can check that the equality in the uncertainty principle (3.1.4) obtains only for the Gaussian function $x(t) = \pi^{-1/4} \exp(-t^2/2)$. Thus the optimal simultaneous time and frequency localization is attained for a Gaussian-shaped signal.

3.2 Windowed Fourier Transform

3.2.1 Forward Windowed Fourier Transform

The uncertainty principle discussed above is a basic law of mathematics and it is impossible to fool nature by measuring the frequency of the incoming signal with an arbitrary precision in a finite time interval. Moreover, for most of the signals we have to deal with in practical problems, such as speech, musical sounds, and radar signals, the situation is often much worse than the basic uncertainty inequality. Nevertheless, it is often possible to process these signals in such a way that, without violation of the uncertainty principle, one can obtain information about the signal's "current" frequency and its time evolution. These various practical signal processing methods are adapted to different kinds of signals and pursue different goals. In this section we will take a look at one of these methods called the *windowed Fourier transform* which is closest perhaps to the spirit of the usual Fourier transform.

In what follows, to better grasp the mechanisms behind the windowed Fourier transform, it will be instructive to test them on a sample signal that we will call the *simplest tune*. Mathematically, it is described by the real part of the complex function

$$x(t) = \exp(j\Phi(t)), \tag{3.2.1}$$

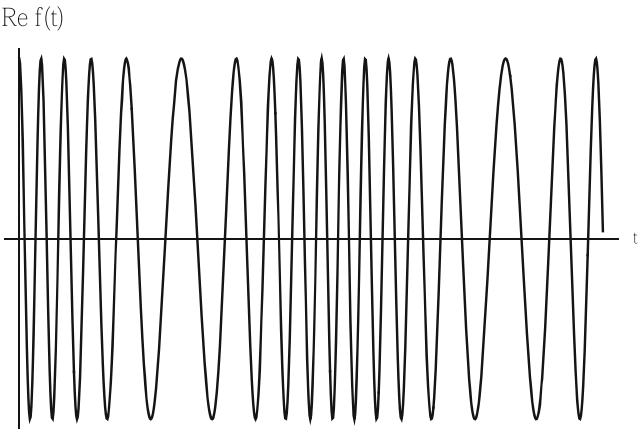


Fig. 3.1 Plot of the simplest tune in case of $f_0 = 10\nu$ and $\beta = \Omega/\nu = 5$

where

$$\Phi(t) = f_0 t + \frac{\Omega}{\nu} \sin(\nu t) \tag{3.2.2}$$

is the signal phase. The simplest tune is plotted in Fig. 3.1.

It is customary to say in the theoretical physics context that the simplest tune has the *instantaneous frequency* (admittedly, an oxymoron)

$$f_{inst}(t) = \frac{d\Phi(t)}{dt} = f_0 + \Omega \cos(\nu t), \tag{3.2.3}$$

which oscillates with period $T = 2\pi/\nu$ between its high value $f_0 + \Omega$ and low value $f_0 - \Omega$. By contrast with a theoretician, an experimenter has to deal not with mathematical formulas but with real signals and his job is to come up with a signal processing method that will discover the existence of frequency oscillations in the simplest tune.

The mathematical tool that is helpful in this situation is called the *windowed Fourier transform* which is just the usual Fourier transform

$$X(f, \tau) = \int_{-\infty}^{\infty} x(t) w(t - \tau) e^{-j2\pi f t}\, dt \tag{3.2.4}$$

of the *time-windowed signal* $x(t)w(t - \tau)$, where $w(t)$ is the *windowing function* that usually is chosen to have value equal to 1 in a vicinity of the origin $t = 0$ (say, inside an interval of length λ), and that either vanishes or has values very close to 0 outside this neighborhood. This windowing function property will assure effective time-localization.

Usually, one defines the windowing function $w(t)$ via a *windowing shape function* $w_0(s)$ of a dimensionless variable s and the formula

$$w(t) = w_0(t/\lambda), \tag{3.2.5}$$

where λ is a scaling parameter. Some typical examples of normalized ($\|w_0\| = 1$) windowing shape functions are (see Fig. 3.2):

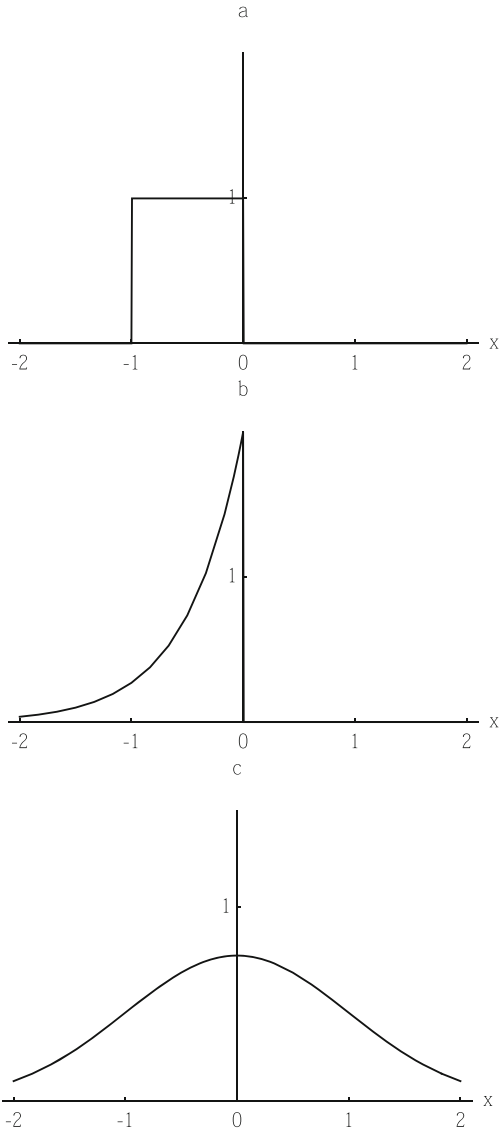


Fig. 3.2 Examples of windowing shape functions. (**a**) Finite memory window; (**b**) Relaxation window; (**c**) Gaussian window

(a) *Finite memory window*

$$w_0(s) = \chi(s+1) - \chi(s); \tag{3.2.6a}$$

(b) *Relaxation window*

$$w_0(s) = 2\chi(-s)\exp(2s); \tag{3.2.6b}$$

(c) *Gaussian window*

$$w_0(s) = \pi^{-1/4}\exp(-s^2/2) \tag{3.2.6c}$$

where $\chi(s)$ is the unit step function.

Shift τ centers the window at different locations on the time-axis t. If $x(t)$ is a time-dependent signal and processing is performed in the *real-time*, then τ is just the current time of the experiment and the time-window $w(t)$ has to satisfy the causality principle, i.e., $w(t) \equiv 0$ for $t > 0$. So, in this case, the finite memory and relaxation windows are appropriate but the Gaussian window is not. If the whole signal is recorded before processing, or the variable t has other interpretation (e.g., space or angle variable), then the experimenter has more freedom in selecting the windowing shape function, and very often the Gaussian window is a good candidate.

3.2.2 Frequency Localization

The time-window $w(t)$ was designed to separate well the time-localized pieces of duration λ of the incoming signal $x(t)$. Luckily, it turns out that the Fourier image of the time-window $w(t)$ can help in frequency localization. To see how this happens let us express the original signal $x(t)$ through its Fourier transform:

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ftf}\,df, \tag{3.2.7}$$

and substitute it into the right-hand side of (3.2.4). Note that, in the case of the simplest tune (3.2.1), $X(f)$ exists only in the distributional sense. The change of the integration order gives that

$$X(f,\tau) = e^{-j2\pi f\tau}\int_{-\infty}^{\infty} X(f')W(f-f')e^{j2\pi f'\tau}\,df'. \tag{3.2.8}$$

Remarkably, except for the nonessential factor in front of the integral, this expression looks like the symmetric counterpart of (3.2.4) in the frequency domain. Now, the role of the signal is played by its Fourier image $X(f)$ and the time-window has been replaced by the *frequency-window* $W(f)$, that is the Fourier transform of the time window $w(t)$.

The uncertainty principle (3.1.4) tells us that if the effective duration of the time-window is λ, then one can expect the effective width of the frequency-window to be of order at least $1/\lambda$. In terms of the dimensionless window shapes $w_0(t)$ and $W_0(f)$ where, similar to (3.2.5),

$$W(f) = \lambda W_0(\lambda f), \tag{3.2.9}$$

both $w_0(t)$ and $W_0(f)$ have to have a similar effective widths ~ 1.

However, the actual situation is a bit more complicated than the above juggling of the uncertainty principle may indicate. When the engineers talk about effectively localized frequency-window, they think about the compact support of the frequency-windowing shape function $W_0(f)$ or at least about its rapid decay outside a finite frequency band. However, we know from the properties of the Fourier transform that it is impossible for both the function and its Fourier transform to have compact supports. Furthermore, the frequency windowing shape function will decay rapidly for $|f| > 1$ only if the time windowing shape function is smooth. This fact eliminates time windowing shape functions (3.2.6a) and (3.2.6b), which have good time-localization properties, as good candidates for good frequency localization by their Fourier transforms. Abrupt truncations in them introduce discontinuities of the first kind which slow the decay of their Fourier transforms. For example, the modulus of Fourier image of the relaxation window (3.2.6b)

$$|W_0(f)| = \frac{1}{\pi\sqrt{4+f^2}} \tag{3.2.10}$$

decays to zero slowly as $|W_0(f)| \sim 1/(\pi|f|)$, $(f \to \infty)$.

So, to achieve better frequency localization one has to take smoother windowing shape functions.

Example 3.2.1 (Compact Time Window and Power-Law Decay of the Frequency Window) Take the windowing shape function

$$w_0(t) = \frac{8}{3}\Big[\chi(t+2) - \chi(t)\Big]\sin^2\left(\frac{\pi t}{2}\right). \tag{3.2.11}$$

normalized appropriately and shifted to satisfy the causality principle. Its frequency counterpart

$$W_0(f) = e^{jf}\frac{4}{3}\pi\frac{\sin f}{f(\pi^2 - f^2)} \tag{3.2.12}$$

decays as $1/|f|^3$, faster than (3.2.11), which produces tolerable frequency localization while preserving perfect time localization. The power law of the frequency windowing shape (3.2.12) decay was caused by hidden discontinuities (in the second derivative) of the time windowing shape (3.2.11).

Example 3.2.2 (Gaussian Time and Frequency Windows; Gabor Transform) Since the Fourier image of a Gaussian time windowing shape gives a Gaussian frequency windowing shape, in this case we have excellent localization in both time and frequency domains. Indeed, if $w_0(t)$ is given by (3.2.6c), then, using the symmetric definition of the Fourier transform, and its inverse,

$$X(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t) e^{-jft}\, dt, \qquad x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (X(f)) e^{jft}\, df, \tag{3.2.13}$$

we obtain

$$W_0(f) \equiv w_0(f).$$

The windowed Fourier transform

$$G(f, \tau) = \pi^{-1/4} \int_{-\infty}^{\infty} x(t) e^{jft - (t-\tau)^2/2}\, dt \tag{3.2.14}$$

based on the Gaussian window is called the *Gabor transform* in honor of the physicist who introduced it for studying quantum-mechanical problems.

3.3 Continuous Wavelet Transforms

3.3.1 Definition and Properties of Continuous Wavelet Transform

In this section we take a general look at the continuous wavelet transform both theoretically and as it relates to physical and engineering problems. Mathematical questions concerning particular wavelet systems will be dealt with in the last three sections of this chapter.

The *continuous wavelet image* of signal $x(t)$ is defined by

$$\hat{X}(\lambda, \tau) = A(\lambda) \int x(t) \psi^* \left(\frac{t - \tau}{\lambda} \right) dt, \qquad \lambda > 0, \tag{3.3.1}$$

where $\psi(z)$ is a certain function called the *mother wavelet* and λ and τ are called, respectively, the *scale variable* and the *location variable*. Function $A(\lambda)$ will be specified later. Note that to distinguish it from the Fourier transform

$$X(f) = \int x(t) e^{-j2\pi ft}\, dt, \tag{3.3.2}$$

the continuous wavelet transform will be denoted by applying a "hat" to X.

A compression and dilation of the mother wavelet are accomplished for the continuous wavelet transform by the scaling parameter λ. In a sense, one can interpret the value of the continuous wavelet transform $\hat{X}(\lambda, \tau)$ as a measure of the contribution of the rescaled by λ mother wavelet $\psi((t-\tau)/\lambda)$ to the signal $x(t)$.

The coefficient $A(\lambda)$ can be selected arbitrarily as to magnify or reduce sensitivity of the transform to different scales. However, very often it is simply selected as

$$A(\lambda) = 1/\sqrt{\lambda}, \tag{3.3.3}$$

so that

$$\hat{X}(\lambda, \tau) = \frac{1}{\sqrt{\lambda}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-\tau}{\lambda}\right) dt. \tag{3.3.4}$$

This choice guarantees that the arbitrary rescaling of the mother wavelet preserves the mother wavelet's L^2-norm. Indeed,

$$\left\| \frac{1}{\sqrt{\lambda}}\psi^*\left(\frac{t-\tau}{\lambda}\right)\right\|_2^2 = \frac{1}{\lambda}\int_{-\infty}^{\infty}\left|\psi\left(\frac{t-\tau}{\lambda}\right)\right|^2 dt = \int_{-\infty}^{\infty}|\psi(s)|^2\, ds = \|\psi(s)\|_2^2. \tag{3.3.5}$$

One could say that with this choice of $A(\lambda)$ all the scales carry equal weight.

As we already mentioned in the previous section, for all its great features discussed at length in Chap. 2, the Fourier transform has from the point of view of a physicist one essential shortcoming: its "mother wavelet" $\exp(j2\pi f t)$ has unbounded support. As a result, based on information contained in the Fourier image $X(f)$ it is difficult to assess where signal $x(t)$ (or its special features) is located on the t axis and where it is equal to 0. In particular, this type of information is totally lost in the "spectral density" of the distribution of harmonic components over the frequency f axis. That drawback will be removed in the continuous wavelet transform by selecting a *localized* mother wavelet $\psi(z)$ which decays rapidly to zero as $z \to \pm\infty$. Consequently, in the continuous wavelet transform, in addition to the scale parameter λ, there appears another primary parameter—the location shift τ. Varying it we can track the time t evolution of the "events."

Example 1 (Morlet Wavelets) The often encountered in practical application mother wavelet

$$\psi(z) = e^{jQz}\varphi(z), \tag{3.3.6}$$

with the Gaussian windowing function

$$\varphi(z) = \exp(-z^2/2), \tag{3.3.7}$$

is traditionally called the *complex-valued Morlet wavelet* (the plot of its real part, for $Q = 10$, is shown in Fig. 3.3). As a result, the Fourier image of $\psi(z)$ is also

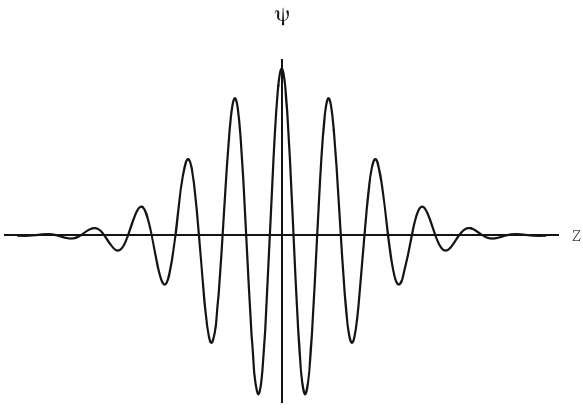


Fig. 3.3 The plot of the real part of the Morlet mother wavelet for $Q = 10$

Gaussian:

$$\Psi(f) = \frac{1}{\sqrt{2}\pi} \exp\left(-\frac{(f-Q)^2}{2}\right). \tag{3.3.8}$$

Recall that the Gaussian shape of the windowing function is the minimizer in the uncertainty principle discussed in Sect. 3.1, and, consequently, it optimizes the joint resolution in time t and frequency ω. Indeed, the continuous wavelet image

$$\begin{aligned}\hat{X}(\lambda,\tau) &= A(\lambda)\int_{-\infty}^{\infty} x(t)\varphi\left(\frac{t-\tau}{\lambda}\right)\exp\left(-j\frac{Q}{\lambda}(t-\tau)\right)dt \\ &= A(\lambda)\int_{-\infty}^{\infty} x(t)\exp\left(-j\frac{Q}{\lambda}(t-\tau)-\frac{(t-\tau)^2}{2\lambda^2}\right)dt \end{aligned} \tag{3.3.9}$$

contains information about the original (not too fast increasing) function $x(t)$ in the window of effective length $\sim \lambda/\sqrt{2}$. Expressing $\hat{X}(\lambda,\tau)$ through the Fourier images of the analyzed functions we get

$$\begin{aligned}\hat{X}(\lambda,\tau) &= 2\pi\lambda A(\lambda)\int_{-\infty}^{\infty} X(\omega)\varphi(\lambda\omega - Q)e^{i\omega\tau}\,d\omega \\ &= \sqrt{2\pi}\lambda A(\lambda)\int_{-\infty}^{\infty} X(\omega)\exp\left(-\frac{\lambda^2}{2}\left(\omega-\frac{Q}{\lambda}\right)^2\right)e^{i\omega\tau}\,d\omega. \end{aligned} \tag{3.3.10}$$

This means that $\hat{X}(\lambda,\tau)$ depends on the values of the Fourier image $X(\omega)$ in the frequency band of width $\sigma[\Psi] = 1/\lambda\sqrt{2}$ centered at the frequency

$$\Omega = Q/\lambda. \tag{3.3.11}$$

In other words, $\hat{X}(\lambda, \tau)$ supplies information about the spectral properties of the original function with resolution $1/\lambda\sqrt{2}$. The arbitrary parameter Q entering in the definition (3.3.6) of the Morlet wavelet could be called the *efficiency factor* of the Morlet wavelet since the quantity $Q/2\pi$ is of the order of the number of Morlet wavelet's periods contained in its window.

Just as the first automobiles of the last century took inspiration from and mimicked the horse-drawn carriages, and only later developed their own identity, the wavelets underwent a similar evolution which started with their identity as "improved" versions of the Fourier transform and only gradually developed into being recognized for their own outstanding capabilities. These capabilities, still far from being fully tapped, are related to the fact that the mathematical theory of wavelets, as we will see later on, imposes very few restrictions on the choice of the mother wavelet's shape. We will illustrate them on concrete applications in the rest of this section.

One of the powerful applications of the continuous wavelet transform is the study of open and hidden singularities in the incoming signal $x(t)$. Usually, the singularities are caused by physical (biological, economic, etc.) laws, whose validity the experimenter is trying to confirm, or come from the existence of the sharp boundaries between the regions where the process $x(t)$ evolves smoothly. The mother wavelets that are useful in this context are quite unlike the Morlet wavelet (3.3.6)–(3.3.7).

Example 2 (Mexican Hat Wavelet) Differentiation can bring to the surface function's hidden singularities. For this reason one often selects mother wavelets so that the corresponding continuous wavelet transform converges, for $\lambda \to 0$, to a desired derivative of the function being analyzed. One of such examples is the *Mexican hat* mother wavelet,

$$\psi(z) = -\frac{d^2}{dz^2}\varphi(z) = (1 - z^2)\exp\left(-\frac{z^2}{2}\right), \tag{3.3.12}$$

which is just the second derivative of the Gaussian function (3.3.7) (Fig. 3.4). Its Fourier image is[1]

$$\Psi(\omega) = \frac{\omega^2}{\sqrt{2\pi}}\exp\left(-\frac{\omega^2}{2}\right). \tag{3.3.13}$$

Substituting (3.3.12) into (3.3.1), and integrating by parts twice, we get

$$\hat{X}(\lambda, \tau) = -A(\lambda)\lambda^2 \int_{-\infty}^{\infty} \varphi\left(\frac{t-\tau}{\lambda}\right)\frac{d^2}{dt^2}x(t)\,dt. \tag{3.3.14}$$

[1]From now onwards, but only in this chapter, we will use the variable $\omega = 2\pi f$ in the Fourier transform.

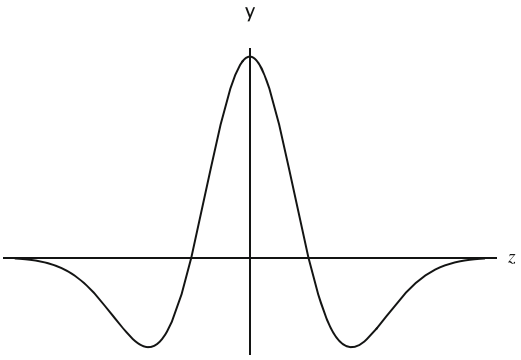


Fig. 3.4 The Mexican hat mother wavelet

It is customary to select

$$A(\lambda) = 1/\lambda^3\sqrt{2\pi}, \tag{3.3.15}$$

so that $\hat{X}(\lambda, \tau)$ converges, for $\lambda \to 0$, to exactly $x''(\tau)$.

Another property of the continuous wavelet transform essential to understand its mechanism is based on the Cauchy-Schwartz inequality

$$\left| \int_{-\infty}^{\infty} x(t) y^*(t)\, dt \right|^2 \leq \int_{-\infty}^{\infty} |x(t)|^2\, dt \int_{-\infty}^{\infty} |y(t)|^2\, dt, \tag{3.3.16}$$

which applied to the function

$$y(t) = \frac{1}{\sqrt{\lambda}} \psi \left(\frac{t - \tau}{\lambda} \right) \tag{3.3.17}$$

yields the inequality

$$|\hat{X}(\lambda, \tau)|^2 \leq \|f\|_2^2 \|\psi\|_2^2. \tag{3.3.18}$$

The inequality provides an upper bound on possible values of the modulus of the continuous wavelet transform (3.3.4) of $x(t)$. Let us assume, without loss of generality, that both the signal and the mother wavelet are normalized so that $\|x\|_2 = \|\psi\|_2 = 1$.

It is clear that the maximum values are achieved, and the inequality (3.3.18) becomes an equality, if the original function $x(t)$ is equal, for certain $\lambda = \lambda_0$ and $\tau = \tau_0$, to the wavelet

$$f(t) = \frac{1}{\sqrt{\lambda_0}} \psi \left(\frac{t - \tau_0}{\lambda_0} \right). \tag{3.3.19}$$

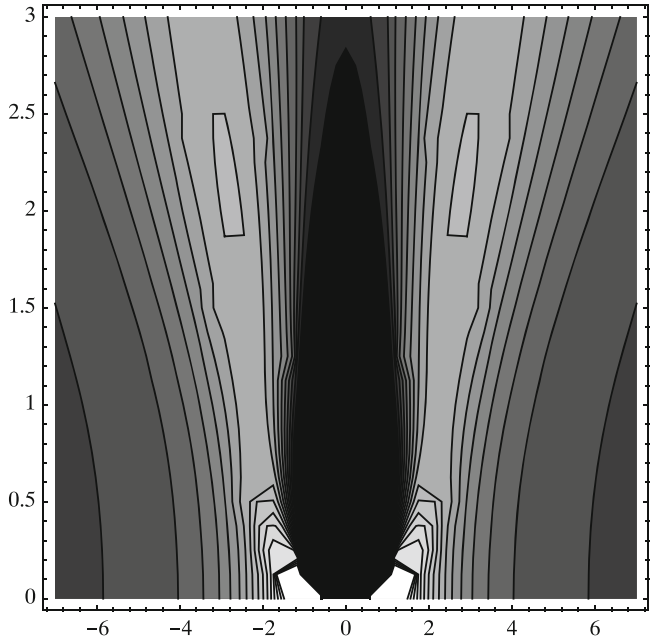


Fig. 3.5 The gray-scale plot of the wavelet image of $x(t) = \exp(-|t|)$ in the case of the Mexican hat mother wavelet. The horizontal axis represents the τ-variable and the vertical—λ-variable. The gray-scale level changes from black to white as the values of the wavelet image increase. The black oval spot in the lower middle portion of the plot is a consequence of the singularity of the original function's second derivative at $t = 0$

Informally, we can say that the continuous wavelet transform is best tuned to, or resonates with signals that have shapes similar to that of the mother wavelet (Fig. 3.5).

Note that the more *complex-structured* the mother wavelet (3.3.17) and the resonating signal (3.3.19) are, the more pronounced the above resonance property of the corresponding continuous wavelet transform is. To make things a bit more formal let us define the signal as *complex-structured* if its time (3.1.2) and frequency (3.1.3) localizations satisfy the *"strong" uncertainty principle*:

$$\sigma^2[x] \cdot \sigma^2[X] \gg 1/4. \tag{3.3.20}$$

Example 4 (Complex-Structured Signal) Let us consider signal $f(t)$ whose Fourier image is the familiar Gaussian function

$$X(\omega) = \frac{\pi^{-1/4}}{\sqrt{2\pi\mu}} \exp\left(-\frac{\omega^2}{2\mu^2}(1 + j\gamma)\right), \tag{3.3.21}$$

where γ is a real number and the constant μ (with the dimension of frequency) has the meaning of effective width of the Fourier image. The coefficient in front of the exponential function has been selected so that the normalization condition

$$\|fx\|_2^2 = 2\pi \|X\|_2^2 = 1$$

is satisfied. Applying the inverse Fourier transform we obtain

$$x(t) = \frac{\sqrt{\mu}}{\pi^{1/4}\sqrt{1+j\gamma}} \exp\left(-\frac{t^2\mu^2}{2(1+j\gamma)}\right). \tag{3.3.22}$$

In turn, using the integral formula

$$\int_{-\infty}^{\infty} t^2 \exp(-r^2t^2)\, dt = \frac{\sqrt{\pi}}{2r^3},$$

we find the frequency and time localizations of the complex-valued signal (3.3.22):

$$\sigma^2[X] = \frac{\mu^2}{2}, \qquad \sigma^2[x] = \frac{1}{2\mu^2}(1+\gamma^2). \tag{3.3.23}$$

Substituting these expressions into (22) we get the following condition for the signal $x(t)$ (3.3.23) to be complex-structured:

$$\gamma \gg 1. \tag{3.3.24}$$

Remark 1 To better see reasons why signal (3.3.22) turned out to be complex-structured let us write the complex-valued Fourier image $X(\omega)$ of an arbitrary signal $x(t)$ in the exponential form

$$X(\omega) = A(\omega)\exp(-j\Phi(\omega)), \tag{3.3.25}$$

where $A(\omega) = |X(\omega)|$ is the nonnegative amplitude and $\Phi(\omega)$—the real phase of the complex Fourier image $X(\omega)$. The amplitude and phase of the Fourier image (3.3.21) of signal (3.3.22) are

$$A(\omega) = \frac{\pi^{-1/4}}{\sqrt{2\pi\mu}} \exp\left(-\frac{\omega^2}{2\mu^2}\right), \qquad \text{and} \qquad \Phi(\omega) = \frac{\gamma\omega^2}{2\mu^2}. \tag{3.3.26}$$

The complex structure of signal (3.3.22) was conditioned on the fast nonlinear variation of the phase of the Fourier image (3.3.21) as a function of ω. Indeed, according to (3.3.3), the signal can be written in the form

$$x(t) = \int A(\omega) \exp\left(j\left(\omega t - \gamma\frac{\omega^2}{2\mu^2}\right)\right) d\omega.$$

Employing the stationary phase method, asymptotically ($\gamma \to \infty$), the value of signal $x(t)$ at a given instant t is determined by the integral contribution in the small neighborhood of the stationary point, in our case $\Omega = 2t\mu^2/\gamma$. Substituting here, instead of Ω, the effective width μ of the Fourier image we shall find the effective duration of the complex-structured signal:

$$T \approx \gamma/\mu, \qquad (T\mu \gg 1).$$

Remark 2 The approximate estimate of the signal (3.3.22) duration obtained above via the stationary phase method may seem unnecessary at first sight since we already know the exact form of the signal and the exact formula for its time localization:

$$\sigma[x] = \sqrt{(1+\gamma^2)/2\mu^2}. \tag{3.3.27}$$

Nevertheless, the above argument has a heuristic value, emphasizing the principal role of the phase in complex-structure signal formation. It also shows a universal method of calculation of its form and duration.

Example 5 (Complex-Structured Mother Wavelet) As another example of mother wavelet let us take function $\psi(z)$ coinciding with the complex-structured signal $x(t)$ (3.3.22). The continuous wavelet image $\hat{X}(\lambda, \tau)$ of function $x(t)$ to which the mother wavelet is perfectly tuned is

$$K(\lambda, \tau) = \frac{1}{\sqrt{\lambda}} \int x(t) x^* \left(\frac{t-\tau}{\lambda} \right) dt. \tag{3.3.28}$$

Recall that the form (5) of the continuous wavelet transform selected here guarantees that, for any λ, the normalization condition

$$\left\| \frac{1}{\sqrt{\lambda}} x \left(\frac{t-\tau}{\lambda} \right) \right\| = 1$$

is satisfied. Notice that we also introduced special notation $K(\lambda, \tau)$ for the special continuous wavelet image of the mother wavelet itself. Function $K(\lambda, \tau)$ is sometimes called the *wideband ambiguity function* of the mother wavelet and it plays an important role in wavelet theory. In terms of the Fourier images

$$K(\lambda, \tau) = 2\pi\sqrt{\lambda} \int_{-\infty}^{\infty} X(\omega) X^*(\omega\lambda) e^{j\omega\tau}\, d\omega, \tag{3.3.29}$$

so that substituting (3.3.21) we obtain

$$K(\lambda, \tau) = \frac{1}{\mu}\sqrt{\frac{\lambda}{\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\rho \frac{\omega^2}{\mu^2} + j\omega\tau \right) d\omega, \tag{3.3.30}$$

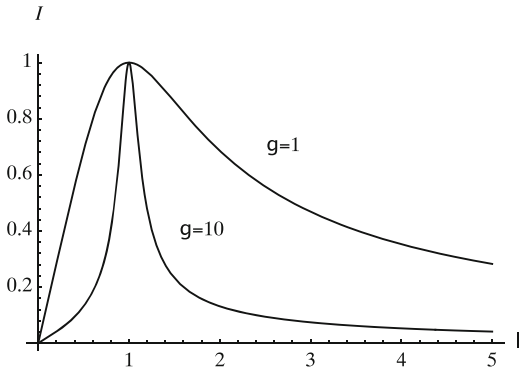


Fig. 3.6 Graphs of function $I(\lambda)$ for different complexity-structure of signals as measured by γ

where

$$\rho = (1 + \lambda^2) + j\gamma(1 - \lambda^2). \tag{3.3.31}$$

Finally, evaluation of the integral (33) gives

$$K(\lambda, \tau) = \sqrt{\frac{2\lambda}{\rho}} \exp\left(-\frac{\tau^2\mu^2}{2\rho}\right).$$

The above function has a maximum at $\tau = 0$, and its modulus square has the following dependence on λ :

$$I(\lambda) = |K(\lambda, 0)|^2 = \frac{2\lambda}{|\rho|} = \frac{2\lambda}{\sqrt{(1 + \lambda^2)^2 + \gamma^2(1 - \lambda^2)^2}}.$$

It is natural to interpret function $I(\lambda)$ as a sort of resonance curve which characterizes the response efficiency of the continuous wavelet transform as a function of the scale parameter λ. Figure 3.6 shows graphs of function $I(\lambda)$ for signals of different complexity, as measured by parameter γ. It is clear from the illustrations that the resonance is best emphasized for large values of γ, that is for signals of large complexity.

The maximal value of $I(\lambda)$ is achieved for $\lambda = 1$. It is related to the fact that for $\lambda = 1$ function (3.3.28) becomes the *autocorrelation function*

$$k(\tau) = \int x(t)x^*(t - \tau)\, dt$$

of the original signal. The autocorrelation function has some remarkable properties. In particular, it transforms any signal, however complex, into a simple signal whose Fourier image,

$$K(\omega) = 2\pi|X\omega)|^2, \tag{3.3.33}$$

is real and nonnegative, with the phase $\Phi(\omega) \equiv 0$. In electrical engineering one often says that all the harmonics of the autocorrelation function $K(t)$ have identical phases.

The autocorrelation function achieves its maximum at $\tau = 0$ and decays relatively rapidly as $|\tau|$ increases. In particular, it is easy to see that its localization properties are determined by

$$\sigma[k] = \frac{1}{\mu\sqrt{2}},$$

so that, in view of (3.3.27), it is clear that for $\gamma \gg 1$ the autocorrelation function $k(\tau) = k(\lambda = 1, \tau)$ is much better localized on the τ-axis than the original signal $x(t)$ (3.3.22) on the t-axis. ∎

3.3.2 *Inversion of the Continuous Wavelet Transform*

As for any other integral transform the basic question is: Does the continuous wavelet image $\hat{X}(\lambda, \tau)$ contain sufficient information permitting recovery of the original function $x(t)$? In more practical terms: Does there exist an inversion formula for the continuous wavelet transform?

To answer these questions let us multiply equality (3.3.1) by $\psi((\theta - \tau)/\lambda)$ and integrate over all τ. The result is the auxiliary integral

$$I(\lambda, \theta) = \int_{-\infty}^{\infty} \hat{X}(\lambda, \tau)\psi\left(\frac{\theta - \tau}{\lambda}\right) d\tau. \tag{3.3.34}$$

Equivalently,

$$I(\lambda, \theta) = A(\lambda) \int_{-\infty}^{\infty} dt\, x(t) \int_{-\infty}^{\infty} d\tau\, \psi^*\left(\frac{t - \tau}{\lambda}\right) \psi\left(\frac{\theta - \tau}{\lambda}\right). \tag{3.3.35}$$

It is easy to see that the inner integral can be expressed via the autocorrelation function (3.3.33)

$$k(z) = \int_{-\infty}^{\infty} \psi(s)\psi^*(s - z)\, ds \tag{3.3.36}$$

of the mother wavelet as follows:

$$\int_{-\infty}^{\infty} \psi^*\left(\frac{t - \tau}{\lambda}\right) \psi\left(\frac{\theta - \tau}{\lambda}\right) d\tau = \lambda k\left(\frac{\theta - t}{\lambda}\right). \tag{3.3.37}$$

As a result,

$$I(\lambda,\theta) = \lambda A(\lambda)\int_{-\infty}^{\infty} f(t)k\left(\frac{\theta - t}{\lambda}\right)dt. \tag{3.3.38}$$

To solve this integral equation for $f(t)$ let us multiply (3.3.41) by a function $B(\lambda)$, to be selected later, and integrate over all λ:

$$\int_0^{\infty} I(\lambda,\theta)B(\lambda)\,d\lambda = \int_{-\infty}^{\infty} f(t)g(\theta - t)\,dt, \tag{3.3.39}$$

where

$$g(s) = \int_0^{\infty} k(s/\lambda)C(\lambda)\,d\lambda, \tag{3.3.40}$$

and

$$C(\lambda) = \lambda A(\lambda)B(\lambda). \tag{3.3.41}$$

Clearly, the right-hand side of (3.3.39) would be reduced to $x(\theta)$, thus solving Eq. (3.3.38) for function $x(t)$ if

$$g(s) = \int_0^{\infty} k(s/\lambda)C(\lambda)\,d\lambda = \delta(s). \tag{3.3.42}$$

Let us find $C(\lambda)$ for which the distributional Eq. (3.3.42) is satisfied. Remembering that the Fourier image of the autocorrelation function $k(z)$ is $2\pi|\tilde{\psi}(\omega)|^2$, we get the equation

$$G(\omega) = 2\pi\int_0^{\infty} |\Psi(\omega\lambda)|^2\lambda C(\lambda)\,d\lambda = 1/2\pi, \tag{3.3.43}$$

equivalent to Eq. (3.3.42). To eliminate the dependence of the above integral on ω we shall select $C(\lambda)$ so that

$$\lambda C(\lambda) = 1/D\lambda. \tag{3.3.44}$$

In this case, (3.3.43) becomes

$$\frac{4\pi^2}{D}\int_0^{\infty} |\Psi(\omega\lambda)|^2\frac{d\lambda}{\lambda} = 1, \tag{3.3.45}$$

where

$$D = 4\pi^2\int_0^{\infty} |\Psi(\kappa)|^2\frac{d\kappa}{\kappa} \tag{3.3.46}$$

is the normalizing constant that can be calculated from (45) by introducing the new variable of integration $\kappa = \omega\lambda$ to get

$$\frac{4\pi^2}{D}\int_0^\infty |\Psi(\kappa)|^2 \frac{d\kappa}{\kappa} = 1, \tag{3.3.47}$$

Putting together (3.3.41), (3.3.44), and (3.3.46) we get that

$$B(\lambda) = \frac{1}{D\lambda^3 A(\lambda)}, \tag{3.3.48}$$

so that, from (3.3.48) and (3.3.39),

$$\frac{1}{D}\int_0^\infty \frac{I(\lambda,\theta)\,d\lambda}{\lambda^3 A(\lambda)} = x(\theta).$$

Substituting expression (3.3.34) for $I(\lambda, t)$ we finally obtain the inverse continuous wavelet transform

$$x(t) = \frac{1}{D}\int_0^\infty \frac{d\lambda}{\lambda^3 A(\lambda)}\int_{-\infty}^\infty d\tau\, \hat{X}(\lambda,\tau)\psi\left(\frac{t-\tau}{\lambda}\right). \tag{3.3.49}$$

In particular, if the continuous wavelet transform is defined by (3.3.4)–(3.3.5), then the inversion formula takes the form

$$x(t) = \frac{1}{D}\int_0^\infty \frac{d\lambda}{\lambda^2\sqrt{\lambda}}\int_{-\infty}^\infty d\tau\, \hat{X}(\lambda,\tau)\psi\left(\frac{t-\tau}{\lambda}\right). \tag{3.3.50}$$

However, the above inversion formulas require several caveats.

Remark 3 The observant reader would have noticed that the passage from (3.3.45) to (3.3.47) is justified only if $|\Psi(\omega)|^2$ is an even function. For that reason formulas (3.3.49)–(3.3.50) are valid only for *two-sided* mother wavelets, as mother wavelets with even square modulus are called. To this class belong all the purely real-valued mother wavelets such as the Mexican hat. On the other hand, the complex-valued Morlet wavelet is not of this type. For that reason mathematicians often work with *one-sided* mother wavelets whose Fourier image is

$$\Psi(\omega) \equiv 0, \qquad \omega \le 0. \tag{3.3.51}$$

For such mother wavelets, instead of (3.3.46) we have the equality

$$G(\omega) = \frac{2\pi}{D}\int_0^\infty |\Psi(\omega\lambda)|^2 \frac{d\lambda}{\lambda} = \frac{1}{2\pi}\chi(\omega). \tag{3.3.52}$$

To explain its consequences let us express the right-hand side of (3.3.39) in terms of the Fourier images $X(\omega)$ and $G(\omega)$:

$$\int_{-\infty}^{\infty} I(\lambda,\theta)B(\lambda)\,d\lambda = 2\pi \int_{-\infty}^{\infty} X(\omega)G(\omega)e^{i\omega\theta}\,d\omega.$$

Substituting here (3.3.34), (3.3.48), and (3.3.52), we arrive at the relation that replaces equality (3.3.49) for one-sided mother wavelets:

$$\int_{-\infty}^{\infty} X(\omega)\chi(\omega)e^{i\omega\theta}\,d\omega = \frac{1}{D}\int_0^{\infty} \frac{d\lambda}{\lambda^3 A(\lambda)} \int_{-\infty}^{\infty} d\tau\,\hat{X}(\lambda,\tau)\psi\left(\frac{\theta-\tau}{\lambda}\right).$$

As we have shown before, the Fourier integral on the left-hand side is, up to coefficient 1/2, equal to the analytic signal

$$x_1(t) = \frac{2}{D}\int_0^{\infty} \frac{d\lambda}{\lambda^3}A(\lambda) \int_{-\infty}^{\infty} d\tau\,\hat{X}(\lambda,\tau)\psi\left(\frac{\theta-\tau}{\lambda}\right) \tag{3.3.53}$$

corresponding to the original signal $x(t)$. Remembering that the real part of the analytic signal coincides with $x(t)$, we arrive at the inversion formula for the continuous wavelet transform for one-sided mother wavelets:

$$x(t) = \frac{2}{D}\,\mathrm{Re}\int_0^{\infty} \frac{d\lambda}{\lambda^3}A(\lambda) \int d\tau\,\hat{X}(\lambda,\tau)\psi\left(\frac{\theta-\tau}{\lambda}\right). \tag{3.3.54}$$

Example 6 (Poisson Wavelets) As an example of one-sided mother wavelets consider

$$\psi_m(z) = (1-jz)^{-m-1}, \qquad m > 0, \tag{3.3.55}$$

which are called *Poisson wavelets.* Their Fourier images

$$\Psi_m(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-j\omega z}\,dz}{(1-iz)^{m+1}} \tag{3.3.56}$$

can be calculated by means of the residues method to be

$$\Psi_m(\omega) = \frac{1}{\Gamma(m+1)}\omega^m e^{-\omega}\chi(\omega). \tag{3.3.57}$$

Poisson wavelets can be used to identify open and hidden singularities of signal $x(t)$ and, for $m = 2$, like the Mexican hat, in the search for edges between different regimes of the original function $x(t)$. Indeed, for $m = 2$, the Poisson wavelet

$$\psi_2(z) = -\frac{1}{2}\frac{d^2}{dz^2}\frac{1}{1-jz}. \tag{3.3.58}$$

Its real part

$$\text{Re } \psi_2(z) = -\frac{1}{2}\frac{d^2}{dz^2}\frac{1}{1+z^2}$$

has a shape similar to that of the Mexican hat and possesses, for $\lambda \to 0$, the same differentiating properties.[2]

3.4 Haar Wavelets and Multiresolution Analysis

In this section we will take a look at a special (one can say "digital") series representation for real-valued signals in terms of the so-called *Haar wavelets.* This idealized system provides a good easy introduction to the concepts of *wavelet transforms* and *multiresolution analysis.* Each term of the expansion will provide information about both the time and the frequency localization of the signal. The Haar wavelets will be obtained from a single prototype—a *mother wavelet*—by translations in time and frequency, although the explicit shift in frequency will be replaced by a more natural in this case *dilation* (rescaling, stretching) in time. This will guarantee that all the wavelets have the same shape. To eliminate redundancy and overdetermination, we will make the wavelet system orthogonal.

The *Haar mother wavelet* is defined as follows:

$$\psi(t) = \begin{cases} 1, & \text{for } 0 \le t < 1/2; \\ -1, & \text{for } 1/2 \le x < 1; \\ 0, & \text{otherwise.} \end{cases} \tag{3.4.1}$$

The *Haar wavelet*

$$\psi_{m,n}(t) := 2^{m/2}\psi(2^m(t - 2^{-m}n)), \tag{3.4.2}$$

of order (m, n), $m, n = \ldots, -1, 0, 1, \ldots$, is obtained by rescaling (dilating or compressing) the time in the mother wavelet $\psi(t)$ by a factor of 2^m and then translating the resulting wavelet by an integer n multiplicity of 2^{-m}. The dilation makes the wavelet $\psi_{m,n}(t)$ fit in the interval of length 2^{-m}, and the translation places its support finally in the interval $[2^{-m}n, 2^{-m}(n+1)]$ (see Fig. 3.7). We will call parameter m—the *level of resolution* of the wavelet, and parameter n—the *location* parameter of the wavelet. Then the number 2^{-m} can be seen as its *resolution*, and $2^{-m}n$—as its *location*.

[2]For a more detailed analysis of the theoretical properties of wavelets, see, e.g., A.I. Saichev and W.A. Woyczyński, *Distributions in the Physical and Engineering Sciences, Volume 1, Distributions and Fractal Calculus, Integral Transforms and Wavelets,* 1997.

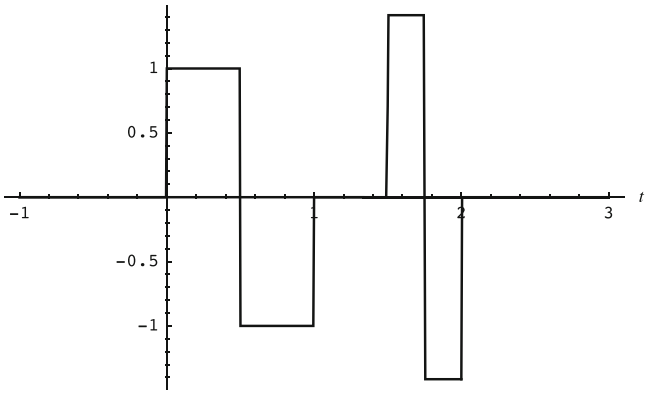


Fig. 3.7 The Haar mother wavelet and a wavelet of order (1,3)

The coefficient $2^{m/2}$ in the definition (3.4.2) was selected to make all the Haar wavelets normalized in $L^2(\mathbf{R})$, that is, to compensate for the dilation operation to guarantee that

$$\|\psi_{m,n}\|_2^2 = \int_{-\infty}^{\infty} \psi_{m,n}^2(t)\, dt = 1. \tag{3.4.3}$$

It turns out that:

The system of Haar wavelets

$$\psi_{m,n}(t), \qquad m, n = \ldots -2, -1, 0, 1, 2, \ldots \tag{3.4.4}$$

is orthogonal, that is

$$\langle \psi_{j,k}, \psi_{m,n} \rangle = \int_{-\infty}^{\infty} \psi_{j,k}(t)\psi_{m,n}(t)\, dt = 0, \quad \text{if} \quad (j,k) \neq (m,n), \tag{3.4.5}$$

and complete in $L^2(\mathbf{R})$. *The latter means that any function* $x(t)$ in $L^2(\mathbf{R})$ *has an* L^2*-convergent representation*

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} w_{m,n}\psi_{m,n}(t) \tag{3.4.6}$$

where, in view of the orthonormality, the expansion coefficients

$$w_{m,n} = w_{m,n}[x] = \langle x, \psi_{m,n} \rangle = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)\, dt. \tag{3.4.7}$$

The above properties of orthogonality and completeness parallel properties of the trigonometric system of functions on a finite interval (say, $[0, 2\pi]$) which give rise to the usual Fourier series expansions.

The orthogonality (3.4.5) can be shown as follows. For a fixed resolution level $j = m$, if location parameters k, n are different, then wavelets $\psi_{j,k}(t)$ and $\psi_{j,n}(t)$ have disjoint supports, and the integral of their product is clearly zero. At different resolution levels, say $j < m$, either the supports of $\psi_{j,k}(t)$ and $\psi_{m,n}(t)$ are disjoint and the previous argument applies, or the support of $\psi_{m,n}(t)$ sits entirely within the interval where $\psi_{j,k}(t)$ is constant (either $+2^{j/2}$ or $-2^{j/2}$), and again the integral of their product vanishes because

$$\int_{-\infty}^{\infty} \psi_{j,k}(t)\psi_{m,n}(t)\,dt = \pm 2^{j/2} \int_{-\infty}^{\infty} \psi_{m,n}(t)\,dt = 0. \qquad \blacksquare$$

The completeness of the Haar wavelet system (4) is more difficult to establish and the proof relies on demonstrating that if all the wavelet coefficients $w_{m,n}[x] = 0$, then function $x(t)$ is necessarily 0 in L^2. We will give a flavor of the proof by showing that this is indeed the case if $x \in L^1 \cap L^2$. So, assume that $w_{m,n} = 0,\ m, n = \ldots -1, 0, 1, \ldots$.

Since $w_{0,0} = 0$ then

$$\int_0^{1/2} x(t)\,dt = \int_{1/2}^{1} x(t)\,dt = \frac{1}{2}\int_0^1 x(t)\,dt.$$

However, since $w_{-1,0} = 0$,

$$\int_0^{1} x(t)\,dt = \int_{1}^{2} x(t)\,dt = \frac{1}{2}\int_0^2 x(t)\,dt$$

and, by induction, for any n

$$\int_0^{1/2} x(t)\,dt = \frac{1}{2^{n+1}}\int_0^{2^n} x(t)\,dt = \lim_{n\to\infty} \frac{1}{2^{n+1}}\int_0^{2^n} x(t)\,dt = 0,$$

since we assumed the finiteness of the integral $\int |x(t)|\,dt$ ($f \in L_1$). Clearly, the same argument can be repeated for any dyadic interval of the form $[2^{-m}n, 2^{-m}(n+1)]$, so that, by approximation, for any interval $[a, b]$

$$\int_{[a,b]} x(t)\,dt = 0.$$

This implies that $x = 0$ in $L^1 \cap L^2$, and the proof of completeness of the Haar wavelets is done. $\blacksquare$

Remark 1 Observe a seemingly paradoxical nature of expansion (3.4.6), where an arbitrary square integrable function in L^1, for which in general $\int x(t)\,dt \neq 0$, has an expansion into a series of Haar wavelets for which $\int \psi_{m,n}(t)\,dt = 0$. The explanation is that the convergence of the series (3.4.6) is in L^2 (that is in the mean square sense) so that the integrals themselves need not be preserved in the limit. To avoid this phenomenon one sometimes considers functions $I_{[0,1]}(t-n), n = \ldots, -1, 0, 1, \ldots$ in combination with the Haar wavelet subsystem $\psi_{m,n}$ for $m \geq 0$. We will return to this theme later.

Note that the inner series in the expansion (6) consists of wavelets of fixed *resolution* 2^{-m}, that is, it represents a function with constant values on dyadic intervals $[2^{-m}n, 2^{-m}(n+1)]$, $n = \ldots, -1, 0, 1, \ldots$, and gives the *contents of function* $x(t)$ *at fixed resolution level* m (see Fig. 3.8) Then the partial sum

$$x_{R,S}(t) = \sum_{m=R}^{S} \sum_{n=-\infty}^{\infty} w_{m,n}\psi_{m,n}(t) \tag{3.4.8}$$

of the expansion (3.4.6) gives an approximation of function $x(t)$ at resolutions finer than 2^{-R} and coarser than 2^{-S} (see Fig. 3.9).

Thus expansion (3.4.6) may be interpreted as a *multiresolution analysis* of the function space $L^2(\mathbf{R})$.

Remark 2 (Scaling Function) We have already observed in Remark 1 that the multiresolution analysis of functions in $L^2(\mathbf{R})$ can be accomplished by means of a slightly different system that starts out with the *scaling function*

$$\varphi(t) = I_{[0,1]}(t)$$

and its integer translates

$$\varphi_n(t) = \varphi(t-n), \quad n = \ldots, -1, 0, 1, \ldots,$$

and supplements them with Haar functions $\psi_{m,n}(t)$ with nonnegative resolution levels $m = 0, 1, 2, \ldots$ and arbitrary integer location parameter n. Note that the resulting system is still orthonormal and complete, and gives a multiresolution expansion of a function $f \in L^2(\mathbf{R})$ of the form

$$f = \sum_{n=-\infty}^{\infty} w_n\varphi_n(t) + \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} w_{m,n}\psi_{m,n}, \tag{3.4.9}$$

with coefficients

$$w_n = w_n(f) = \int_{-\infty}^{\infty} f(t)\varphi_n(t)\,dt, \tag{3.4.10}$$

and $w_{m,n}$ as in formula (3.4.7).

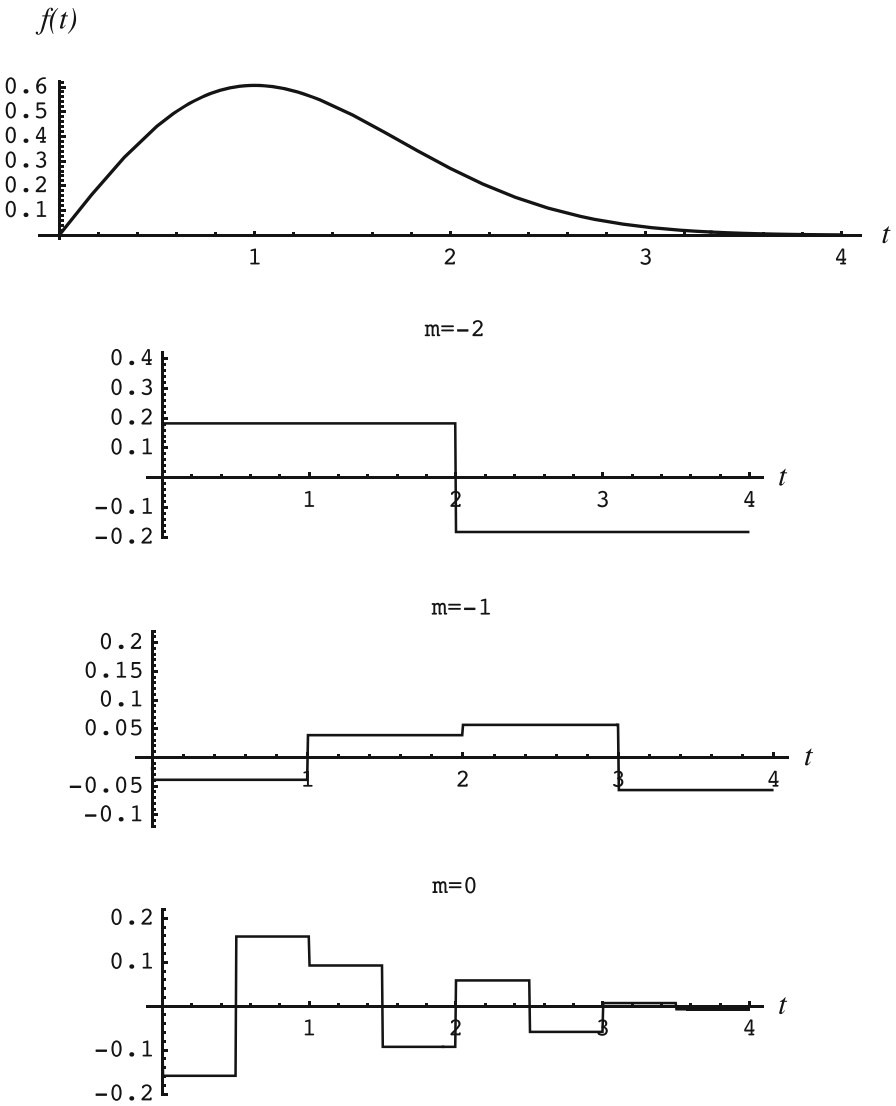


Fig. 3.8 Function $f(t) = t\exp(-t^2/2)$ and its contents at resolution levels $m = -2, -1, 0$

Remark 3 (Self-Similar (Fractal) Properties of Haar Wavelets) The crucial observation for the general theory of wavelets (to be discussed in the next section) is that the scaling function $\varphi(t)$ (the indicator function of the interval [0,1]) is *self-similar* in the sense that it satisfies the *scaling relation*

$$\varphi(x) = \varphi(2x) + \varphi(2x-1), \tag{3.4.11}$$

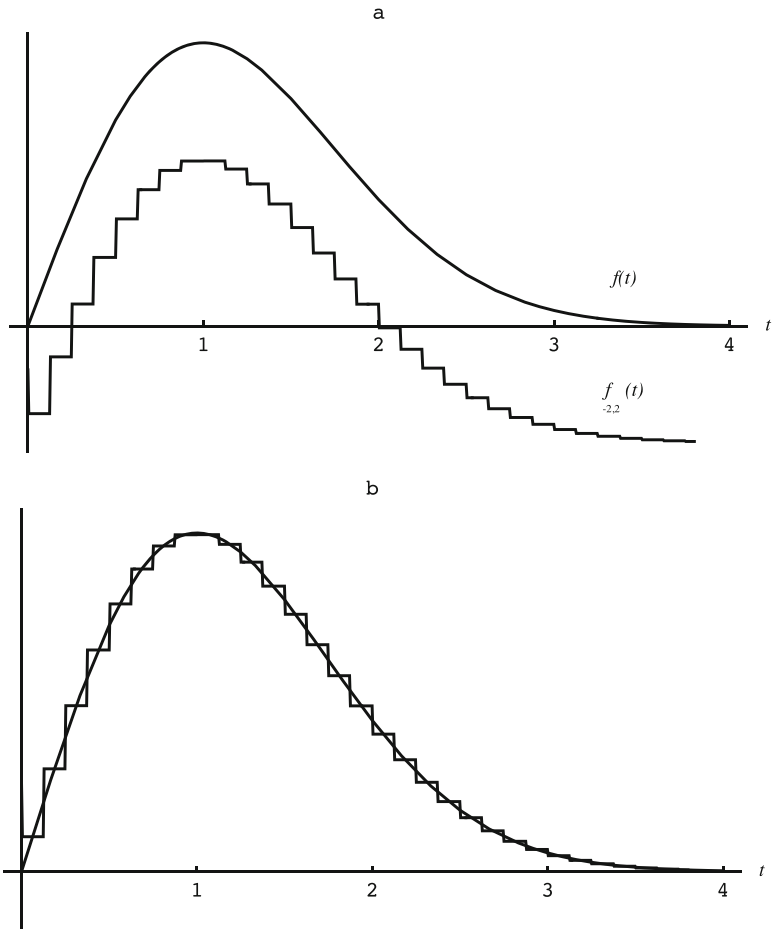


Fig. 3.9 (**a**) Approximation of function $x(t) = t\exp(-t^2/2)$ with resolution finer than 2^2 and coarser than 2^{-2}. In view of the definition (1) of the Haar mother wavelet, $\int_0^4 x_{-2,2}(t)dt = 0$, while $\int_0^4 x(t)\,dt \neq 0$, which leaves a vertical gap between x and its approximation $x_{-2,2}$ (see Remarks 1 and 2). (**b**) Addition of the constant $c_R = \sum_{m=-\infty}^{-R-1} w_{m,0}2^{m/2}$ (in our case, $R = 2$) removes the gap

and that the mother wavelet $\psi(t)$ can be obtained from the scaling function via the formula

$$\psi(x) = \varphi(2x) - \varphi(2x-1). \tag{3.4.12}$$

The scaling relation (3.4.11) asserts that the scaling function is a certain linear combination of its own dilations and translations. It completely characterizes the indicator function $\varphi(t)$ up to a constant multiplier. Indeed, given values of $\varphi(t)$ at $t = 0$ and 1, the scaling relation (3.4.11) permits computation of values of φ at all dyadic rationals, i.e., real numbers of the form $2^{-m}n$.

3.5 Continuous Daubechies' Wavelets

The Haar wavelets discussed in the previous section enjoyed many useful properties such as orthonormality, completeness, compact support, and self-similarity but, as elegant as their construction was, they were anything but smooth. As a matter of fact they were not even continuous—a property important in many applications. So, in the present section we will explore the possibility of constructing smoother wavelets.

Since the scaling relation (3.4.11) characterizes the indicator scaling function I and thus the Haar wavelets, more complex scaling relations will have to be allowed. It turns out that one can find smooth scaling functions which satisfy a scaling relation

$$\varphi(t) = \sum_{k=0}^{N} a_k \varphi(2t - k)$$

for some positive integer $N > 2$ and coefficients a_k (by (3.4.11), $N = 2$ was necessary and sufficient for Haar wavelets). Then the mother wavelet can be selected to be

$$\psi(t) = \sum_{k=0}^{N} (-1)^k a_{N-k} \varphi(2t - k),$$

and the corresponding wavelet system can be built with its help via formula (3.4.2). Such an approach was suggested by Ingrid Daubechies in 1988, and the resulting wavelets are called *Daubechies wavelets*.

Conceptually, the above construction is a clear-cut generalization of the construction of Haar wavelets from the scaling function I provided in the previous section. However, for $N > 2$, the selection of coefficients a_k becomes highly nontrivial. Also, as a rule, the smoother one wants the wavelets to be, the larger N has to be selected.

Below, we provide a sketch of the relatively simple construction of continuous Daubechies wavelets which is due to David Pollen (1992). Their scaling function $\varphi(t)$ satisfies the *scaling relation*

$$\varphi(t) = a\varphi(2t) + (\overline{1-a})\varphi(2t-1) + (1-a)\varphi(2t-2) + \bar{a}\varphi(2t-3), \qquad (3.5.1)$$

where

$$a = \frac{1+\sqrt{3}}{4}, \qquad (3.5.2)$$

and where, for real numbers of the form $\alpha + \beta\sqrt{3}$ with (dyadic) rational α, β, the overline indicates the "conjugation" operation,

$$\overline{\alpha + \beta\sqrt{3}} = \alpha - \beta\sqrt{3}.$$

The support of the resulting $\varphi(t)$ is contained in the interval [0, 3] and, additionally,

$$\sum_{k=-\infty}^{\infty} \varphi(k) = 1. \tag{3.5.3}$$

Assume that there exists a scaling function $\varphi(t)$ supported by [0,3] and satisfying (3.5.1) and (3.5.3) for integer values of the argument t. The scaling relation (3.5.1) written for $t = 0, 1, 2, 3$ becomes a matrix equation

$$\begin{pmatrix} \varphi(0) \\ \varphi(1) \\ \varphi(2) \\ \varphi(3) \end{pmatrix} = \begin{pmatrix} a & 0 & 0 & 0 \\ (1-a) & \overline{1-a} & a & 0 \\ 0 & \bar{a} & (1-a) & \overline{1-a} \\ 0 & 0 & 0 & \bar{a} \end{pmatrix} \begin{pmatrix} \varphi(0) \\ \varphi(1) \\ \varphi(2) \\ \varphi(3) \end{pmatrix}$$

which, in view of condition (3.5.3), has exactly one solution:

$$\varphi(0) = 0, \quad \varphi(1) = \frac{1+\sqrt{3}}{2}, \quad \varphi(2) = \frac{1-\sqrt{3}}{2}, \quad \varphi(3) = 0.$$

Starting with these prescribed values and using the scaling relation (3.5.1) one can produce values of the scaling function $\varphi(t)$ for any dyadic rational t. For example,

$$\varphi(1/2) = \frac{2+\sqrt{3}}{4}, \quad \varphi(3/2) = 0, \quad \varphi(5/2) = \frac{2-\sqrt{3}}{4},$$

and so on.

The values of $\varphi(t)$ for dyadic t are clearly of the form $\alpha + \beta\sqrt{3}$ with dyadic α and β. One can also prove that they also satisfy two extended *partition of unity* (see also (3.5.3)) formulas

$$\sum_{k=-\infty}^{\infty} \varphi(t-k) = 1$$

and

$$\sum_{k=-\infty}^{\infty} \left(\frac{3-\sqrt{3}}{2} + k \right) \varphi(t-k) = t.$$

Since the support of $\varphi(t)$ is contained in [0, 3] the above properties also give the interval translation properties for dyadic $t \in [0, 1]$:

$$2\varphi(t) + \varphi(t+1) = t + \frac{1+\sqrt{3}}{2},$$

$$2\varphi(t+2) + \varphi(t+1) = -t + \frac{3-\sqrt{3}}{2},$$

$$\varphi(t) - \varphi(t+2) = t + \frac{-1+\sqrt{3}}{2}.$$

Combining them with the scaling relation (3.5.1) gives the scaling relations for dyadic $t \in [0, 1]$:

$$\varphi\left(\frac{0+t}{2}\right) = a\varphi(t);$$

$$\varphi\left(\frac{1+t}{2}\right) = \bar{a}\varphi(t) + at + \frac{2+\sqrt{3}}{4};$$

$$\varphi\left(\frac{2+t}{2}\right) = a\varphi(1+t) + \bar{a}t + \frac{\sqrt{3}}{4};$$

$$\varphi\left(\frac{3+t}{2}\right) = \bar{a}\varphi(1+t) - at + \frac{1}{4}; \tag{3.5.4}$$

$$\varphi\left(\frac{4+t}{2}\right) = a\varphi(2+t) - \bar{a}t + \frac{3-2\sqrt{3}}{4};$$

$$\varphi\left(\frac{5+t}{2}\right) = \bar{a}\varphi(2+t).$$

Compared with the original scaling relation (3.5.1), they have a clear advantage: the values of $\varphi(t)$ at the next resolution level depend only on one value at the previous resolution level (instead of four in (3.5.1)).

The above formulas form a basis for the following recursive construction of the continuous version of the scaling function on the whole interval [0,3]. Start with function $g_0(t)$ which is equal to $\varphi(t)$ at integers 0,1,2,3, and which linearly interpolates φ in-between these integers. Clearly, $g_0(t)$ is continuous. In the next step, form $g_1(t)$ at the second resolution level by applying the (right-hand sides of) scaling relations (3.5.4) to g_0. More precisely, for $t \in [0, 1]$, define

$$g_1\left(\frac{0+t}{2}\right) = ag_0(t);$$

$$g_1\left(\frac{1+t}{2}\right) = \bar{a}g_0(t) + at + \frac{2+\sqrt{3}}{4};$$

$$g_1\left(\frac{2+t}{2}\right) = ag_0(1+t) + \bar{a}t + \frac{\sqrt{3}}{4};$$

$$g_1\left(\frac{3+t}{2}\right) = \bar{a}g_0(1+t) - at + \frac{1}{4};$$

$$g_1\left(\frac{4+t}{2}\right) = ag_0(2+t) - \bar{a}t + \frac{3-2\sqrt{3}}{4};$$

$$g_1\left(\frac{5+t}{2}\right) = \bar{a}g_0(2+t).$$

Outside $[0, 3]$ set $g_1(t) = 0$. Function $g_1(t)$ is continuous and coincides with $\varphi(t)$ at dyadic points with resolution 2^{-1} (in-between, it again provides a linear interpolation). Continuing this procedure we obtain a sequence g_n of continuous, piecewise linear functions (zero outside $[0, 3]$) which agree with $\varphi(t)$ at dyadic points of the form $k2^{-n}$.

Notice that functions $|g_n(t)| \leq 3$ for all $n = 1, 2, \ldots$, and since $0 \leq |\bar{a}| \leq a < 1$ (see (3.5.2)), we get that

$$\max_t |g_k(t) - g_{k+j}(t)| \leq a^k \max_t |g_0(t) - g_j(t)| \leq 6a^k.$$

Hence the sequence of functions $g_k(t)$ satisfies uniformly the Cauchy condition, and the limit

$$\varphi(t) = \lim_{n\to\infty} g_n(t)$$

is a continuous function. This is the scaling function we were searching for.

Remark 1 Note that the scaling function $\varphi(t)$ is not differentiable because

$$\lim_{j\to\infty} \frac{\varphi(2^{-j}) - \varphi(0)}{2^{-j}} = \lim_{j\to\infty} \frac{\varphi(2^{-j})}{2^{-j}} = \lim_{j\to\infty} \frac{a^j\varphi(1)}{2^{-j}} = \lim_{j\to\infty} (2a)^j\varphi(1) = \infty,$$

since $2a > 1$ and $\varphi(0) \neq 0$.

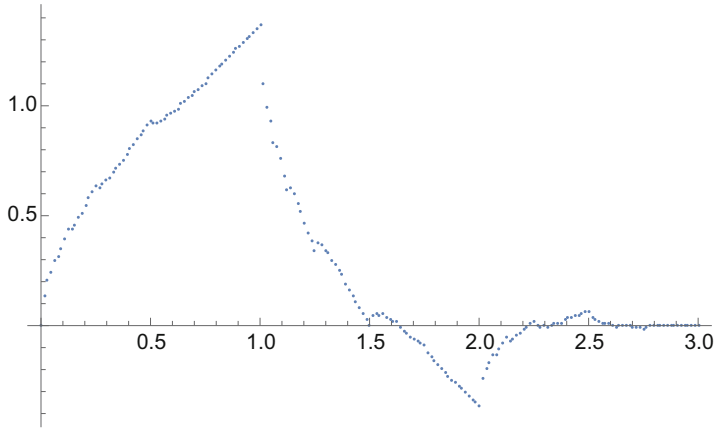


Fig. 3.10 Values of the Daubechies' scaling function computed at dyadic points $t = n \cdot 2^{-6}$, $0 \le t \le 3$, via the scaling relation (3.5.1)

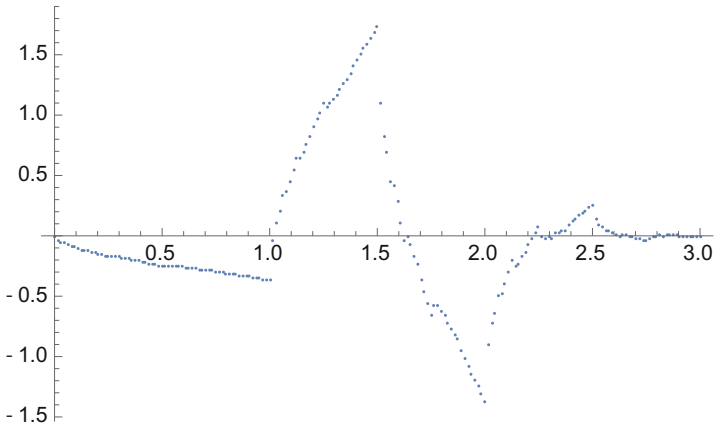


Fig. 3.11 Values of the Daubechies' mother wavelet computed at dyadic points $t = n \cdot 2^{-6}$, $0 \le t \le 3$, via the formula (3.5.5)

With some additional work one can now establish that

$$\int_{-\infty}^{\infty} \varphi(t)\, dt = 1,$$

and that the integer translations of $\varphi(t)$ form an orthonormal system, that is (Figs. 3.10 and 3.11)

$$\int_{-\infty}^{\infty} \varphi(t)\varphi(t-k)\, dt = \begin{cases} 0, & \text{if } k \neq 0; \\ 1, & \text{if } k = 0. \end{cases}$$

Following the general scheme explained in detail for the Haar wavelets in Sect. 3.4, we can now define the mother wavelet $\psi(t)$ via equality

$$\psi(t) = -\bar{a}\varphi(2t) + (1-a)\varphi(2x-1) - (\overline{1-a})\varphi(2x-2) + a\varphi(2x-3) \qquad (3.5.5)$$

and check that

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0.$$

The integer shifts of the mother wavelet are orthonormal, that is

$$\int_{-\infty}^{\infty} \psi(t)\psi(t-k)\, dt = \begin{cases} 0, & \text{if} k \neq 0; \\ 1, & \text{if } k = 0. \end{cases}$$

Moreover, the scaling function φ and the mother wavelet ψ are orthogonal as well, that is

$$\int_{-\infty}^{\infty} \varphi(t)\psi(t-k)\, dt = 0.$$

Thus, again, by an argument similar to that used for the Haar wavelets, the set of Daubechies wavelets

$$\psi_{m,n}(t) = 2^{m/2}\psi(2^m - n), \quad m, n = \ldots -1, 0, 1 \ldots$$

forms an orthonormal complete basis in $L^2(\mathbf{R})$, and so does the set of functions

$$\varphi_n(t) = \varphi(t-n), \quad n = \ldots, -1, 0, 1, \ldots,$$

$$\psi_{m,n}(t) = 2^{m/2}\psi(2^m - n), \quad m = \ldots 0, 1, 2 \ldots, \quad n = \ldots, -1, 0, 1, \ldots.$$

3.6 Exercises

Windowed Fourier Transform

1 Let $\tilde{f}(\omega, \tau)$ be the windowed Fourier transform of the signal $f(t)$. Denote by $\tilde{f}'(\omega, \tau)$ the windowed Fourier transform of the derivative $f'(t)$. Express $\tilde{f}'$ in terms of $\tilde{f}$.

2 Signal $x(t)$ is a solution of the differential equation

$$\frac{dx(t)}{dt} + hx(t) = f(t),$$

where $f(t)$ is a signal with known windowed Fourier image $\tilde{f}(\omega, \tau)$ and h is a (real or complex) constant. Express the windowed Fourier image of $x(t)$ in terms of $\tilde{f}(\omega, \tau)$.

3 Let $\tilde{f}(\omega, \tau)$ be the windowed Fourier image of signal $f(t)$ (i.e., $f(t) \mapsto \tilde{f}(\omega, \tau)$). Find the windowed Fourier images of signals (a) $f(t)e^{i\omega_0 t}$, and (b) $f(t + \theta)$.

4 Find the windowed Fourier transform $\tilde{f}(\omega, \tau)$ of function $f(t) = e^{vt}$.

5 Utilizing results of the previous exercises, find the windowed Fourier image of the signal $f(t) = e^{vt} \cos \omega_0 t$.

Wavelets

6 Obtain, in the case of two-sided mother wavelets, a formula connecting $|\hat{X}|^2$ and $|x|^2$, analogous to the Parseval formula for the ordinary Fourier transform.

7 Denote by $\hat{X}_a(\lambda, \tau)$ the continuous wavelet image of signal $x(at)$, $a > 0$, compressed ($a > 1$) or dilated ($a < 1$) in comparison with the original signal $x(t)$. Find out how $\hat{X}_a(\lambda, \tau)$ is related to the continuous wavelet image of signal $x(t)$ itself in the case of wavelet transform definition (3.3.4).

8 Find $\hat{X}(\lambda, \tau)$ (Fig. 7.10) for the self-similar signal $x(t) = |t|^\alpha$.

9 Let

$$x(t) = (5 - 9t + 4t^2)/(5 - 12t + 8t^2)^3.$$

Find numerically and graphically the Haar wavelet expansion of $x(t)$ with resolution level coarser than 0 and finer than 6. Graph the resolution level n contents of $x(t)$ for $n = 0, 1, \ldots, 6$. Use your computer in order to estimate numerically the maximum error of your approximation.

10 Use your computer and the defining scaling relations to produce *numerical* values of the Daubechies scaling function and mother wavelet at the dyadic points up to resolution level 6.

Chapter 4
Random Quantities and Random Vectors

Abstract By definition, values of random signals at a given sampling time are random quantities which can be distributed over a certain range of values. The tools for the precise, quantitative description of those distributions are provided by the classical *probability theory*. However natural, it's development has to be handled with care since the overly heuristic approach can easily lead to apparent paradoxes.[1] But the basic intuitive idea, that, for independently repeated experiments, probabilities of their particular outcomes correspond to their relative frequencies of appearance, is correct. Although the concept of probability is more elementary than the concept of cumulative probability distribution function, we assume that the reader is familiar with the former at the high school level, and start our exposition with the latter which not only applies universally to all types of data, both discrete and continuous, but also gives us a tool to immediately introduce the probability calculus ideas, including the physically appealing probability density function.

4.1 Discrete, Continuous, and Singular Random Quantities

Think here about an electrical engineer whose responsibility is to monitor the voltage on the electrical outlets in the university's circuits laboratory. The record of a month worth of daily readings on a very sensitive voltmeter may look as follows:

```
109.779, 109.37, 110.733, 109.762, 110.364, 110.73,
109.906, 110.378, 109.132, 111.137, 109.365, 108.968,
111.275, 110.806, 110.99, 111.522, 110.728, 109.689,
111.163, 107.22, 109.661, 108.933, 111.057, 111.055,
112.392, 109.55, 111.042, 110.679, 111.431, 112.06
```

Not surprisingly, the voltage varies slightly and irregularly from day to day, and this variability is visualized in Fig. 4.1.

[1] See, e.g., Problem 4.7.25.

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_4

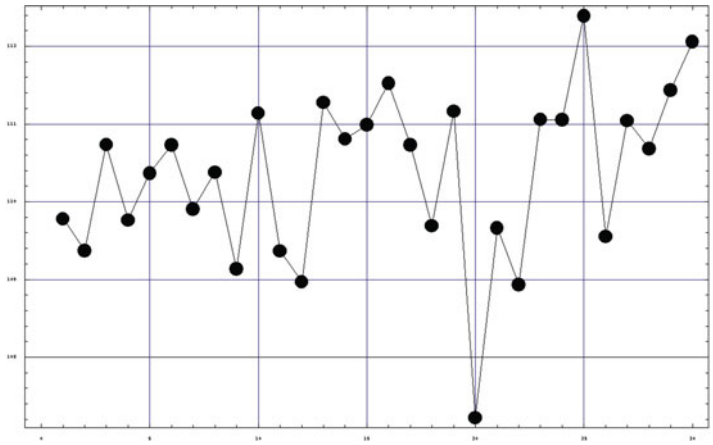


Fig. 4.1 Variability of daily voltage readings on an electrical outlet

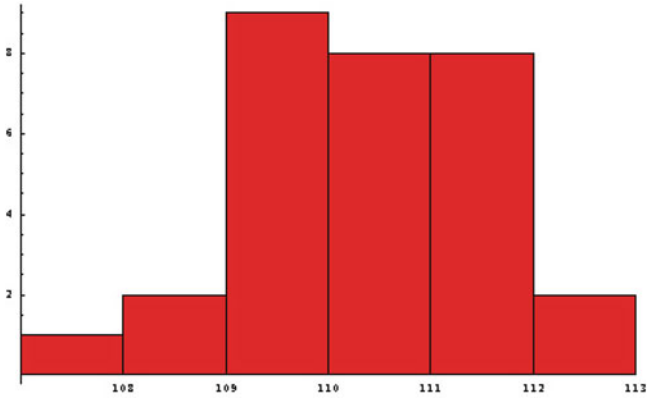


Fig. 4.2 The histogram of daily voltage readings on an electrical outlet

In the presence of such uncertainty he may want to get a better idea of how the voltage values are distributed within its range and he is likely to visualize this information in the form of a histogram shown in Fig. 4.2.

In this chapter we will discuss analytical tools for study of such random quantities. The discrete and continuous random quantities are introduced, but we also show that, in the presence of fractal phenomena, the above classification is not exhaustive.

For the purposes of these lectures, *random quantities* (also called *random variables* in the literature), denoted by capital letters X, Y, etc., will symbolize measurements of experiments with uncertain outcomes.

A random quantity X will be fully characterized by its *cumulative distribution functions (c.d.f.)*, denoted $F_X(x)$, which gives the probability, $\mathbf{P}(X \leq x)$, of the

outcomes of experiment X do not exceeding number x:

$$F_X(x) := \mathbf{P}(X \leq x). \tag{4.1.1}$$

Necessarily,

$$F_X(-\infty) = 0, \qquad F_X(\infty) = 1, \tag{4.1.2}$$

function $F_X(x)$ is nondecreasing,

$$F_X(x) \leq F_X(y), \qquad \text{if} \qquad x < y, \tag{4.1.3}$$

and probability of the measurement being contained in the interval $(a, b]$ is

$$\mathbf{P}(a < X \leq b) = F_X(b) - F_X(a). \tag{4.1.4}$$

If $a < b < c$, we thus have

$$\mathbf{P}(a < X \leq c) = F_X(c) - F_X(a) = [F_X(b) - F_X(a)] + [F_X(c) - F_X(b)]$$

$$= \mathbf{P}(a < X \leq b) + \mathbf{P}(b < X \leq c).$$

This fundamental property of probabilities, called *additivity*, can be extended from disjoint intervals to more general disjoint[2] sets A and B, yielding the formula

$$\mathbf{P}(X \in A \cup B) = \mathbf{P}(X \in A) + \mathbf{P}(X \in B).$$

In other words, probability behaves like the area measure of planar sets.

Discrete Probability Distributions A random quantity X with a discrete probability distribution takes on only (finitely or infinitely many) discrete values, say $x_1, x_2, \ldots$, so that

$$\mathbf{P}(X = x_i) = p_i, \quad i = 1, 2, \ldots, \tag{4.1.5a}$$

where

$$0 < p_i < 1. \qquad \sum p_i = 1. \tag{4.1.5b}$$

In the discrete case, c.d.f.

[2]Recall that sets A and B are called *disjoint* if their intersection is the empty set, i.e., $A \cap B = \emptyset$.

$$F_X(x) = \sum_{i=1}^{\infty} p_i u(x - x_i), \tag{4.1.6}$$

where $u(x)$ is the Unit Step Function. In other words, c.d.f. has jumps of size p_i at locations x_i and is constant at other points of the real line.

Example 4.1.1 (Bernoulli Distribution) In this case the values of X, that is the possible outcomes of the experiment, are assumed to be either 1 or 0 (think about it as a model of an experiment in which "success" or "failure" are the only possible outcomes), with $\mathbf{P}(X = 1) = p > 0, \quad \mathbf{P}(X = 0) = q > 0$, with p, q satisfying condition $p + q = 1$. The c.d.f. of the Bernoulli random quantity is

$$F_X(x) = \begin{cases} 0, & \text{for} x < 0; \\ q = 1 - p, & \text{for} 0 \le x < 1; \\ 1, & \text{for} 1 \le x. \end{cases}$$

The Bernoulli family of distributions has one parameter p which must be a number between 0 and 1. Then $q = 1 - p$ (Fig. 4.3).

Example 4.1.2 (Binomial Distribution) The binomial random quantity X can take values $0, 1, \ldots, n$, with corresponding probabilities

$$p_k = \mathbf{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

where the binomial coefficient is defined by

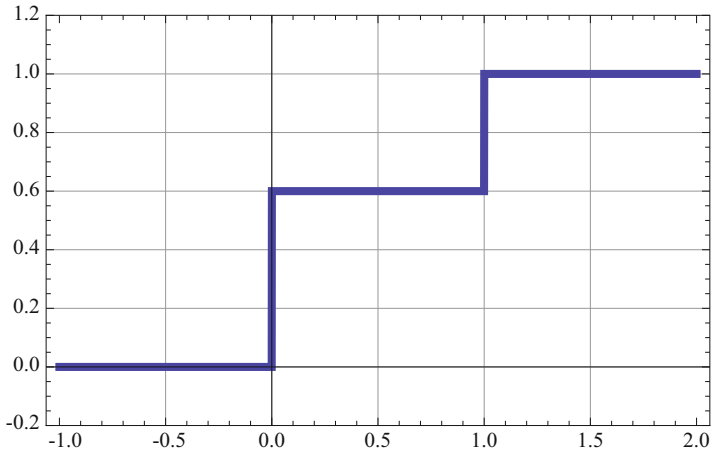


Fig. 4.3 Cumulative distribution function $F_X(x)$ of a Bernoulli random quantity X with parameter $p = 0.4$ has a jump of size $q = 1 - 0.4 = 0.6$ at $x = 0$, and a jump of size $p = 0.4$ at $x = 1$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Recall that the name "binomial coefficient" comes from the elementary *binomial formula*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

familiar in the special cases:

$$(a+b)^2 = a^2 + 2ab + b^2,$$

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,$$

and so on (Fig. 4.4).

Probabilities $p_k = p_k(n, p)$ in the binomial probability distribution are probabilities that exactly k "successes" occur in n independent[3] Bernoulli experiments, each with probability of "success" equal to p.

The normalization condition $\sum_k p_k = 1$ (4.5.1b) is here satisfied because, in view of the above-mentioned binomial formula,

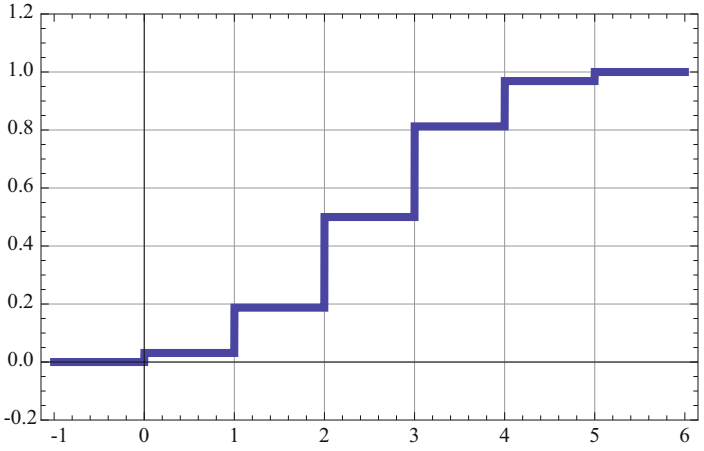


Fig. 4.4 Cumulative distribution function $F_X(x)$ of a binomial random quantity X with parameters $p = 0.5$ and $n = 5$

[3] A rigorous definition of the concept of independence of random quantities will be discussed later on in this chapter.

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1.$$

The binomial family of distributions has two parameters: p, which must be between 0 and 1, and n, which can be an arbitrary positive integer.

Example 4.1.3 (Poisson Distribution) The values of a Poisson random quantity X can be arbitrary nonnegative integers $0, 1, 2, \ldots$, and their probabilities are defined by the formula

$$p_k = \mathbf{P}(X = k) = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \ldots$$

The normalization condition $\sum_k p_k = 1$ is satisfied in this case because of the power series expansion for the exponential function:

$$\sum_{k=0}^{\infty} e^{-\mu} \frac{\mu^k}{k!} = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} e^{\mu} = 1.$$

The family of Poisson distributions has one parameter $\mu > 0$. Poisson random quantities often are used as models of numbers of arrivals of "customers" in queuing systems (an internet web site, a line at the check-out counter, etc.) within a given time interval.

Continuous Distributions A random quantity X is said to have a continuous probability distribution[4] if its c.d.f. $F_X(x)$ can be written as an integral of a certain nonnegative function $f_X(x)$ which traditionally is called the *probability density function* (p.d.f.) of X, that is

$$F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f_X(z)\, dz. \tag{4.1.7}$$

Then, of course, the probability of the random quantity to assume values between a and b is just the integral of the p.d.f. over the interval $[a, b]$, see Fig. 4.5, where $f_X(x)$ was selected to be $(3/5\sqrt{\pi})e^{-x^2} + (2/5\sqrt{\pi})e^{-(x-2)^2}$. Note that in the continuous case it does not matter whether the interval between a and b is open or closed since the probability of the random quantity taking a particular value is always zero. Thus we have

[4]Strictly speaking, c.d.f.s that admit the integral representation (4.1.7), that is have densities, are called *absolutely continuous distributions* as there exist continuous c.d.f.s which do not admit this integral representation, see an example of a singular c.d.f. later in this section, and, e.g., M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity and Chaotic behavior in Engineering and Science*, Birkhäuser-Boston, 1998.

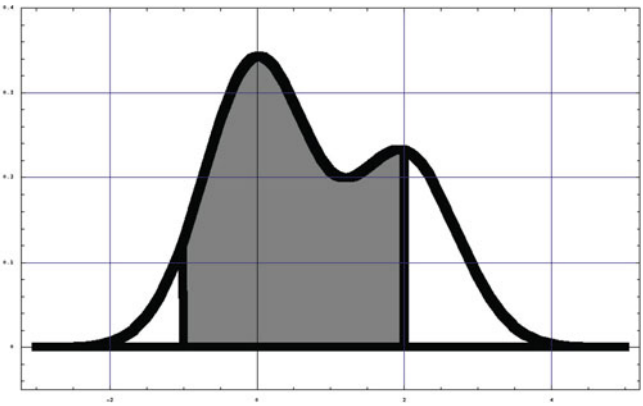


Fig. 4.5 The shaded area under $f_X(x)$, and above the interval $[-1, 2]$ is equal to the probability that a random quantity X with p.d.f. $f_X(x)$ takes values in the interval $[-1, 2]$

$$\mathbf{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(z)\, dz. \tag{4.1.8}$$

Also, necessarily, we have the normalization condition,

$$\int_{-\infty}^{\infty} f_X(x)\, dx = F_X(+\infty) = 1, \tag{4.1.9}$$

and, in view of (4.1.7), and the Fundamental Theorem of Calculus, we can obtain the p.d.f. $f_X(x)$ by differentiation of the c.d.f. $F_X(x)$:

$$\frac{d}{dx} F_X(x) = f_X(x).$$

Example 4.1.4 (Uniform Distribution) The density of a uniformly distributed random quantity X is defined to be a positive constant within a certain interval, say $[c, d]$, and zero outside this interval. Thus, because of the normalization condition (4.1.9),

$$f_X(x) = \begin{cases} (d-c)^{-1}, & \text{for} c \leq x \leq d; \\ 0, & \text{elsewhere.} \end{cases}$$

The family of uniform densities is parametrized by two parameters c and d, with $c < d$.

The c.d.f. of a uniform random quantity is

$$F_X(x) = \begin{cases} 0, & \text{for} x < c; \\ (x-c)/(d-c), & \text{for} c \le x \le d; \\ 1, & \text{for} d \le x. \end{cases}$$

Example 4.1.5 (Exponential Distribution) An exponentially distributed random quantity X has the p.d.f. of the form

$$f_X(x) = \begin{cases} 0, & \text{for} x < 0; \\ e^{-x/\mu}/\mu, & \text{for} x \ge 0. \end{cases}$$

There is one parameter, $\mu > 0$. The c.d.f. in this case is easily computable:

$$F_X(x) = \begin{cases} 0, & \text{for} x < 0; \\ 1 - e^{-x/\mu}, & \text{for} x \ge 0. \end{cases}$$

An exponential p.d.f. and the corresponding c.d.f. are pictured in Fig. 4.5.

Exponential p.d.f.s often appear in applications as probability distributions of random waiting times between Poisson events discussed earlier in this section. For example, under certain simplifying assumptions, it can be proven that the random time intervals between consecutive hits at a web site have an exponential probability distribution. For this reason, exponential p.d.f.s play a crucial role in the analysis of the internet traffic and other queuing networks.

Example 4.1.6 (Gaussian (Normal) Distribution) The density of a Gaussian (also called normal) random quantity X is defined by the formula

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

There are two parameters, μ, which is a real number, and $\sigma > 0$, and this distribution is often denoted $N(\mu, \sigma^2)$ p.d.f. (N for "normal"). The Gaussian c.d.f. is of the form, see Fig. 4.6,

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(z-\mu)^2}{2\sigma^2}}\, dz,$$

but, unfortunately, the integral cannot be expressed in terms of the elementary functions of the variable x. Thus the values of this c.d.f., and the probabilities of a Gaussian random quantity taking values within a given interval, have to be evaluated numerically, using tables (provided at the end of this chapter), or mathematical software such as *Matlab, Maple*, or *Mathematica*; see Example 4.1.6 (continued) below (Fig. 4.7).

However, the normalization condition for the Gaussian p.d.f. can be verified directly analytically by a clever trick that replaces the square of the integral by a

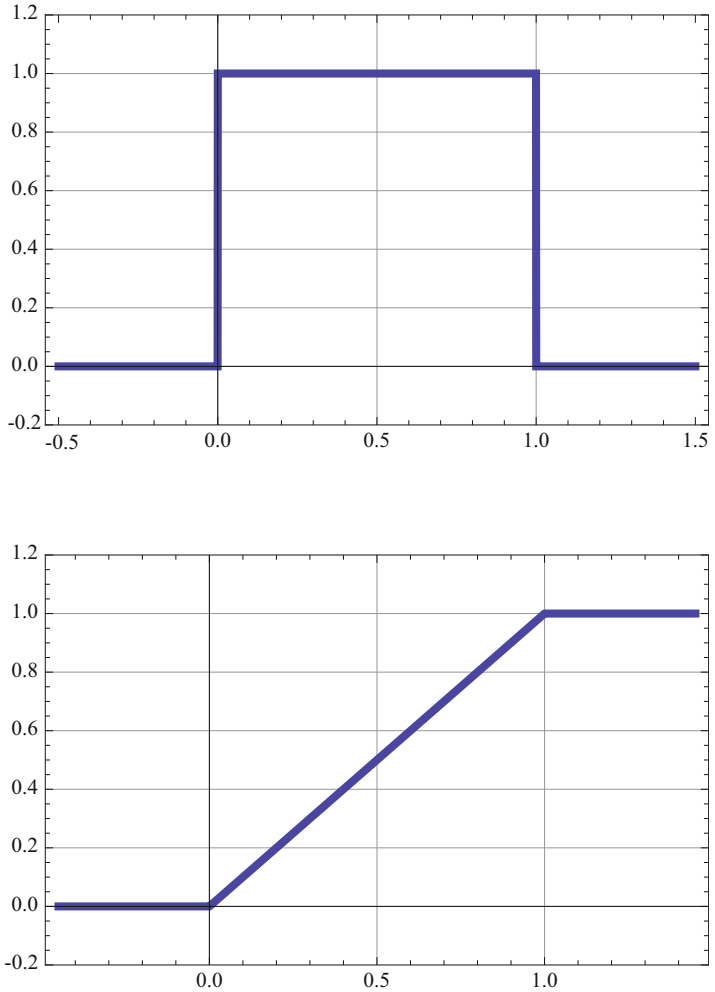


Fig. 4.6 (Top) Probability density function (p.d.f) $f_X(x)$ for a random quantity with values uniformly distributed over the interval [0, 1]. (Bottom) C.d.f. $F_X(x)$ for the same random quantity

double integral which is then evaluated in polar coordinates r, θ. We carry out this calculation in the special case $\mu = 0, \sigma^2 = 1$:

$$\left(\int_{-\infty}^{\infty} f_X(x)\,dx\right)^2 = \int_{-\infty}^{\infty} f_X(x)\,dx \cdot \int_{-\infty}^{\infty} f_X(y)\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x)\cdot f_X(y)\,dx\,dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-x^2-y^2}{2}}\,dx\,dy = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{\frac{-r^2}{2}} r\,dr\,d\theta = 1.$$

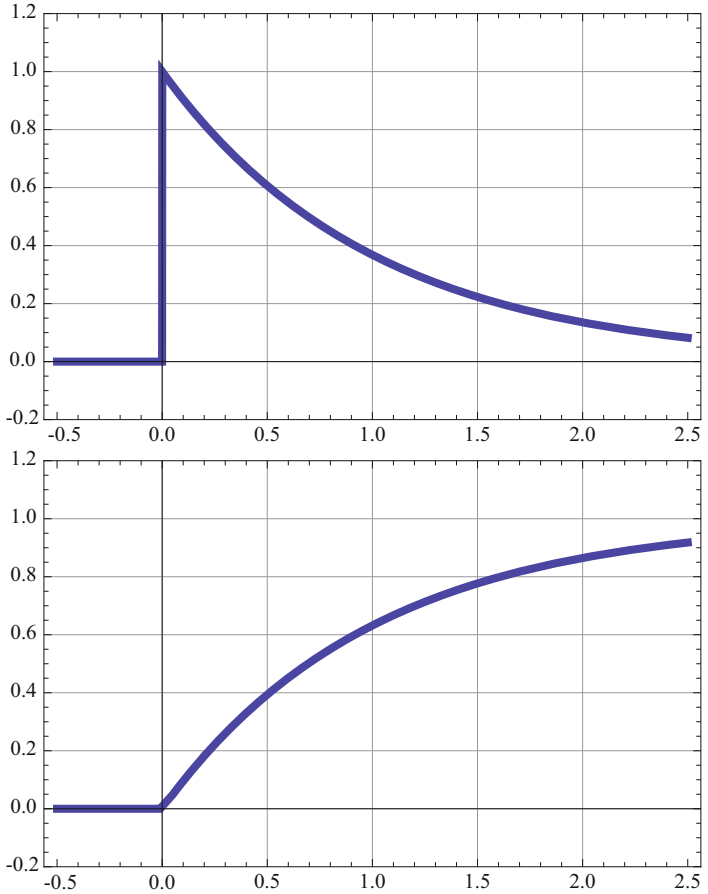


Fig. 4.7 (Top) Probability density function (p.d.f) $f_X(x)$ of an exponentially distributed random quantity with parameter $\mu = 1$. (Bottom) Cumulative distribution function (c.d.f.) $F_X(x)$ for the same random quantity

Example 4.1.6 ((Continued): Calculations of N(0,1) Probabilities) The values of the Gaussian $N(0, 1)$ cumulative distribution, traditionally denoted $\Phi(x)$, are tabulated at the end of this chapter. They are listed only for positive values of variable x, because, in view of the symmetry of the $N(0, 1)$ density, we have

$$\Phi(-x) = 1 - \Phi(x).$$

Thus

$$\mathbf{P}(-1.53 < X < 2.11) = \Phi(2.11) - \Phi(-1.53) = \Phi(2.11) - (1 - \Phi(1.53))$$

$$= 0.9826 - (1 - 0.9370) = 0.9196.$$

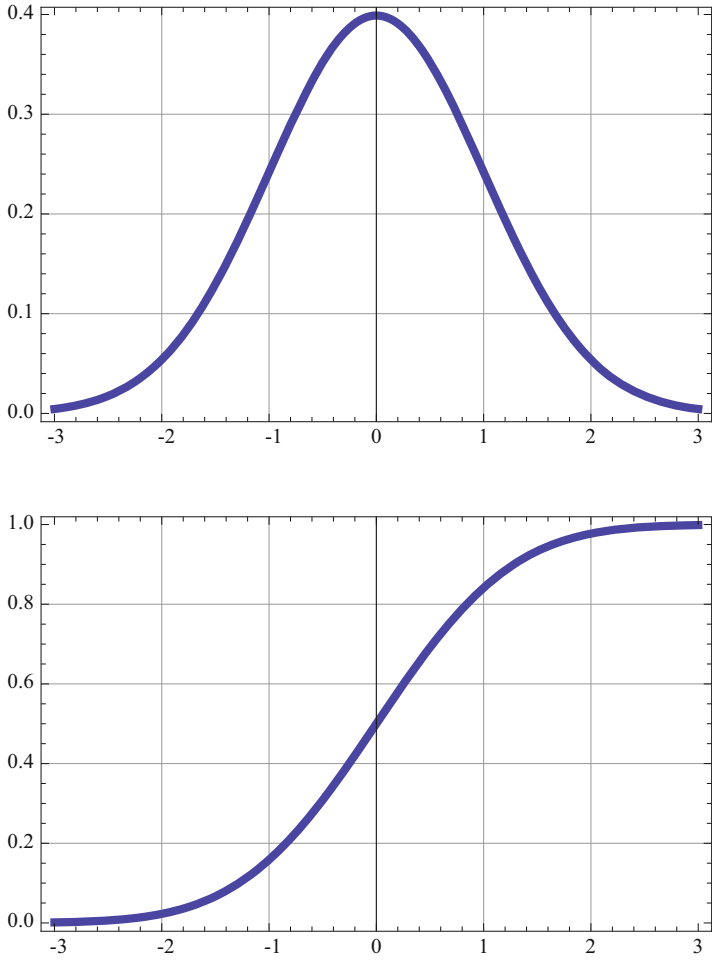


Fig. 4.8 (Top) Probability density function (p.d.f) $f_X(x)$ for a Gaussian random quantity with parameters $\mu = 0,\ \sigma = 1$. (Bottom) Cumulative distribution function (c.d.f.) $F_X(x)$ for the same random quantity

This leaves unanswered the question of how to calculate the general $N(\mu, \sigma^2)$ probabilities. For a solution, see Example 4.1.9 (Fig. 4.8).

Remark 4.1.1 (Importance of the Gaussian Distribution) The fundamental importance of the Gaussian probability distribution stems from the *Central Limit Theorem* (see Sect. 4.5) which asserts that, for a large number of independent repetitions of experiments with random outcomes, the fluctuations (errors) of the outcomes around their mean value have, approximately, a Gaussian p.d.f. At a more fundamental

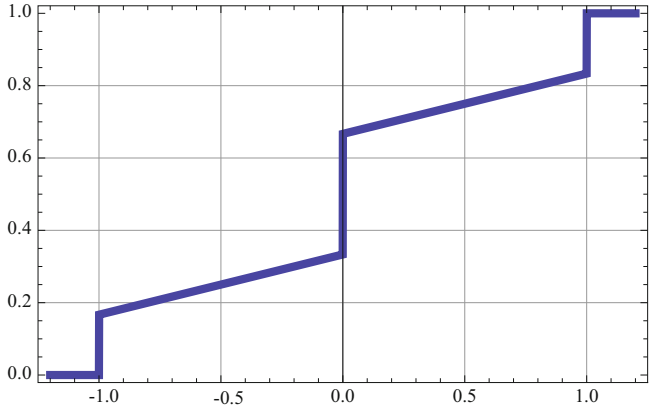


Fig. 4.9 Cumulative distribution function (c.d.f.) $F_X(x)$ of mixed type described by formula (4.1.10). This distribution has both discrete and continuous components

level, this result is related to the invariance of Gaussian densities under the Fourier transformation, see Example 2.4.1.

Mixed and Singular Distributions A random quantity is said to have a c.d.f. of *mixed type* if it has both discrete and continuous components. The c.d.f. thus has both discrete jumps, perhaps infinitely (but countably) many, as well as points of continuous increase where its derivative is well defined. For example, the c.d.f.

$$F_X(x) = \begin{cases} 0, & \text{for} x < -1; \\ x/6 + 2/6, & \text{for} -1 \le x < 0; \\ x/6 + 4/6, & \text{for} 0 \le x < 1; \\ 1, & \text{for} 1 \le x \end{cases} \tag{4.1.10}$$

represents a random quantity X which is uniformly distributed on the intervals $(-1, 0) \cup (0, 1)$ with probability 1/3, but also takes the discrete values $-1, 0, 1$, with positive probabilities equal to the jump sizes of the c.d.f at those points. Think here about a cloud of particles randomly, and uniformly distributed over the intervals $(-1, 0) \cup (0, 1)$, with absorbing boundaries at $x = \pm 1$, and a sticky trap at $x = 0$; the probability of finding a particle at those discrete points is positive, 1/6 at $x = \pm 1$ and 1/3 at $x = 0$ (Fig. 4.9).

Thus, for example,

$$\mathbf{P}\left(-\frac{1}{2} < X \le \frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) - F_X\left(-\frac{1}{2}\right) = \left(\frac{1}{12} + \frac{4}{6}\right) - \left(-\frac{1}{12} + \frac{2}{6}\right) = \frac{1}{2},$$

and

$$\mathbf{P}(X=0)=\lim_{\epsilon\to 0}\mathbf{P}(-\epsilon<X\le\epsilon)=\lim_{\epsilon\to 0}(F_X(\epsilon)-F_X(-\epsilon))$$

$$=\lim_{\epsilon\to 0}[(\epsilon/6+4/6)-(-\epsilon/6+2/6)]=1/3.$$

Similarly,

$$\mathbf{P}(X=-1)=1/6,\quad \mathbf{P}(X=0)=2/6,\quad \mathbf{P}(X=1)=1/6.$$

Remark 4.1.2 (Mixture of Gaussian p.d.f.s.) The reader will notice that the example of a p.d.f. which appeared in Fig. 4.5 is a mixture of two Gaussian p.d.f.s.

It is tempting to venture a guess that all c.d.f.s have to be either discrete, continuous, or of mixed type. This, however, is not the case.

The limit of the so-called "devil's staircase" c.d.f.s shown in Fig. 4.10 is an example of a c.d.f. which, although continuous, and differentiable "almost everywhere," does not have a p.d.f.

Observe that inside the interval [0,1] its derivative is 0 on the union of the infinite family of disjoint intervals whose lengths add up to 1. Indeed, as is clear from the construction displayed in Fig. 4.10, this set has the linear measure

$$\lim_{n\to\infty}\left(\frac{1}{3}+2\cdot\frac{1}{3^2}++\ldots+2^{n-1}\cdot\frac{1}{3^n}\right)=\frac{1}{3}\sum_{i=0}^{\infty}\left(\frac{2}{3}\right)^i=\frac{1}{3}\cdot\frac{1}{1-2/3}=1,$$

in view of the formula for the sum of a geometric series. Thus integration of this derivative cannot possibly give a c.d.f. which must grow from 0 to 1. Distributions of this type are called *singular* and they arise in studies of fractal phenomena. One can prove that the set of points of increase of the limit "devil's staircase," i.e., the set of points on which the probability is concentrated, has a fractional dimension equal to $\ln 2/\ln 3=0.6309\ldots$[5]

Distributions of Functions of Random Quantities One often measures random quantities through devices that distort the original quantity X to produce a new random quantity, say, $Y=g(X)$, and the natural question is how the c.d.f. $F_X(x)$ of X is affected by such a transformation. In other words, the question is: Can $F_Y(y)$ be expressed in terms of g and $F_X(x)$? In the case when the transforming function $g(x)$ is *monotonically increasing* the answer is simple:

$$F_{g(X)}(y)=\mathbf{P}(g(X)\le y)=\mathbf{P}\Big(X\le g^{-1}(y)\Big)=F_X(g^{-1}(y)),\tag{4.1.11}$$

[5]See, for example, M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity and Chaotic Behavior in Engineering and Science*, Birkhäuser-Boston, 1998.

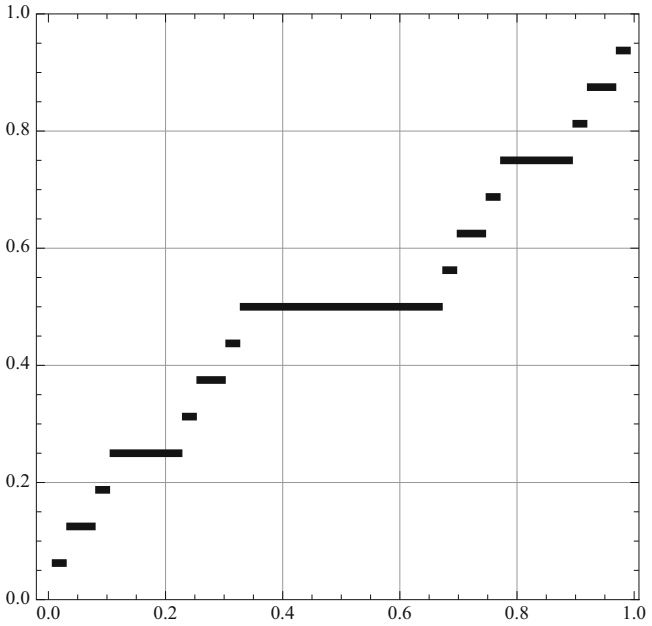


Fig. 4.10 The construction of the singular "devil's staircase" c.d.f. $F_X(x)$. It continuously grows from 0, at $x = 0$, to 1, at $x = 1$, and yet it has no density; its derivative is equal to 0 on disjoint intervals whose lengths add up to 1

where $g^{-1}(y)$ is the inverse function of $g(x)$, that is $g^{-1}(g(x)) = x$, or, equivalently, if $y = g(x)$, then $x = g^{-1}(y)$.

Remembering the chain rule of the elementary calculus, and the formula for the derivative of the inverse function $g^{-1}(y)$, we also immediately obtain, in the case of monotonically increasing $g(x)$, the *expression of the p.d.f. of* $Y = g(X)$ *in terms of the p.d.f. of* X *itself:*

$$f_{g(X)}(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \cdot \frac{1}{g'(g^{-1}(y))}. \tag{4.1.12}$$

Example 4.1.7 (Linear Transformation of a Standard Gaussian Random Quantity) Recall that a Gaussian random quantity X is called *standard* (or, $N(0, 1)$) if its p.d.f. is of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}.$$

It is a special case of the general Gaussian p.d.f. introduced in Example 4.1.6, with parameters μ and σ specified to be 0 and 1, respectively. Consider now a new random quantity Y obtained from X by a linear transformation

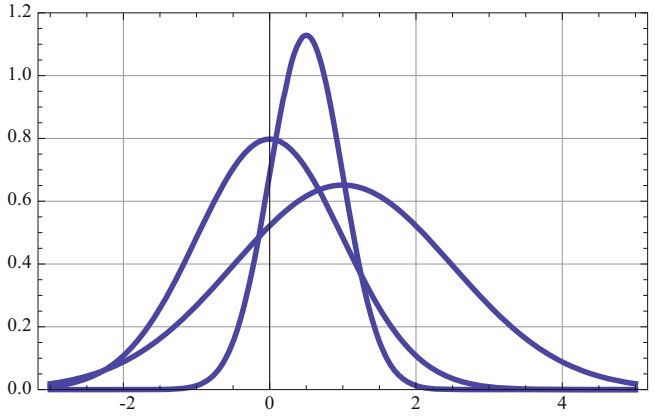


Fig. 4.11 Probability density functions of $N(0, 1)$, $N(0.5, 0.25)$, and $N(1, 2.25)$ random quantities (from left to right)

$$Y = aX + b, \quad a > 0.$$

Think about this transformation as representing the change in units of measurement and the choice of the origin (like changing the temperature measurements from degrees Celsius to Fahrenheit: if X represents temperature measurements in degrees Celsius, then $Y = 1.8 \cdot X + 32$ gives the same measurements in degrees Fahrenheit).

The transforming function in this case, $y = g(x) = ax + b$, is monotonically increasing, and

$$g'(x) = a, \qquad \text{and} \qquad g^{-1}(y) = (y - b)/a.$$

Formula (4.1.12) gives now the following expression for the p.d.f. of Y:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-((y-b)/a)^2}{2}} \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi a^2}} e^{\frac{-(y-b)^2}{2a^2}}.$$

The conclusion is that the transformed random quantity Y has also a Gaussian p.d.f., but with parameters $\mu = b$ and $\sigma^2 = a^2$; in other words, Y is $N(b, a^2)$-distributed. Several examples of Gaussian p.d.f.s are shown in Fig. 4.11.

Example 4.1.7 ((Continued): Calculation of General $N(\mu, \sigma^2)$ Probabilities) The relationship established in Example 4.1.7 permits utilization of tables of the $N(0, 1)$ distributions supplied at the end of this chapter to calculate $N(\mu, \sigma^2)$ probabilities for arbitrary values of parameters μ and $\sigma^2 > 0$. Indeed, if a random quantity Y has the $N(\mu, \sigma^2)$ distribution, then it is of the form

$$Y = \sigma X + \mu,$$

where X has the $N(0, 1)$ distribution with c.d.f. $F_X(x) = \Phi(x)$, so that

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(\sigma X + \mu \le y)$$

$$= \mathbf{P}(X \le (y - \mu)/\sigma) = \Phi\left(\frac{y - \mu}{\sigma}\right), \tag{4.1.13}$$

and the values of the latter can be taken from Table 4.1. For example, if Y is Gaussian with parameters $\sigma = 1.8$ and $\mu = 32$, then

$$\mathbf{P}(30 < Y < 36) = \Phi\left(\frac{36 - 32}{1.8}\right) - \Phi\left(\frac{30 - 32}{1.8}\right)$$

$$= \Phi(2.22) - (1 - \Phi(-1.11)) = 0.9868 - (1 - 0.8665) = 0.8533.$$

In the next two examples we will consider the quadratic transformation $Y = X^2/2$ corresponding to calculation of the (random) kinetic energy[6] Y of an object of unit mass $m = 1$, traveling with random velocity X.

Example 4.1.8 (Kinetic Energy of a Unit Mass Traveling with Random, Exponentially Distributed Velocity) Suppose that the random quantity X has an exponential c.d.f. and p.d.f. given in Example 4.1.5, with parameter $\mu = 1$. It is transformed by a quadratic "device" $g(x) = x^2/2$ into the random quantity $Y = X^2/2$. Note that the exponential p.d.f. is concentrated on the positive half-line and that the transforming function $g(x)$ is monotonically increasing in that domain. Then the c.d.f. $F_Y(y) = 0$, for $y \le 0$, and, for $y > 0$, we can repeat the argument from formula (4.1.11) to obtain

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X^2/2 \le y)$$

$$= \mathbf{P}(X \le \sqrt{2y}) = F_X(\sqrt{2y}) = 1 - e^{-\sqrt{2y}}.$$

Similarly, using (4.1.12), one gets the p.d.f. of $X^2/2$:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0, & \text{for } y \le 0; \\ e^{-\sqrt{2y}}/\sqrt{2y}), & \text{for } y > 0. \end{cases}$$

Note that this p.d.f. has a singularity at the origin; indeed, $f_Y(y) \uparrow +\infty$ as $y \downarrow 0+$. Observe, however, that the singularity does not affect the p.d.f. normalization condition $\int_{-\infty}^{\infty} f_Y(y)\, dy = 1$.

[6] Recall that an object of mass m traveling with velocity v has kinetic energy $E = mv^2/2$.

Table 4.1 Gaussian $N(0, 1)$ c.d.f.: $\Phi(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-x^2/2}\, dx$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5395	0.5438	.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6296	0.6331	0.6366	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6884	0.6879
0.5	0.6915	0.6956	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7857	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8075	0.8106	0.8133
0.9	0.8195	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8503	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8613	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8796	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8977	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9359	0.9370	0.9382	0.9309	0.9404	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9606	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9666	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9891	0.9861	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.4	0.9918	0.9820	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

If the transforming function $y = g(x)$ is *not monotonically increasing* (or, decreasing, see Problem 4.7.26, and Sects. 8.1 and 8.2) over the range of the random quantity X (as, for example, $g(x) = x^2$ in the case when X takes both positive and negative values), then a more subtle analysis is required to find the p.d.f. of the random quantity $Y = g(X)$.

Example 4.1.9 (Square of a Standard Gaussian Random Quantity) Assume that X has the standard $N(0, 1)$ Gaussian p.d.f. and that the transforming function is quadratic: $y = g(x) = x^2$. The quadratic function is monotonically increasing only over the positive half-line; it is monotonically decreasing over the negative half-line). So, we have to proceed with caution and start with an analysis of the c.d.f. of $Y = X^2$ taking advantage of the symmetry of the Gaussian p.d.f.:

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X^2 \le y)$$
$$= 2\mathbf{P}(0 \le X \le \sqrt{y}) = 2(F_X(\sqrt{y}) - 1/2).$$

The above formula, obviously, is valid only for $y > 0$; on the negative half-line the c.d.f. of $Y = X^2$ vanishes. Thus the p.d.f. of $Y = X^2$ is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0, & \text{for } y \le 0; \\ e^{-y/2}/(\sqrt{2\pi y}), & \text{for } y > 0. \end{cases}$$

This p.d.f. is traditionally called the *chi-square* probability density function. We'll see its importance in Sect. 4.6, where it plays the central role in the statistical parameter estimation problems.

Random Quantities as Functions on a Sample Space For those who insist on mathematical precision, the above introduction of random quantities via their probability distributions should be preceded by their formal definition as functions on a sample space. This approach had been pioneered by A.N. Kolmogorov[7] and it has become a commonly accepted, mainstream approach to the mathematical probability theory.

The definition starts with an introduction of the triple $(\Omega, \mathcal{B}, \mathbf{P})$, where the *sample space* Ω is an arbitrary set[8] consisting of *sample points* ω. They should be thought of as labels for different (not necessarily numerical) outcomes of a random experiment being modeled. The *field* $\mathcal{B}$ consists of subsets of the sample space Ω which are called *random events*. To make the logical operations (such as "not," "or," and "and") on random events possible it is assumed that $\mathcal{B}$ contains the whole sample space Ω, and the empty set $\emptyset$, and is closed under complements, unions, and intersections. In other words, one imposes on $\mathcal{B}$ the following *axioms*:

A1.1. $\Omega, \emptyset \in \mathcal{B}$;
A1.2. If $B \in \mathcal{B}$, then its complement $\Omega \setminus B \in \mathcal{B}$;
A1.3. If $A, B \in \mathcal{B}$, then $A \cup B \in \mathcal{B}$ and $A \cap B \in \mathcal{B}$.

[7] See his fundamental *Grundbegriffe der Wahrscheinlichkeisrechnung*, Berlin 1933, but also an earlier work in the same direction by H. Steinhaus, *Studia Mathematica*, 1923, and Bibliographical Comments at the end of this volume.

[8] Without loss of generality one can always take as Ω the unit interval [0, 1], see Remark 4.1.2.

The *probability measure* is then defined as a function $\mathbf{P} : \mathcal{B} \mapsto [0, 1]$, assigning to any random event B a real number between 0 and 1, so that it is normalized to 1 on the whole sample space, and is additive on mutually exclusive (disjoint) random events. In other words one imposes on the probability measure the following axioms:

A2.1. $\mathbf{P}(\Omega) = 1$ (normalization);
A2.2. If $A \cap B \in \emptyset$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ (additivity).

Finally, a *random quantity (variable)* X is any function on the sample space Ω which assigns to each sample point ω (that is to each outcome of a random experiment) a real number $X(\omega)$ in such a way that determining probabilities of $X(\omega)$ taking values in any given interval on the real line is possible. In other words, one demands that the function, $X : \Omega \mapsto \mathbf{R}$, is *measurable*, i.e., it satisfies the following axiom:

A3.1. For each $a, b \in \mathbf{R}$, the set of sample points $\{\omega : a < X(\omega) \leq b\} \in \mathcal{B}$.

The consequence is that $B = \{\omega : a < X(\omega) \leq b\}$ is always a (measurable) random event and the probability $\mathbf{P}(B)$ thereof is well defined. This permits now an introduction of the cumulative distribution function of the random quantity X (and brings us back to the beginning of Section 4.2) via the formula

$$F_X(x) = \mathbf{P}(\{\omega : -\infty < X(\omega) \leq x\}), \qquad x \in \mathbf{R}.$$

To permit limit operations on random events and random quantities one usually extends the above axioms to guarantee that infinite unions are permitted in axioms **A1.3** and **A2.2**. A wide spectrum of examples of sample spaces can be encountered in research practice; we provide three, the first very simple, and the third, rather complex.

Example 4.1.10 (Coin Toss—A Small Sample Space) In this case the outcomes can be labeled H (heads), and T (tails), and the sample space is $\Omega = \{H, T\}$ has only two sample points, H and T. The field of random events can be taken to be $\mathcal{B} = \{\emptyset, \{H\}, \{T\}, \Omega\}$. For any number $p \in [0, 1]$, the probability measure $\mathbf{P}$ on all random events in $\mathcal{B}$ can now be defined as follows:

$$\mathbf{P}(\emptyset) = 0, \quad \mathbf{P}(H) = p, \quad \mathbf{P}(T) = 1 - p, \quad \mathbf{P}(\Omega) = 1.$$

Now one can define a variety of random quantities on $(\Omega, \mathcal{B}, \mathbf{P})$. If in the game you are playing one wins \$1 if heads come up and nothing if tails come up, then the corresponding random quantity X is a function on Ω defined by the equalities,

$$X(H) = 1, \qquad X(T) = 0,$$

and its probability distribution is

$$\mathbf{P}(\{\omega : X(\omega) = 1\}) = p, \qquad \mathbf{P}(\{\omega : X(\omega) = 0\}) = 1 - p.$$

However, if in the game you are playing one wins one dollar if heads come up and one loses one dollar if tails come up, then the corresponding random quantity X is a function on Ω defined by the equalities,

$$X(H) = +1, \qquad X(T) = -1,$$

and its probability distribution is

$$\mathbf{P}(\{\omega : X(\omega) = +1\}) = p, \qquad \mathbf{P}(\{\omega : X(\omega) = -1\}) = 1 - p.$$

Example 4.1.11 (Coin Toss—A Larger Sample Space) The above "natural" choice of the "minimal" sample space is not unique. For example, one can choose $\Omega = [0, 1]$ with $\mathbf{P}$ being the length measure of the subsets of the unit interval. Then take

$$X(\omega) = \begin{cases} 0, & \text{for } \omega \in [0, 1 - p]; \\ 1, & \text{for } \omega \in (1 - p, 1]. \end{cases}$$

Then, obviously, $\mathbf{P}(\{\omega : X(\omega) = 0\}) = 1 - p$, and $\mathbf{P}(\{\omega : X(\omega) = 1\}) = p$.

Example 4.1.12 (Gas of Particles—A Large Sample Space) Consider a gas consisting of $6 \cdot 10^{23}$ (Avogadro number) of particles (say, of mass 1) moving in $\mathbf{R}^3$ according to the Newtonian mechanics. The sample space consists of all possible configurations (states) of the gas described by particles' positions (x^1, x^2, x^3) and velocities (v^1, v^2, v^3). Hence, each sample point

$$\omega = (x_1^1, x_1^2, x_1^3, v_1^1, v_1^2, v_1^3, \ldots, x_N^1, x_N^2, x_N^3, v_N^1, v_N^2, v_N^3)$$

is a $6 \cdot 6 \cdot 10^{23}$-dimensional vector, and the sample space is of the same huge dimension:

$$\Omega = \mathbf{R}^{6 \cdot 6 \cdot 10^{23}}.$$

The field $\mathcal{B}$ of random events here is also huge and consists of all the subsets of Ω that are defined by imposing upper and lower bounds on the components of the positions and velocities of all $6 \cdot 10^{23}$ particles.

Various probability measures on $\mathcal{B}$ can then be defined. In statistical mechanics the standard way to define it is by assigning energy $E(\omega)$ to each configuration ω and then demanding that the probability (fraction of all particles) of the system being in state ω is proportional to $\exp(-\beta)E_\omega$. The resulting probability measure on Ω is called the *Gibbs-Boltzmann measure*.

If the random quantity of interest is just the kinetic energy (temperature) of the configuration,

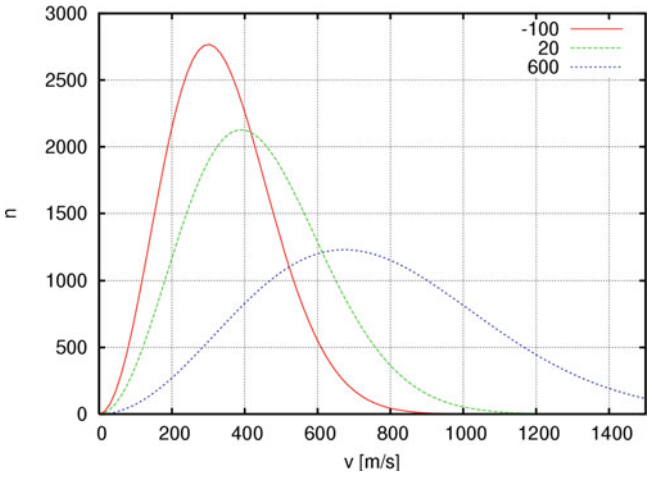


Fig. 4.12 Distribution of particle speed for 10^6 oxygen particles at -100, 20, and 600 degrees Celsius (left to right). Speed distribution was derived from Maxwell-Boltzmann distribution (from http://en.wikipedia.org/wiki/Maxwell-Boltzmann_statistics)

$$E_\omega = \frac{1}{2}\sum_{i=1}^{N}\left[(v_i^1)^2 + (v_i^2)^2 + (v_i^3)^2\right],$$

then the above Gibbs-Boltzman distribution correctly gives the classical Maxwell probability density function of gas particle speeds $s = \sqrt{(v^1)^2 + (v^2)^2 + (v^3)^2}$:

$$f_S(s) = (2/\pi)^{1/2}(kT)^{-3/2}s^2\exp(-s^2/2kT), \qquad s \geq 0,$$

where k is the Boltzmann constant, and T is the absolute temperature. To accommodate different types of particles additional parameters are usually included in the above formula, see Fig. 4.12 for an example of plots of p.d.f.s of particle speeds for oxygen particles.[9]

Remark 4.1.3 (Unit Interval as a Universal Sample Space) For any random quantity X one can always choose the unit interval $[0, 1]$ as the underlying sample space Ω (although, not always this is the most natural selection) with sample points $\omega \in \Omega$ being just numbers between 0 and 1. Indeed, equip Ω with the Lebesgue (length) measure as the underlying probability $\mathbf{P}$. That is, if $A = [a, b] \subset \Omega = [0, 1]$, then we set

[9]See, e.g., Carter, Ashley H., "Classical and Statistical Thermodynamics", Prentice-Hall, Inc., 2001, New Jersey.

$$\mathbf{P}(A) = \mathbf{P}(\{\omega : a \leq \omega \leq b\}) = b - a,$$

and let $\omega = F_X(x)$ be the cumulative distribution function of X. Since, the above c.d.f. is not necessarily a strictly increasing function we will define its inverse $F_X^{-1}(\omega)$, $\omega \in [0, 1]$, as the reflection in the diagonal, $x = \omega$, in the (x, ω)-plane, of the plot of the p.d.f. $F_X(x)$. More precisely, we uniquely define the (generalized) inverse of the c.d.f. by the equality,[10]

$$F_X^{-1}(\omega) = \min\{x : F_X(x) \geq \omega\}.$$

Of course, if $F_X(x)$ is strictly increasing, then the above definition yields the usual inverse function satisfying the conditions,

$$F_X^{-1}(F_X(x)) = x, \qquad \text{and} \qquad F_X(F_X^{-1}(\omega)) = \omega.$$

In the next step, define

$$X(\omega) = F_X^{-1}(\omega), \qquad \omega \in \Omega = [0, 1].$$

Clearly, $X(\omega)$ defined in this fashion has the correct c.d.f.,

$$\mathbf{P}(\{\omega \in [0, 1] : X(\omega) \leq x\}) = \mathbf{P}(\{\omega \in [0, 1] : F_X^{-1}(\omega) \leq x\})$$

$$= \mathbf{P}(\{\omega : 0 \leq \omega \leq F_X(x)\}) = F_X(x).$$

For instance, Example 4.1.11, defines the Bernoulli random quantity as a function on [0, 1] via the above "generalized" inverse of the Bernoulli c.d.f. shown in Fig. 4.3. As an example of the strictly increasing c.d.f., we can take the Cauchy random quantity X with the c.d.f.,

$$F_X(x) = \frac{1}{\pi}\left(\arctan(x) + \frac{\pi}{2}\right),$$

which continuously increases from 0 to 1 as X ranges from $-\infty$ to $+\infty$. Solving the equation $F_X(x) = \omega$ yields the inverse function

$$F_X^{-1}(\omega) = \tan\left(\pi\omega - \frac{\pi}{2}\right) = X(\omega)$$

and a representation of the Cauchy random quantity as a function on the unit interval. Both the Cauchy c.d.f. and its inverse are shown in Fig. 4.13.

[10] Traditionally the inverse of the c.d.f. of a random quantity is called its quantile function, see Sect. 4.6.

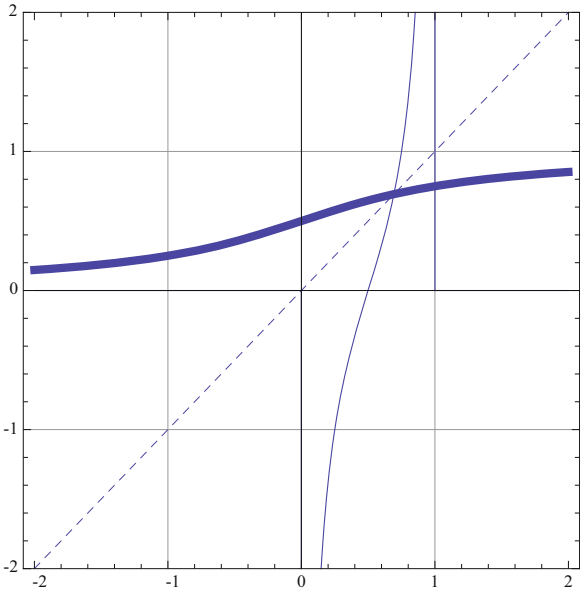


Fig. 4.13 Cumulative distribution function, $F_X(x)$, of a Cauchy random quantity (thick line), and its inverse, $X(\omega) = F_X^{-1}(\omega)$ (thin line), providing a representation of the Cauchy random quantity $X(\omega)$ as a function on the sample space $\Omega = [0, 1]$

4.2 Expectations and Moments of Random Quantities

The *expected value* or, in brief, the *expectation* of a random quantity X is its mean value (or, for a physics-minded reader, the center of the probability mass) with different values of X given weights equal to their probabilities. The expectation of X will be denoted $\mathbf{E}X$, or $\mathbf{E}(X)$, whichever is more convenient. So, for a discrete random quantity X with $\mathbf{P}(X = x_i) = p_i$, $\sum_i p_i = 1$, we have

$$\mathbf{E}X = \sum_i x_i \, p_i, \tag{4.2.1}$$

and for an (absolutely) continuous random quantity with probability density $f_X(x)$

$$\mathbf{E}X = \int_{-\infty}^{\infty} x f_X(x) \, dx. \tag{4.2.2}$$

More generally, one can consider the expectation of a function $g(X)$ of a random quantity X which is defined by the formulas,

$$\mathbf{E}[g(X)] = \begin{cases} \sum_i g(x_i) p_i, & \text{in the discrete case;} \\ \int_{-\infty}^{\infty} g(x) f_X(x) \, dx, & \text{in the continuous case.} \end{cases} \tag{4.2.3}$$

In particular, if $g(x) = x^k$, $k = 1, 2, \ldots$, then the numbers

$$\mu_k(X) = \mathbf{E}g(X) = \mathbf{E}X^k = \begin{cases} \sum_i x_i^k p_i, & \text{inthediscretecase;} \\ \int_{-\infty}^{\infty} x^k f_X(x)\,dx, & \text{inthecontinuouscase} \end{cases} \tag{4.2.4}$$

are called k-th *moments* of X. The first moment $\mu_1 = \mu_1(X)$ is just the expectation of $\mathbf{E}X$ of the random quantity X.

If $g(x) = |x|^\alpha,\ -\infty < \alpha < \infty$, then

$$m_k(X) = \mathbf{E}|X|^\alpha$$

are called α-th *absolute moments*, and for $g(x) = |x - \mu_1|^\alpha$, the numbers

$$\mathbf{E}|X - \mu_1|^\alpha = \mathbf{E}|X - \mathbf{E}X|^\alpha,$$

are called α-th *absolute central moments* of X. The latter measure the mean value of the α-th power of the deviation of the random quantity X from its expectation $\mathbf{E}X$. In other words, they provide a family of parameters which measure how the values of the random quantity are spread around its "center of mass."

In the special case $\alpha = 2$, the second central moment

$$\mathbf{E}(X - \mathbf{E}X)^2 = \begin{cases} \sum_i (x_i - \mu_1)^2 p_i, & \text{indiscretecase;} \\ \int_{-\infty}^{\infty} (x - \mu_1)^2 f_X(x)\,dx, & \text{incontinuouscase} \end{cases} \tag{4.2.5}$$

is called the *variance* of the random quantity X and denoted $\mathrm{Var}(X)$. Again, for a physically minded reader, it is worth noticing that the variance is just the moment of inertia of the probability mass distribution. A simple calculation gives the formula

$$\mathrm{Var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2, \tag{4.2.6}$$

which is sometimes simpler computationally then (4.2.5); the variance is thus the difference between the second moment (sometimes called also *mean square* of a random quantity) and the square of the first moment. This rule is then often phrased: *variance is equal to the mean square minus the squared mean*.

Example 4.2.1 (Moments of the Bernoulli Distribution) For the Bernoulli random quantity X, with distribution given in Example 4.1.1, all the moments are

$$\mu_k(X) = 1^k \cdot p + 0^k \cdot (1 - p) = p,$$

and the variance is

$$\mathrm{Var}(X) = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p).$$

Example 4.2.2 (Mean and Variance of the Uniform Distribution) A uniformly distributed random quantity X (see Example 4.1.4) has expectation

$$\mathbf{E}X = \int_c^d x \frac{1}{d-c}\, dx = \frac{d+c}{2}.$$

Its variance is

$$\mathrm{Var}\,(X) = \int_c^d \left(x - \frac{d+c}{2}\right)^2 \frac{1}{d-c}\, dx = \frac{(d-c)^2}{12}.$$

Notice that the *expectation* or *expected value* $\mathbf{E}X$ *of a random quantity* X *scales linearly*, that is,

$$\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X), \qquad -\infty < \alpha < \infty, \tag{4.2.7}$$

so that the change of scale of the measurements affects the expectations proportionally: if, for example, X is measured in meters, then $\mathbf{E}X$ is also measured in meters. Indeed, in the continuous case,

$$\mathbf{E}(\alpha X) = \int_{-\infty}^{\infty} (\alpha x) f_X(x)\, dx = \alpha \int_{-\infty}^{\infty} x f_X(x)\, dx = \alpha \mathbf{E}(X),$$

and the discrete case can be verified in an analogous fashion.

On the other hand, the *variance* $\mathrm{Var}(X)$ *has the quadratic scaling*

$$\mathrm{Var}(\alpha X) = \alpha^2 \mathrm{Var}(X). \tag{4.2.8}$$

This follows immediately from the linear scaling of the expectations (4.2.7) and the formula (4.2.6). Thus the mean-square deviation has a somewhat unpleasant nonlinear property which implies that if X is measured, say, in meters, then its variance is measured in meters square.

For this reason, one often considers the *standard deviation* $\mathrm{Std}(X)$ of random quantity X which is defined as the square root of the variance:

$$\mathrm{Std}(X) = \sqrt{\mathrm{Var}\,(X)}. \tag{4.2.9}$$

The standard deviation scales linearly, at least for positive α, since

$$\mathrm{Std}(\alpha X) = |\alpha|\, \mathrm{Std}(X), \qquad -\infty < \alpha < \infty. \tag{4.2.10}$$

This means that changing the measurement units affects the standard deviation proportionately as well. If a random quantity is measured in meters, then its standard deviation is also measured in meters.

Additionally, observe that the *expectation is additive with respect to constants*, that is, for any constant β, $-\infty < \beta < \infty$,

$$\mathbf{E}(X + \beta) = \mathbf{E}(X) + \beta. \tag{4.2.11}$$

The verification is again immediate and follows from the additivity property of the integrals (or, in the discrete case, sums):

$$\mathbf{E}(X+\beta) = \int_{-\infty}^{\infty} (x+\beta) f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} \beta f_X(x)\, dx = \mathbf{E}(X) + \beta,$$

because $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

Finally, the *variance is invariant under translations*, that is, for any constant β, $-\infty < \beta < \infty$,

$$\mathrm{Var}(X+\beta) = \mathrm{Var}(X). \tag{4.2.12}$$

Indeed,

$$\mathrm{Var}(X+\beta) = \mathbf{E}\Big((X+\beta) - \mathbf{E}(X+\beta)\Big)^2 = \mathbf{E}\Big(X+\beta - \mathbf{E}(X) - \beta\Big)^2 = \mathrm{Var}(X).$$

The above properties indicate that any random quantity X can be *standardized* by, first, centering it, and then by rescaling it, so that the standardized random quantity has expectation 0 and variance 1. Indeed, if

$$Z = \frac{X - \mathbf{E}X}{\mathrm{Std}(X)}, \tag{4.2.13}$$

then it immediately follows from (4.2.10) and (4.2.11) that $\mathbf{E}Z = 0$ and $\mathrm{Var}(Z) = 1$.

Example 4.2.3 (Mean and Variance of the Gaussian Distribution) Let us begin with a random quantity X with the standard $N(0, 1)$ p.d.f. Its expectation

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 0$$

because the integrand is an odd function and is integrated over the interval $(-\infty, \infty)$ which is symmetric about the origin. Thus its variance is just the second moment (mean square) of X which can be evaluated easily by integration-by-parts[11]

$$\mathrm{Var}\,(X) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot (x e^{-x^2/2})\, dx.$$

[11] Recall the integration-by-parts formula: $\int f(x)g'(x)\, dx = f(x)g(x) - \int f'(x)g(x)\, dx$.

$$= \frac{1}{\sqrt{2\pi}} \left(-x \cdot e^{-x^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2}\, dx \right) = 1,$$

because $\lim_{x \to \pm\infty} x \cdot e^{-x^2/2} = 0$ and $(1/\sqrt{2\pi}) \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = 1$.

Now, let us consider a general Gaussian random quantity Y with $N(\mu, \sigma^2)$ p.d.f.,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

In view of Example 4.1.7,

$$Y = \sigma X + \mu.$$

The above properties of the expectation and the variance ((4.2.7) and (4.2.8), and (4.2.11) and (4.2.12)) immediately give

$$\mathbf{E}(Y) = \mathbf{E}(\sigma X + \mu) = \sigma \mathbf{E}(X) + \mu = \mu,$$

and

$$\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \mathrm{Var}(\sigma X) = \sigma^2 \mathrm{Var}(X) = \sigma^2.$$

Thus the parameters μ and σ^2 in the Gaussian $N(\mu, \sigma^2)$ p.d.f. are, simply, its expectation and variance.

Remark 4.2.1 (Sums of Random Quantities?) Note that the discussions carried out in the previous two section permitted us, in principle, to determine the probability distributions (and thus expectations, moments, etc.) of functions $g(X)$, once the distribution of X itself was known. However, an effort to determine the distribution of the sum $X + Y$ if the separate distributions of X and Y are known is bound to end up in failure; there is simply not enough information about how the values of X and Y are paired up. This is one of the reasons why one must study the distribution of the pair (X, Y) viewed as the distribution a single random vector. This will be done in the next section.

4.3 Random Vectors, Conditional Probabilities, Statistical Independence, and Correlations

A *random vector* $\boldsymbol{X}$ has components $X_1, X_2, \ldots, X_d$, which are scalar random quantities, that is

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_d),$$

where d is the dimension of the random vector. For the sake of simplicity of notation we shall consider first the case of dimension $d = 2$, and we shall write $\boldsymbol{X} = (X, Y)$.

Statistical properties of random vectors are characterized by their *joint probability distributions*. In the discrete case, for a random vector $\boldsymbol{X}$ taking discrete values $\boldsymbol{x} = (x, y)$, the joint probability distribution is

$$\mathbf{P}(\boldsymbol{X} = \boldsymbol{x}) = \mathbf{P}(X = x, Y = y) = p_{\boldsymbol{X}}(x, y), \tag{4.3.1}$$

and

$$\sum_{(x,y)} p_{\boldsymbol{X}}(x, y) = 1. \tag{4.3.2}$$

Example 4.3.1 (A Bernoulli Random Vector) The random vector (X, Y) takes values $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, with the following joint probabilities:

$$p_{(X,Y)}(0, 0) = (1 - p)^2, \qquad p_{(X,Y)}(0, 1) = p(1 - p),$$

$$p_{(X,Y)}(0, 1) = (1 - p)p, \qquad p_{(X,Y)}(1, 1) = p^2.$$

It is easy to check that

$$\sum_{x=0}^{1} \sum_{y=0}^{1} p_{(X,Y)}(x, y) = 1.$$

In the special case $p = 1/2$ all four possible values of this random vector are taken with the same probability equal to 1/4.

A continuous random vector is characterized by its *joint p.d.f.* $f_{(X,Y)}(x, y)$, which is a nonnegative function of two variables x, y, such that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dx\, dy = 1. \tag{4.3.3}$$

In this case the probability that the random vector (X, Y) takes values in a certain domain A of the 2D space is calculated by evaluating the double integral of the joint p.d.f. over the domain A:

$$\mathbf{P}((X, Y) \in A) = \int \int_A f_{(X,Y)}(x, y)\, dx\, dy. \tag{4.3.4}$$

For example, if the domain A is a rectangle $[a, b] \times [c, d] = \{(x, y) : a \le x \le b, c \le y \le d\}$, then

$$\mathbf{P}((X, Y) \in A) = \mathbf{P}(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f_{(X,Y)}(x, y)\, dy\, dx. \tag{4.3.5a}$$

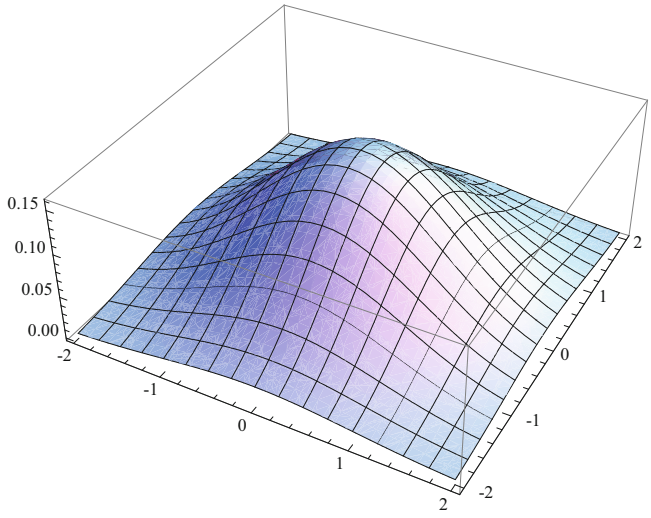


Fig. 4.14 Plot of the surface representing a 2D Gaussian joint p.d.f. (4.3.6) in the case $\sigma_x, \sigma_y = 1$ and $\mu_x, \mu_y = 0$

If the domain $B = \{(x, y) : x^2 + y^2 \leq R^2\}$ is a centered disk of radius R, then

$$\mathbf{P}((X, Y) \in B) = \mathbf{P}(X^2 + Y^2 \leq R^2) = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} f_{(X,Y)}(x, y)\, dy\, dx. \tag{4.3.5b}$$

The graph of a 2D joint p.d.f. is a surface over the (x, y)-plane such that the volume underneath it is equal to 1, see (4.3.3).

Example 4.3.2 (A 2D Gaussian Random Vector) An example of the 2D Gaussian joint p.d.f. is given by the formula

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right], \tag{4.3.6}$$

where $\sigma_x, \sigma_y > 0$ and μ_x, μ_y are arbitrary real numbers. Figure 4.14 shows the plot of the surface representing a 2D Gaussian joint p.d.f. in the case $\sigma_x, \sigma_y = 1$ and $\mu_x, \mu_y = 0$.

Calculation of the probabilities $\mathbf{P}(a \leq X \leq b, c \leq Y \leq d)$ is here reduced to calculation of one-dimensional Gaussian probabilities since the joint 2D density in this case is the product of two 1D Gaussian densities, one depending only on x,

and the other on y,[12] and the double integral splits into the product of two single integrals. To obtain numerical values, tables of (or software for) 1D $N(0, 1)$ c.d.f. have to be used; see Sect. 4.5.

In the special case of equal variances $\sigma_x^2 = \sigma_y^2 = \sigma^2$, the probability that the above Gaussian random vector takes values in a disk of radius R centered at (μ_x, μ_y) can, however, be carried out explicitly by calculation of the integral in polar coordinates (θ, r):

$$\mathbf{P}\Big((X-\mu_x)^2+(Y-\mu_y)^2 \le R^2\Big) = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right] dy\, dx$$

$$= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} \int_0^R \exp\left[-\frac{r^2}{2\sigma^2}\right] r\, dr\, d\theta$$

$$= \frac{1}{\sigma^2}\left[-\sigma^2 \exp\left[-\frac{r^2}{2\sigma^2}\right]\right]_0^R = 1 - e^{-R^2/2\sigma^2}.$$

Because the joint p.d.f. gives complete information about the random vector (X, Y), it yields also complete information about the probability distributions of each of the component random quantities. These distributions are called *marginal distributions* of the random vector.

In particular, for a discrete random vector, the marginal distribution of the component X is

$$p_X(x) = \sum_y p_{(X,Y)}(x, y). \tag{4.3.7}$$

To find the probability of X taking a particular value x_0 we simply need to sum, over all possible y's, the probabilities of (X, Y) taking values (x_0, y). For a continuous random vector the marginal p.d.f. of the component X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dy. \tag{4.3.8}$$

It is important to observe that the marginal distributions of components of a random vector *do not* determine its joint distribution. Indeed, the example provided below shows that it is quite possible for random vectors to have the same marginal probability distributions of their components while their joint probability distributions are different.

[12] We will have more to say about joint p.d.f.s of this type in the next few pages. The multiplicative property is equivalent to the concept of statistical independence of components of a random vector.

Example 4.3.3 (Different Random Vectors with the Same Marginal Probability Distributions) A random vector (X, Y) has components X and Y which take values 1, 2, and 3, and 1 and 2, respectively. The joint probability distribution of this random vector is given in the following table:

Y\ X	1	2	3	Y
1	30/144	24/144	18/144	6/12
2	30/144	24/144	18/144	6/12
X	5/12	4/12	3/12	$\sum = 1$

So, for example, $\mathbf{P}((X, Y) = (3, 2)) = 3/24$. The last row in the above table gives the marginal probability distribution for the component X, and the last column, the marginal probability distribution for the component Y.

Consider now another random vector (W, Z) with components W and Z which also take values 1, 2, and 3, and 1 and 2, respectively. The joint distribution of this random vector is given by a different table:

Z\ W	1	2	3	Z
1	1/12	2/12	3/12	6/12
2	4/12	2/12	0	6/12
W	5/12	4/12	3/12	$\sum = 1$

This time $\mathbf{P}((X, Y) = (3, 2)) = 0$. The last row in the above table gives the marginal probability distribution for the component W, and the last column, the marginal probability distribution for the component Z. The marginal probability distributions for vectors (X, Y) and (W, Z) are the same, while their joint distributions are different.

Conditional Probabilities Knowledge of the joint p.d.f. permits us also to introduce the concept of the conditional probability (in the discrete case) and the conditional density (in the continuous case). Thus, the conditional probability of the component X taking value x, given that the second component Y took value y, is given by the formula[13]

$$p_{X|Y}(x|y) \equiv \mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)} = \frac{p_{(X,Y)}(x, y)}{p_Y(y)}, \qquad (4.3.9)$$

[13]The notation $p_{X|Y}(x|y) \equiv \mathbf{P}(X = x|Y = y)$ reads: probability of $X = x$, given $Y = y$.

and the conditional probability density function of X given $Y = y$ is given by the formula

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}. \tag{4.3.10}$$

In other words, conditional probability distributions are distributions of values of one component of a random vector calculated under the assumption that the value of the other component has already been determined.

Conditional probabilities are *bona fide* probabilities as they satisfy the normalization property. Indeed, say, in the continuous case, for each fixed y,

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = \frac{\int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1,$$

in view of the formula (4.3.8) which calculates the marginal density from the joint density.

If the component X of random vector (X, Y) takes on distinct values $x_1, x_2, \ldots, x_n$, then the additive property of probabilities immediately gives the following *total probability formula:*

$$\mathbf{P}(Y = y) = \sum_{i=1}^{n} \mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i).$$

Example 4.3.4 (How to Avoid Running into a Bear?) Heuristically, one can think about conditional probabilities as probabilities obtained under additional constraints. Think here about probability of your running into a bear during a hike. Given that you are hiking in the Cleveland Metroparks, the probability of the event may be only 0.0001; in Yellowstone the similar conditional probability may be as high as 0.75. Now assume you participate, with 51 of your classmates, in a raffle and the prize is a trip to Yellowstone; the consolation prize is a group hike in the Metroparks. The total probability of your running into a bear would then be $0.0001 \cdot (51/52) + 0.75 \cdot (1/52) \approx 0.015$.

One of the corollaries of the total probability formula is the celebrated *Bayes formula for reverse conditional probabilities* which, loosely speaking, computes the conditional probability of X, given Y, in terms of the conditional probabilities of Y, given X:

$$\mathbf{P}(X = x_i|Y = y) = \frac{\mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i)}{\sum_{i=1}^{n} \mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i).}$$

Indeed,

$$\mathbf{P}(X = x_i | Y = y) = \frac{\mathbf{P}(X = x_i, Y = y)}{\mathbf{P}(Y = y)} \cdot \frac{\mathbf{P}(X = x_i)}{\mathbf{P}(X = x_i)}$$

$$= \frac{\mathbf{P}(Y = y | X = x_i) \cdot \mathbf{P}(X = x_i)}{\mathbf{P}(Y = y)},$$

and an application of the total probability formula immediately gives the final result.

Example 4.3.5 (Transmission of a Binary Signal in the Presence of Random Errors) A channel transmits binary symbols 0 and 1 with random errors. The probability that the symbols 0 and 1 appear at the input of the channel are, respectively, 0.45 and 0.55. Because of transmission errors, if the symbol 0 appears at the input, then the probability of it being received as 0 at the output is 0.95. The analogous conditional probability is 0.9, for the symbol 1 to be received, given that it was transmitted. Our task is to find the reverse conditional probability that the symbol 1 was transmitted given that 1 was received.

The random vector here is (X, Y), where X is the input signal and Y is the output signal. The problem's description contains the following information:

$$\mathbf{P}(X = 0) = 0.45, \qquad \mathbf{P}(X = 1) = 0.55,$$

and

$$\mathbf{P}(Y = 0 | X = 0) = 0.95, \qquad \mathbf{P}(Y = 1 | X = 1) = 0.9,$$

so that

$$\mathbf{P}(Y = 1 | X = 0) = 0.05, \qquad \mathbf{P}(Y = 0 | X = 1) = 0.1.$$

We are seeking $\mathbf{P}(X = 1 | Y = 1)$ and the Bayes formula gives the answer:

$$\mathbf{P}(X = 1 | Y = 1)$$

$$= \frac{\mathbf{P}(Y = 1 | X = 1) \cdot \mathbf{P}(X = 1)}{\mathbf{P}(Y = 1 | X = 0) \cdot \mathbf{P}(X = 0) + \mathbf{P}(Y = 1 | X = 1) \cdot \mathbf{P}(X = 1)}$$

$$= \frac{0.9 \cdot 0.55}{0.05 \cdot 0.45 + 0.9 \cdot 0.55} \approx 0.9565.$$

Statistical Independence Components X and Y of a random vector $\boldsymbol{X} = (X, Y)$ are said to be *statistically independent* if the conditional probabilities of X given Y are independent of Y and *vice versa*. In the discrete case, this means that, for all x and y,

$$\mathbf{P}(X = x | Y = y) = \mathbf{P}(X = x),$$

which is equivalent to the statement that the joint p.d.f. is the product of the marginal p.d.f.s. Indeed, the above independence assumption and the formula defining the conditional probabilities yield

$$\begin{aligned}\mathbf{P}(X = x, Y = y) &= \mathbf{P}_{(X,Y)}(x, y)\\ &= \mathbf{P}_X(x) \cdot \mathbf{P}_Y(y) = \mathbf{P}(X = x) \cdot \mathbf{P}(Y = y).\end{aligned} \tag{4.3.11}$$

In the continuous case the analogous definition of independence of X and Y can be stated via the multiplicative formula for the joint p.d.f.:

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y). \tag{4.3.12}$$

Note that both the 2D Bernoulli distribution of Example 4.3.1 and the 2D Gaussian distribution of Example 4.3.2 have statistically independent components X and Y. Also, components of the random vector (X, Y) in Example 4.3.3 are independent since the table was actually obtained by multiplying the marginal probabilities in the corresponding rows and columns. However, the components W and Z of random vector (W, Z) in Example 4.3.3 are not statistically independent. To see this it is sufficient to observe that

$$\mathbf{P}(W = 3, Z = 2) = 0,$$

but

$$\mathbf{P}(W = 3) \cdot \mathbf{P}(Z = 2) = 3/12 \cdot 6/12 = 18/144 \neq 0.$$

Moments of Random Vectors and Correlations If a random quantity Z is a function of a random vector (X, Y), say,

$$Z = g(X, Y),$$

then, as in Section 4.3, we can calculate the mean of Z using the joint p.d.f. Indeed,

$$\mathbf{E}Z = \sum_x \sum_y g(x, y) p_{(X,Y)}(x, y), \tag{4.3.13}$$

in the discrete case, and

$$\mathbf{E}Z = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{(X,Y)}(x, y)\, dx\, dy, \tag{4.3.14}$$

in the continuous case.

A mixed central second-order moment corresponding to function $g(x, y) = (x - \mu_X)(y - \mu_Y)$ will play a pivotal role in the analysis of random signals. The number,

$$\text{Cov}(X, Y) = \mathbf{E}\Big[(X - \mu_X)(Y - \mu_Y)\Big] = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y), \tag{4.3.15}$$

is called the *covariance* of X and Y. Obviously, the covariance of X and X is just the variance of X:

$$\text{Cov}(X, X) = \mathbf{E}\Big[(X - \mu_X)(X - \mu_X)\Big] = \text{Var}(X). \tag{4.3.16}$$

In the case when the expectations of X and Y are zero,

$$\text{Cov}(X, Y) = \mathbf{E}(X \cdot Y). \tag{4.3.17}$$

By the Cauchy-Schwartz Inequality,[14]

$$|\text{Cov}(X, Y)| \leq \text{Std}(X) \cdot \text{Std}(Y). \tag{4.3.18}$$

This suggests introduction of yet another parameter for a 2D random vector which is called the *correlation coefficient* of X and Y:

$$\text{Cor}(X, Y) \equiv \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\text{Std}(X) \cdot \text{Std}(Y)}. \tag{4.3.19}$$

In view of (4.3.18) the correlation coefficient is always contained between -1 and $+1$:

$$-1 \leq \rho_{X,Y} \leq 1, \tag{4.3.20}$$

and, in view of (4.3.17), if random components X and Y are linearly dependent, that is $Y = \alpha X$, then the correlation coefficient takes its extreme values

$$\rho_{X,\alpha X} = \pm 1, \tag{4.3.21}$$

depending on whether α is positive or negative. In those cases we say that the random quantities X and Y are perfectly (positively or negatively) correlated. If $\rho_{X,Y} = 0$, then the random quantities X and Y are said to be *uncorrelated*.

[14] Recall that if $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$ are two d-dimensional vectors, then the Cauchy-Schwartz inequality says that the absolute value of their scalar (dot) product is not larger than the product of their norms (magnitudes), i.e., $|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\| \cdot \|\mathbf{b}\|$., where $\langle \mathbf{a}, \mathbf{b} \rangle = a_1 b_1 + \cdots + a_d b_d$, and $\|\mathbf{a}\|^2 = a_1^2 + \cdots + a_d^2$; see Sect. 4.7.

The opposite case is that of statistically independent random quantities X and Y. Then, because of the multiplicative property $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ (4.3.11) and (4.3.12) of the joint p.d.f., we always have

$$\mathbf{E}(XY) = \int \int xy f_X(x) f_Y(y)\, dx\, dy = \mathbf{E}X \cdot \mathbf{E}Y, \tag{4.3.22}$$

so that

$$\mathrm{Cov}\,(X, Y) = \mathbf{E}(XY) - \mathbf{E}X \cdot \mathbf{E}Y = 0, \tag{4.3.23}$$

and the correlation coefficient $\rho_{X,Y} = 0$. Thus *independent random quantities are always uncorrelated*. In this context, the correlation coefficient $\rho_{X,Y}$ is often considered as a measure of "independence" of random quantities X and Y; more appropriately it should be interpreted as a measure of the "linear association" of random quantities X and Y.

Remark 4.3.1 (Uncorrelated Random Quantities Need Not Be Independent) Although statistically independent random quantities are always uncorrelated, the reverse implication is not true in general. Indeed, consider an example of a 2D random vector (X, Y) with values uniformly distributed inside the unit circle. Obviously, because of the symmetry, $\mathbf{E}X = \mathbf{E}Y = 0$, and the covariance (calculated in polar coordinates)

$$\mathrm{Cov}(X, Y) = \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} xy \frac{dy\, dx}{\pi} = \int_0^{2\pi} \int_0^1 r^3 \cos\theta \sin\theta \, \frac{dr\, d\theta}{\pi} = 0.$$

So X and Y are uncorrelated. But they are not independent because, for example,

$$\mathbf{P}(\sqrt{2}/2 < X < 1, \sqrt{2}/2 < Y < 1) \neq \mathbf{P}(\sqrt{2}/2 < X < 1) \cdot \mathbf{P}(\sqrt{2}/2 < Y < 1).$$

Indeed, the left-hand side is zero since the square $\{(x, y) : \sqrt{2}/2 < x < 1, \sqrt{2}/2 < y < 1\}$ lies outside the unit circle, but the right-hand side is positive since

$$\mathbf{P}(\sqrt{2}/2 < X < 1) = \int_{\sqrt{2}/2}^{1} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{dy\, dx}{\pi} = \mathbf{P}(\sqrt{2}/2 < Y < 1) > 0;$$

each of the above probabilities is simply the (normalized) area of the sliver of the unit disc to the right of the vertical line $x = \sqrt{2}/2$. However, in certain special cases the reverse implication is true: Gaussian random quantities are independent if and only if they are uncorrelated, see Chap. 9.

Example 4.3.6 (A Discrete 2D Distribution with Nontrivial Correlation) Consider the random vector (W, Z) from Example 4.3.3. The expectations of the components are

$$\mathbf{E}W = 1(5/12) + 2(4/12) + 3(3/12) = 11/6,$$

$$\mathbf{E}Z = 1(6/12) + 2(6/12) = 3/2.$$

The variances are

$$\mathrm{Var}(W) = (1 - 11/6)^2(5/12) + (2 - 11/6)^2(4/12)$$

$$+(3 - 11/6)^2(3/12) = 23/36,$$

$$\mathrm{Var}(Z) = (1 - 3/2)^2(6/12) + (2 - 3/2)^2(6/12) = 1/4.$$

The expectation of the product is

$$\mathbf{E}(WZ) = (1 \cdot 1)(1/12) + (2 \cdot 1)(2/12) + (3 \cdot 1)(3/12)$$

$$+(1 \cdot 2)(4/12) + (2 \cdot 2)(2/12) + (3 \cdot 2)0 = 5/2.$$

Thus the covariance is

$$\mathrm{Cov}(W, Z) = \mathbf{E}(WZ) - \mathbf{E}(W)\mathbf{E}(Z) = 5/2 - (11/6)(3/2) = -1/4,$$

and, finally, the correlation coefficient of W and Z,

$$\mathrm{Cor}(W, Z) = \frac{\mathrm{Cov}\,(W, Z)}{\mathrm{Std}(W) \cdot \mathrm{Std}(Z)} = \frac{-1/4}{\sqrt{23/36} \cdot \sqrt{1/4}} = -\sqrt{3/23} \approx -0.361.$$

Example 4.3.7 (A Continuous 2D Distribution with Nontrivial Correlation) A random vector (X, Y) has a continuous joint p.d.f. of the form

$$f_{(X,Y)}(x, y) = \begin{cases} C(1 - (x + y)), & \text{for } x, y \geq 0, x + y \leq 1; \\ 0, & \text{elsewhere.} \end{cases}$$

The constant C can be determined from the normalization condition,

$$\int_0^1 \int_0^{1-x} C(1 - (x + y))\, dy\, dx = 1,$$

which gives $C = 6$. The plot of the surface representing this density is given in Fig. 4.15.

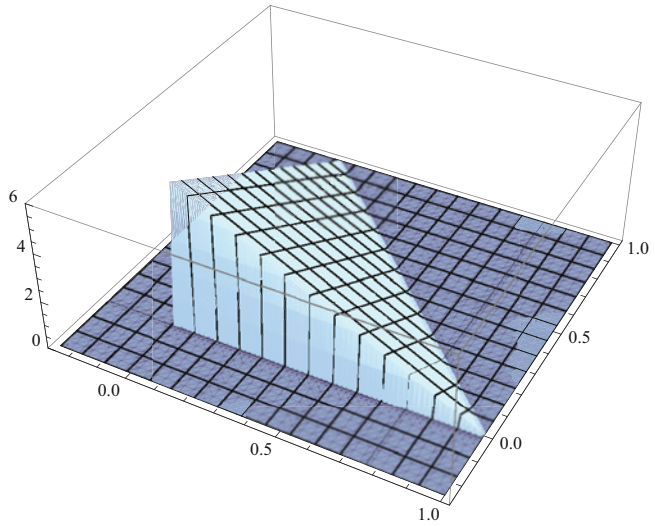


Fig. 4.15 The plot of the surface representing the joint p.d.f. from Example 4.3.7

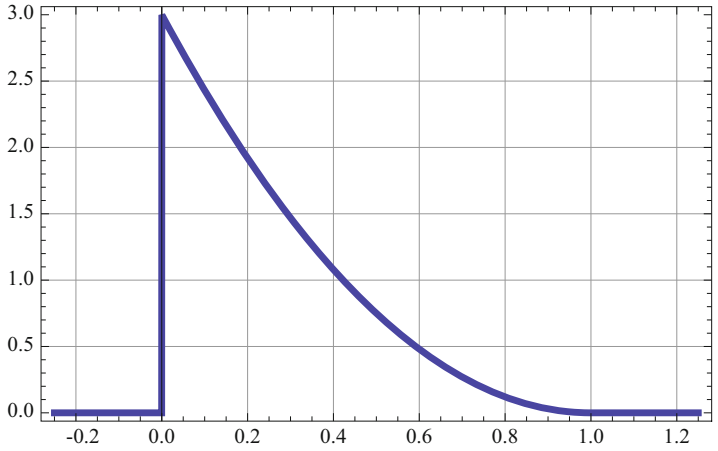


Fig. 4.16 The marginal density $F_X(x)$ of the X component of the random vector from Example 4.3.7

The marginal density of the component X,

$$f_X(x) = \int_0^{1-x} 6(1-(x+y))\,dy = 3(1-x)^2,$$

for $0 < x < 1$. It is equal to 0 elsewhere, and its plot is pictured in Fig. 4.16.

The expectations of X and Y are easily evaluated using the marginal p.d.f.:

$$\mathbf{E}X = \mathbf{E}Y = \int_0^1 x \cdot 3(1-x)^2\, dx = \frac{1}{4}.$$

Similarly, the variances

$$\sigma^2(X) = \sigma^2(Y) = \int_0^1 (x - 1/4)^2 \cdot 3(1-x)^2\, dx = \frac{3}{80}.$$

Finally, the covariance

$$\mathrm{Cov}\,(X, Y) = \int_0^1 \int_0^{1-x} (x - 1/4)(y - 1/4) \cdot 6(1 - (x + y))\, dy\, dx = -\frac{1}{80}.$$

So the random components X and Y are not independent; they are negatively correlated. The correlation coefficient itself is now easily evaluated to be

$$\rho_{X,Y} = \frac{-1/80}{3/80} = -\frac{1}{3}.$$

4.4 The Least Squares Fit, Linear Regression

The roles of the covariance and the correlation coefficient will become better understood in the context of the following *least squares regression* problem. Consider a sample,

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N),$$

of N, 2D, vectors. Its representation in the (x, y) plane is called the *scatter plot* of the sample; see, for example, Fig. 4.17. Our goal is to find a line,

$$y = ax + b,$$

which would provide the best approximation to the scatter plot in the sense of minimizing the sum of the squares of the errors of the approximation measured in the vertical direction. To be more precise the error of the approximation for the i-th sample point is expressed by the formula

$$\epsilon_i(a, b) = |y_i - (ax_i + b)|, \qquad i = 1, 2, \ldots, N,$$

and the sum of the squares of the errors,

$$\sum_{i=1}^{N} \epsilon_i^2(a,b) = \sum_{i=1}^{N} (y_i - (ax_i + b))^2$$

is a nice, differentiable function of two variables a and b. We can find its minimum taking partial derivatives with respect to a and b and equating them to 0[15]:

$$\frac{\partial}{\partial a} \sum_{i=1}^{N} \epsilon_i^2(a,b) = -2 \sum_{i=1}^{N} (y_i - (ax_i + b)) x_i = 0,$$

$$\frac{\partial}{\partial b} \sum_{i=1}^{N} \epsilon_i^2(a,b) = -2 \sum_{i=1}^{N} (y_i - (ax_i + b)) = 0.$$

These two equations, sometimes called the *normal equations*, are linear in a and b and can be easily solved by the substitution method. To make the next step more transparent we will introduce the following simplified notation for different sample means (think here about the means of random quantities with N possible values with each value assigned probability $1/N$). The x and y components of the above data will be treated as N-D vectors, and denoted,

$$\boldsymbol{x} = (x_1, \ldots, x_N), \quad \boldsymbol{y} = (y_1, \ldots, y_N).$$

Various sample means will be denoted as follows:

$$\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{N} x_i, \qquad \overline{\boldsymbol{y}} = \frac{1}{N} \sum_{i=1}^{N} y_i,$$

$$\overline{\boldsymbol{x}^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2, \qquad \overline{\boldsymbol{y}^2} = \frac{1}{N} \sum_{i=1}^{N} y_i^2,$$

$$\overline{\boldsymbol{x}\boldsymbol{y}} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i.$$

Now, the normal equations for a and b can be written in the form

$$a\overline{\boldsymbol{x}} + b - \overline{\boldsymbol{y}} = 0, \quad \text{and} \quad a\overline{\boldsymbol{x}^2} + b\overline{\boldsymbol{x}} - \overline{\boldsymbol{x}\boldsymbol{y}} = 0,$$

[15] This explains why we consider quadratic errors rather than the straight absolute errors; in the latter case the calculus tools would not work so well.

which can be immediately solved to give

$$b = \overline{\boldsymbol{y}} - a\overline{\boldsymbol{x}}, \qquad a = \frac{\overline{\boldsymbol{x}\boldsymbol{y}} - \overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{y}}}{\overline{\boldsymbol{x}^2} - (\overline{\boldsymbol{x}})^2}.$$

The first of the above two equations indicates that the point with coordinates formed by the sample means $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ is located on the regression line. To better see the meaning of the second equation observe that

$$\overline{\boldsymbol{x}\boldsymbol{y}} - \overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{\boldsymbol{x}})(y_i - \overline{\boldsymbol{y}}) = \mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})$$

is just the sample covariance of the x- and y-coordinates of 2D data, and that

$$\overline{\boldsymbol{x}^2} - (\overline{\boldsymbol{x}})^2 = \mathrm{Var}(\boldsymbol{x}), \qquad \overline{\boldsymbol{y}^2} - (\overline{\boldsymbol{y}})^2 = \mathrm{Var}(\boldsymbol{y}).$$

Thus the equation $y = ax + b$ of the regression line becomes now,

$$y = \frac{\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\mathrm{Var}(\boldsymbol{x})}x + \left(\overline{\boldsymbol{y}} - \frac{\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\mathrm{Var}(\boldsymbol{x})}\overline{\boldsymbol{x}}\right),$$

and can be, finally, rewritten in a more elegant and symmetric form,

$$\frac{y - \overline{\boldsymbol{y}}}{\mathrm{Std}(\boldsymbol{y})} = \rho_{\boldsymbol{x},\boldsymbol{y}} \cdot \frac{x - \overline{\boldsymbol{x}}}{\mathrm{Std}(\boldsymbol{x})}, \tag{4.4.1}$$

where

$$\rho_{\boldsymbol{x},\boldsymbol{y}} = \frac{\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\mathrm{Std}(\boldsymbol{x})\mathrm{Std}(\boldsymbol{y})}$$

is the sample correlation coefficient; the standard deviation Std, as usual, denotes the square root of the variance Var. The significance of the form of the regression equation (4.4.1) is now clear: $\rho_{x,y}$ is the slope of the regression line but only after the x- and y-coordinates were standardized (see (4.2.11)), that is, they were centered by the means $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$, and rescaled by the standard deviations $\mathrm{Std}(\boldsymbol{x})$, and $\mathrm{Std}(\boldsymbol{y})$, respectively.

Example 4.4.1 Consider a 2D vector sample of size 10:

x	y
1.05983	1.10539
2.07758	3.36697
3.2816	3.22934
4.13003	6.91638
5.28022	7.65665
6.38872	6.78509
7.11893	8.11736
8.04133	9.94112
9.23407	9.55498
10.3814	10.8697

The coefficients $a = 0.9934$ and $b = 1.0925$, so that the equation of the regression line is

$$y = 0.9934 \cdot x + 1.0925$$

and the correlation coefficient,

$$\rho_{\boldsymbol{x},\boldsymbol{y}} = 0.9503,$$

turns out to be relatively close to 1, indicating strong positive "linear association" between the x- and y-data.

The scatterplot of these data as well as the plot of the regression line (best linear fit) are shown in Fig. 4.17.

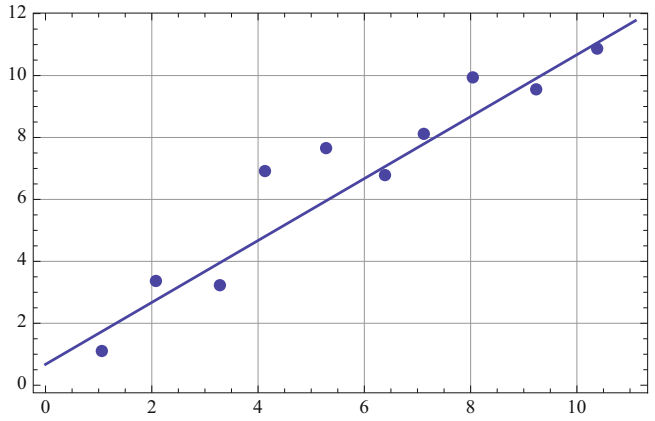


Fig. 4.17 The scatterplot and the least squares fit regression line for data from Example 4.4.1

4.5 The Law of Large Numbers and the Stability of Fluctuations Law

One of the fundamental theorems of statistics, called the *Law of Large Numbers (LLN)*, says that if $X_1, X_2, \ldots, X_n$ are independent random quantities with identical probability distributions, and finite identical expectations $\mathbf{E}X_i = \mu_X$, then, as $n \to \infty$, the averages converge to that expectation, i.e.,

$$\bar{X}_n \equiv \frac{X_1 + X_2 + \ldots + X_n}{n} \longrightarrow \mu_X, \quad \text{as} \quad n \to \infty. \tag{4.5.1}$$

Of course, the immediate issue is what do we mean here by convergence of random variables $\bar{X}_n$. For the purpose of these lectures the convergence of $\bar{X}_n$ to μ_X will mean that the standard deviation of the fluctuations of the averages $\bar{X}_n$ around the mean μ_X, that is the differences $\bar{X}_n - \mu_X$, converge to zero as $n \to \infty$. More formally,

$$\lim_{n\to\infty} \text{Std}\,(\bar{X}_n - \mu_X) = 0. \tag{4.5.2}$$

The statement (4.5.2) can be easily verified if we observe first that:

(a) For any random vector (X, Y) with finite absolute first moments of the components, the expectation

$$\mathbf{E}\,(X + Y) = \mathbf{E}\,(X) + \mathbf{E}\,(Y). \tag{4.5.3a}$$

Indeed, taking $g(x, y) = x + y$ in formulas (4.3.14) and (4.3.15), and using the definition of expectation of functions of random vector we obtain

$$\mathbf{E}\,(X + Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x + y) f_{(X,Y)}(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dx \right) dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy = \mathbf{E}\,(X) + \mathbf{E}\,(Y),$$

in view of the formula (4.3.9) for the marginal p.d.f. of a component of a random vector.[16]

(b) For any random (X, Y) with *independent* components with finite variances, the variance

$$\text{Var}\,(X+Y) = \text{Var}\,(X) + \text{Var}\,(Y). \tag{4.5.3b}$$

This follows immediately from the multiplicative property (4.3.22) of the expectations of independent random variables, see Sect. 4.3.

Now, if X and Y are independent, then $X - \mu_X$ and $Y - \mu_Y$ are also independent, so that, utilizing (4.5.3a) and (4.5.3b),

$$\text{Var}\,(X+Y) = \mathbf{E}((X-\mu_X) + (Y-\mu_Y))^2$$

$$= \mathbf{E}(X-\mu_X)^2 + 2\mathbf{E}(X-\mu_X)\mathbf{E}(Y-\mu_Y) + \mathbf{E}(Y-\mu_Y)^2 = \text{Var}\,(X) + \text{Var}\,(Y),$$

because $\mathbf{E}(X-\mu_X) = \mathbf{E}(Y-\mu_Y) = 0$. Hence,

$$\text{Var}\,(\bar{X}_n - \mu_X) = \text{Var}\left(\frac{X_1-\mu_X}{n} + \ldots + \frac{X_n-\mu_X}{n}\right) = \frac{\text{Var}\,(X)}{n} \tag{4.5.4}$$

which obviously approaches 0 as $n \to \infty$. Thus the Law of Large Numbers (4.5.1), also often called the Law of Averages, is verified, at least in the situation when random quantities X_i have well-defined finite variances.[17]

A more subtle information about the averages is provided by the following *Stability of Fluctuations Law*, usually called the *Central Limit Theorem (CLT)* in the mathematical and statistical literature. It states that as the averages $\bar{X}_n$ fluctuate around the expectation μ_X, the fluctuations, if viewed under a "magnifying glass," turn out to follow, asymptotically as $n \to \infty$, a Gaussian or normal probability distribution. More precisely, the c.d.f. of the standardized (see (4.2.13)) random fluctuations of the averages $\bar{X}_n$ around the mean μ_X,

$$Z_n = \frac{\sqrt{n}}{\text{Std}\,(X)} \cdot (\bar{X}_n - \mu_X), \tag{4.5.5}$$

converges to the standard $N(0, 1)$ Gaussian c.d.f., that is

[16] Note how the knowledge of the joint probability distribution of the random vector (X, Y), and also of $(X_1, X_2, \ldots, X_n)$, is what permits us to study the sums $X + Y$ and $X_1 + X_2 + \cdots + X_n$ as real-valued random quantities with well-defined probability distributions; see Remark 4.2.1 and the following Remark 4.5.1.

[17] Observe that not all random quantities have well-defined, finite variances, see Problem 4.7.28 in Sect. 4.7.

$$\lim_{n\to\infty} \mathbf{P}(Z_n \le z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(x)\,dx, \tag{4.5.6}$$

where the density

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \tag{4.5.7}$$

is that of the standard $N(0, 1)$ Gaussian random quantity. The important assumption of the Central Limit Theorem is that the common variance of X_i's is finite.

Summarizing the above discussion, the Central Limit Theorem can be loosely rephrased as follows:

Standardized random fluctuations of averages of independent and identically distributed random quantities around their common expected value have a limiting standard Gaussian cumulative distribution function.

Remark 4.5.1 (Probability Distribution of a Sum of Independent Random Quantities) It can be immediately verified that all of Z_n's in (4.5.5) have mean zero and variance one, see (4.2.13) and (4.5.3), but the proof of the convergence to a Gaussian limit is more delicate. Without going into the details (for a sketch of the full proof, see Sect. 4.7), it is clear that the proof has to rely on determination of the probability distribution of the sum $Z = X + Y$ of two (or more) of independent random quantities X, and Y. In the case of continuous random quantities (for the derivation in case of discrete random quantities, see Sect. 4.7), it turns out that the p.d.f. of $Z = X + Y$ is the convolution of p.d.f.s of X and Y. Indeed, in view of independence of X and Y, the c.d.f. of Z, for an arbitrary but fixed z, is equal to

$$F_Z(z) = \mathbf{P}(Z \le z) = \mathbf{P}(X + Y \le z) = \int\int_{\{(x,y):x+y\le z\}} f_{(X,Y)}(x, y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_X(x) f_Y(y)\,dx\,dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} f_X(x)\,dx\right) f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f_X(u-y)\,du\right) f_Y(y)\,dy = \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f_X(u-y) f_Y(y)\,dy\right) du,$$

after a change of variables, $x = u - y$, and then, a change of the order of integration. Consequently, the p.d.f.

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy = (f_X * f_Y)(z). \tag{4.5.8}$$

As we have seen in Chap. 2, convolution can be a fairly complex operation even in the case of relatively simple $f_X(x)$ and $f_Y(y)$. Moreover, the distribution of $X_1 + \ldots + X_n$ in (4.5.1) is an n-fold convolution of the p.d.f. $f_X(x)$, and the n is growing

to infinity. So dealing directly with the p.d.f. of the average $\bar{X}_n, n \to \infty$, seems to be a hopeless task. However, in view of Chap. 2, it is obvious that the whole problem would be greatly simplified if instead of dealing with p.d.f.s one could employ their Fourier transforms; the convolution are replaced in the frequency domain by simple point-wise products. This idea is implemented in the sketch of the proof suggested in Problem 4.7.24.

4.6 Estimators of Parameters and Their Accuracy: Confidence Intervals

The Law of Large Numbers can be reinterpreted as follows: If $X_1, X_2, \ldots, X_n$, are independent and identically distributed random quantities representing repeated sampling from a certain probability distribution $F_X(x)$, then, as n increases, the sample means $\bar{X}_n, n = 1, 2, \ldots,$ become better and better estimators of the expectation of that distribution. In statistical terminology the Law of Large Numbers (4.5.1) says that $\bar{X}_n$ is a *consistent estimator* for parameter μ_X.

The Central Limit Theorem (4.5.5) and (4.5.6) permits us to say what is the error of approximation of the theoretical mean μ_X by the sample mean $\bar{X}_n$, or, in other words, to establish the accuracy of the above estimation. Indeed, for a given sample of size n, the CLT says that the difference between the parameter μ_X and its estimator, the sample mean $\bar{X}_n$, is, after normalization by $\sqrt{n}/\mathrm{Std}(X)$, approximately $N(0, 1)$-distributed so that, for large n,

$$\mathbf{P}\left(-\epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}} \le \bar{X}_n - \mu_X \le \epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}}\right) \approx \Phi(\epsilon) - \Phi(-\epsilon) = 2\Phi(\epsilon) - 1, \tag{4.6.1}$$

where $\Phi(z)$ is the c.d.f. of the standard Gaussian ($N(0, 1)$) random quantity tabulated in Table 4.1.

If X itself has a Gaussian p.d.f. the above approximate equality becomes exact for all n. This follows from the fact that the sum of two independent Gaussian random quantities is again a Gaussian random quantities, obviously with the mean and variance being the sums of means and variances, respectively, of the corresponding random summands; see Sect. 4.7.

The contents of the formula (4.6.1) can be rephrased as follows: the true value of *parameter μ_X is contained in the random interval*

$$\left(\bar{X}_n - \epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}}, \bar{X}_n + \epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}}\right)$$

with probability

$$C = C(\epsilon) = 2\Phi(\epsilon) - 1.$$

The above random interval is called the *confidence interval*, and the probability $C = C(\epsilon)$ is called its *confidence level*. The above statement is sometimes abbreviated by writing

$$\mu_X = \bar{X}_n \pm \epsilon \frac{\text{Std}(X)}{\sqrt{n}}$$

at the confidence level C. Note that it is the center of the above random interval that is random; its length is not random unless $\text{Std}(X)$ itself has to be estimated from the sample.

Example 4.6.1 (A 95% Confidence Interval for μ_X with Known $\text{Std}(X)$*)* Sixteen independently repeated measurements of a random quantity X were conducted resulting in $\bar{X}_{16} = 2.56$. Suppose that we know that $\text{Std}(X) = 0.12$. To find the 95%-confidence interval for μ_X using (4.6.1) we need to find ϵ such that $2\Phi(\epsilon) - 1 = 0.95$, i.e., $\Phi(\epsilon) = 0.975$. From Table 4.1 of the Gaussian $N(0, 1)$ c.d.f. we have $\epsilon = 1.96$. Thus, at the 95%-confidence level,

$$2.56 - 1.96\frac{0.12}{\sqrt{16}} \le \mu_X \le 2.56 + 1.96\frac{0.12}{\sqrt{16}},$$

that is,

$$\mu_X = 2.56 \pm 0.059$$

at the 95%-confidence level. The above approximate confidence interval is exact if X has a Gaussian distribution.

Remark 4.6.1 (Error of the Gaussian Approximation in the CLT) To be honest we left open the essential, but delicate question of how good is the approximate equality in the basic formula (4.6.1) or, equivalently, the question of precise estimation of the error in the Central Limit Theorem (4.5.6) which, by itself, only says that the difference

$$\mathbf{P}(Z_n \le z) - \Phi(z) \to 0, \qquad \text{as} \qquad n \to \infty,$$

where

$$Z_n = \frac{(X_1 + \ldots + X_n) - n\mu_X}{\sqrt{n} \cdot \text{Std}(X)}$$

are standardized sums $X_1 + \ldots + X_n$. It turns out that the accuracy in CLT is actually pretty good if X_i's have higher absolute moments finite. In particular, if the third central moment $m_3 = \mathbf{E}|X - \mu_X|^3 < \infty$, then, for all $-\infty < x < \infty$, and $n = 1, 2, \ldots,$

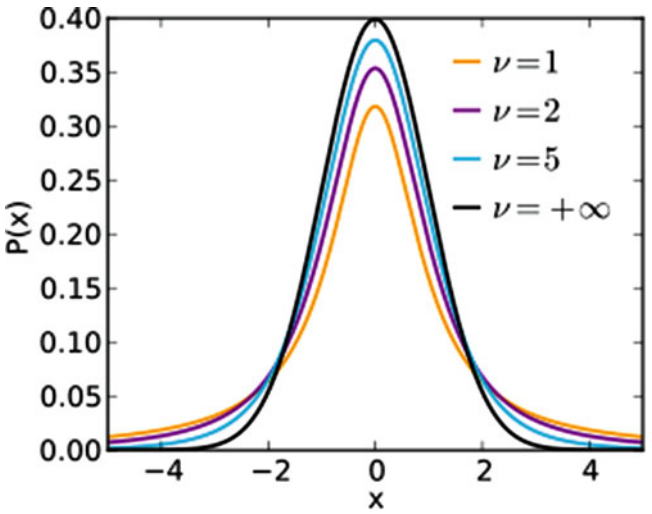


Fig. 4.18 Student's-t p.d.f.s with (bottom to top) 1, 2, 5, 15, and ∞ degrees of freedom (from http://en.wikipedia.org/wiki/Student's-t-distribution)

$$|\mathbf{P}(Z_n \le z) - \Phi(z)| \le \frac{\nu m_3}{\sqrt{n}(\mathrm{Std}(X))^3},$$

where ν is a universal (independent of n and X) constant contained in the interval (0.4097, 0.7975). Its exact value is not known.[18]

Of course, the procedure used in Example 4.6.1 requires advance knowledge of the standard deviation $\mathrm{Std}(X)$. If that parameter is unknown, then the obvious step is to try to estimate it from the sample $X_1, X_2, \ldots, X_n$, itself using the sample variance estimator

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \tag{4.6.2}$$

which is an unbiased estimator for $\mathrm{Var}(X)$, see Problem 4.7.29 (Fig. 4.18).

But in this case, even if X_i's are Gaussian, the standardized random quantity

$$T = \frac{\sqrt{n}}{S_n}(\bar{X} - \mu_X) \tag{4.6.3}$$

[18]This error estimate in CLT is known as the Berry-Esseen Theorem and its proof can be found, for example, in V.V. Petrov's monograph *Sums of Independent Random Variables*, Springer-Verlag, 1975.

is no longer $N(0, 1)$-distributed, so a simple construction of the confidence interval for μ_X using the Gaussian distribution is impossible. However, in the narrower situation of a Gaussian random sample $X_1, X_2, \ldots, X_n$, it is known that the random quantity T has the p.d.f.

$$f_T(x; n-1) = \frac{\Gamma((n)/2)}{\sqrt{n\pi}\Gamma((n-1)/2)} \left(1 + \frac{x^2}{n-1}\right)^{-n/2}, \tag{4.6.4}$$

which, traditionally, is called *Student's-t* p.d.f. with $(n-1)$ degrees of freedom.[19]

The Gamma function $\Gamma(\gamma)$ appearing in the definition of f_T is defined by the formula

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x}\, dx, \qquad \gamma > 0. \tag{4.6.5}$$

It is worth noting that

$$\gamma\Gamma(\gamma) = \Gamma(\gamma+1), \qquad \text{and} \qquad \Gamma(n) = (n-1)!, \tag{4.6.6}$$

if n is a positive integer. So, the Gamma function is an interesting extension of the concept of the factorial to noninteger numbers.

Thus, in the Gaussian case with unknown variance, the confidence interval for μ_X at confidence level $C = (2F_T(\epsilon) - 1)$ is of the form,

$$\left(\bar{X}_n - \epsilon\frac{S_n}{\sqrt{n}},\ \bar{X}_n + \epsilon\frac{S_n}{\sqrt{n}}\right). \tag{4.6.7}$$

Since in practice the goal is often to construct confidence intervals at given confidence levels, instead of tabulating the Student's-t c.d.f.s $F_T(t)$, it is convenient to tabulate the relevant probabilities via the tail quantile function $q(\alpha; n)$ defined by the equality

$$q(\alpha; n) = Q_T(1-\alpha; n),$$

where the quantile function $Q_T(\alpha; n)$, see Remark 3.1.2, is the inverse function to c.d.f. $F_T(t)$, i.e.,

$$F_T(Q_T(\alpha; n)) = \alpha. \tag{4.6.8}$$

[19] See, for example, M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity and Chaotic Behavior in Engineering and Science*, Birkhäuser-Boston 1998, for more details on the statistical issues discussed in this section.

Table 4.2 Tail quantiles $q_T(\alpha; n)$ of student's-t distribution

$n \backslash \alpha$	0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.317	636.61
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.451	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	8.610
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.500	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.813	2.228	2.764	3.169	4.144	4.587
11	1.364	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.141
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.584	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.553	2.879	3.610	3.992
19	1.328	1.729	2.093	2.540	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.849
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.320	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.059	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.312	1.701	2.049	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.311	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Thus the tail quantile $q(\alpha; n)$ is the number such that the probability that Student's-t random quantity with n degrees of freedom is greater than α. Selected tail quantiles $q_T(\alpha; n)$ are provided in Table 4.2.

Using the tail quantiles $q_T(\alpha; n)$ the C-confidence level interval for μ_X can now be simply written in the form

$$\left(\bar{X}_n - q_T\left(\frac{1-C}{2}, n-1\right)\frac{S_n}{\sqrt{n}},\ \bar{X}_n + q_T\left(\frac{1-C}{2}, n-1\right)\frac{S_n}{\sqrt{n}}\right). \tag{4.6.9}$$

The Student's-t p.d.f.s are symmetric about zero, and bell-shaped but flatter than the $N(0,1)$ p.d.f. (why?). For large values of N, say $n > 20$, they are practically indistinguishable from the standard Gaussian p.d.f. (why? see Problem 4.7.18), and the latter can be used in construction of confidence intervals even in the case of unknown variance.

Example 4.6.2 (A 95%-Confidence Interval for μ_X with Unknown Std(X)*)* Sixteen independent measurements of a Gaussian random quantity X resulted in $\bar{X}_{16} = 2.56$ and $S_{16} = 0.12$. With the desired confidence level $C = 0.95$, Table 4.2 yields the tail quantile,

$$q_T((1-0.95)/2;\, 15) = q_T(0.025;\, 15) = 2.13.$$

Hence the 95%-confidence interval for the expectation μ_X is of the form

$$\left(2.56 - 2.13 \cdot \frac{0.12}{\sqrt{16}},\ 2.56 + 2.13 \cdot \frac{0.12}{\sqrt{16}}\right)$$

or, in other words, $\mu_X = 2.56 \pm 0.064$ at the 95% confidence level. Observe that, not surprisingly, in the absence of the precise knowledge of the variance $\mathrm{Var}(X)$ which had to be replaced by the estimator S_{16}, this confidence interval is wider than that in Example 4.6.1 ($\mu_X = 2.56 \pm 0.059$ at the same 95% confidence level) where the value of the variance was assumed to be known exactly.

The final question in this section is: How good is the sample variance estimator S_n^2 introduced in (4.6.2)? Here, again the answer is difficult for a general c.d.f. F_X. However, in the case of a Gaussian $N(\mu_X, \sigma_X^2)$ sample one can prove that the nonnegative random quantity

$$\chi^2 = \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \tag{4.6.10}$$

has the p.d.f. of the form

$$f_{\chi^2}(x; n-1) = \frac{1}{2^{(n-1)/2}\Gamma((n-1)/2)}x^{(n-3)/2}e^{-x/2}, \qquad x \geq 0, \tag{4.6.11}$$

which traditionally is called the chi-square p.d.f. with $(n-1)$ degrees of freedom.

Again, here it is more convenient to tabulate the tail quantiles $q_{\chi^2}(\alpha; n)$ rather than the c.d.f.s themselves; see Table 4.3. Thus a C-confidence level interval for σ_X^2 is of the form

$$\left(\frac{(n-1)S_X^2}{q_{\chi^2}((1-C)/2; n-1)}, \frac{(n-1)S_X^2}{q_{\chi^2}((1+C)/2; n-1)} \right), \tag{4.6.12}$$

if we decide to make a symmetric cutoffs at the top and the bottom of the range of the chi-square p.d.f.

Example 4.6.3 (A 99%-Confidence Interval for $\mathrm{Var}(X)$*)* Twenty-six independent measurements of a Gaussian random quantity X resulted in the estimate $S_{26}^2 = 1.37$ for the variance $\mathrm{Var}(X)$. With $C = 0.99$, Table 4.3 yields

$$q_{\chi^2}((1+0.99)/2; 25) = q_{\chi^2}(0.995; 25) = 10.52,$$

and

$$q_{\chi^2}((1-0.99)/2; 25) = q_{\chi^2}(0.005; 25) = 46.92.$$

Thus the 99%-confidence level interval for the variance σ_X^2 is

$$\left(\frac{25 \cdot 1.37}{46.92}, \frac{25 \cdot 1.37}{10.52} \right) = (0.72,\ 3.25).$$

The interval is relatively large because the confidence level demanded is very high. Note that it is not symmetric about the estimated value $S_{26}^2 = 1.37$.

Remark 4.6.2 (Asymmetry of the Chi-Square Distribution) Both the standard Gaussian and Student's-t distribution are symmetric about the origin; their p.d.f.s are even functions. For that reason, to construct confidence intervals for them at a given (high) confidence level it is sufficient to know their tail quantiles only for small tail probabilities. However, the chi-square distribution is asymmetric, see Fig. 4.19.

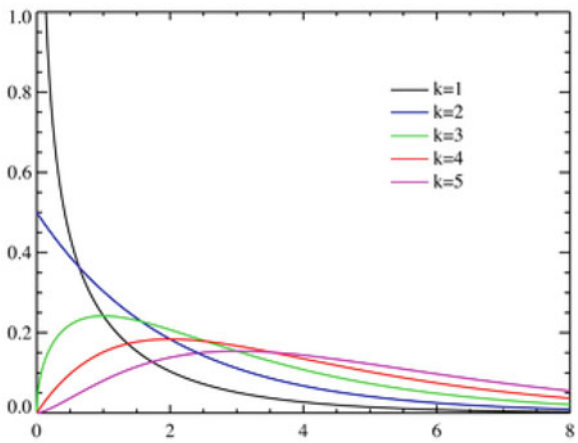


Fig. 4.19 Chi-square p.d.f.s with (top to bottom) 1, 2, 3, 4, and 5 degrees of freedom (from http://en.wikipedia.org/wiki/chi-square-distribution)

Table 4.3 Tail quantiles $q_{\chi^2}(\alpha; n)$ of chi-square distribution

$n\backslash\alpha$	0.9950	0.9900	0.9750	0.9500	0.9000	0.1000	0.0500	0.0250	0.0100	0.0050
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.937
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.160	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.17	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.479	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.462	65.473
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

Thus the tables need to contain tail quantiles for both small and large (close to 1) tail probabilities. This need is on display in the above Example 4.6.3.

4.7 Problems and Exercises

Use *Mathematica*, *Maple*, or *Matlab* as needed throughout this and other problem sections.

1 Plot the c.d.f.s of binomial random quantities X with $p = 0.21$ and $n = 5, 13, 25$. Calculate probabilities that X take values between 1.3 and 3.7. Repeat the same exercise for $p = 0.5$ and $p = 0.9$.

2* Calculate the probability that a random quantity uniformly distributed over the interval $[0, 3]$ takes values between 1 and 3. Do the same calculation for the exponentially distributed random quantity with parameter $\mu = 1.5$, and the Gaussian random quantity with parameters $\mu = 1.5, \sigma^2 = 1$.

3 Prove that $\gamma\Gamma(\gamma) = \Gamma(\gamma + 1)$, and that $\Gamma(n) = (n - 1)!$ Use the integration by parts formula. Verify analytically that $\Gamma(1/2) = \sqrt{\pi}$. Use the idea employed in Example 4.1.6 to prove that the standard Gaussian density is normalized. Then calculate moments of order n of the standard Gaussian distribution.

4* The p.d.f. of a random variable X is expressed by the quadratic function $f_X(x) = ax(1 - x)$, for $0 < x < 1$, and is zero outside the unit interval. Find a from the normalization condition and then calculate $F_X(x)$, $\mathbf{E}X$, $\mathrm{Var}(X)$, $\mathrm{Std}(X)$, the n-th central moment, and $\mathbf{P}(0.4 < X < 0.9)$. Graph $f_X(x)$, and $F_X(x)$.

5 Find the c.d.f and p.d.f. of the random quantity $Y = X^3$, where X is uniformly distributed on the interval $[1, 3]$. Calculate its mean, variance, and higher order moments.

6* Find the c.d.f and p.d.f. of the random quantity $Y = \tan X$, where X is uniformly distributed over the interval $(-\pi/2, \pi/2)$. Find a physical (geometric) interpretation of this result. Show that the second moment of Y (and thus variance) is infinite, and that the expectation $\mathbf{E}(Y)$ is not well defined despite the symmetry of the p.d.f. about zero. Also, see problem 4.7.28.

7 Verify that $\mathrm{Var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2$, see formula (4.2.6).

8 Calculate the expectation and the variance of the binomial distribution from Example 4.1.2.

9 Calculate the expectation and the variance of the Poisson distribution from Example 4.1.3.

10 Calculate the expectation, the variance, and the n-th moment of the exponential distribution from Example 4.1.5.

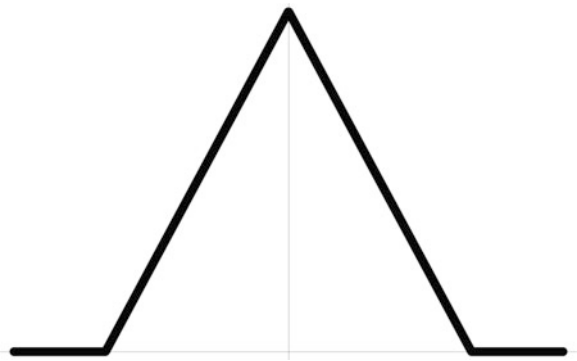


Fig. 4.20 A triangular p.d.f. from Problem 4.7.13

11 Calculate the n-th central moment of the Gaussian distribution from Example 4.1.6.

12 Derive the formula for the binomial distribution from Example 4.1.2 relying on the observation that it is the distribution of the sum of n independent and identically distributed Bernoulli random quantities. Show that if $p = \mu/n$ and $n \to \infty$, then the binomial probabilities converge to the Poisson probabilities.

13* A random quantity X has an even p.d.f., $f_X(x)$, of the triangular shape shown in Fig. 4.20.

(a) How many parameters do you need to describe this p.d.f.? Find an explicit analytic formula for p.d.f. $f_X(x)$ and c.d.f. $F_X(x)$. Graph both of them.
(b) Find the expectation and variance of X.
(c) Let $Y = X^3$. Find the p.d.f. $f_Y(y)$ and graph it.

14 A discrete 2D random vector (X, Y) has the following joint p.d.f.:

$$\mathbf{P}(X = 1, Y = 1) = \frac{2}{12}, \mathbf{P}(X = 2, Y = 1) = \frac{1}{12}, \mathbf{P}(X = 3, Y = 1) = \frac{1}{12},$$

$$\mathbf{P}(X = 1, Y = 3) = \frac{2}{12}, \mathbf{P}(X = 2, Y = 3) = \frac{4}{12}, \mathbf{P}(X = 3, Y = 2) = \frac{2}{12}.$$

Find the marginal distributions of X and Y, their expectations and variances, as well as the covariance and the correlation coefficient of X and Y. Are X and Y independent?

15* Verify the Cauchy-Schwartz Inequality (4.3.18). *Hint:* Take $Z = (X - \mathbf{E}X)/\sigma(X)$ and $W = (Y - \mathbf{E}Y/\sigma(Y)$, and consider the discriminant of the expression $\mathbf{E}(Z + xW)^2$. The latter is quadratic in variable x and necessarily always nonnegative, so it can have at most one root.

16 The following sample of random vector (X, Y) was obtained: $(1, 1.7)$, $(2, 2)$, $(5, 4.3)$, $(7, 5.9)$, $(9, 8)$, $(9, 8.7)$. Produce the scatter plot of the sample and the corresponding least-squares regression line.

17 Using the table of $N(0, 1)$ c.d.f. provided at the end of this chapter calculate $\mathbf{P}(-1 \le Y \le 2)$ if $Y \sim N(0.7, 4)$.

18 Produce graphs of Student's-t p.d.f. $f_T(x, n)$, for $n = 2, 5, 12, 20$, and compare them with the standard normal p.d.f.

19 Produce graphs of the chi-square p.d.f. $f_{\chi^2}(x, n)$ for $n = 2, 5, 12, 20$.

20 Find a constant $c > 0$ such that the function

$$f_X(x) = \begin{cases} c(1+x)^{-4}, & \text{for } x > 0; \\ 0, & \text{for } x \le 0 \end{cases}$$

is a valid p.d.f. Find $\mathbf{P}(1/5 < X < 5)$, $\mathbf{E}(X)$ and the p.d.f., $f_Y(y)$, of $Y = X^{1/5}$.

21 Measurements of voltage V and current I on a resistor yielded the following $n = 5$ paired data: $(1.0, 2.3)$, $(2.0, 4.1)$, $(3.0, 6.4)$, $(4.0, 8.5)$, $(5.0, 10.5)$. Draw the scatter plot and find the regression line providing the least squares fit for the data.

22 Independent measurements of the leakage current I on a capacitor yielded the following data: 2.71, 2.66, 2.78, 2.67, 2.71, 2. 69, 2.70, 2.73 mA. Assuming that the distribution of the random quantity I is Gaussian, find the 95% confidence intervals for the expectation $\mathrm{e}I$ and the variance σ_I^2.

23 Verify that the random quantities $Z_n, n = 1, 2, \ldots,$ defined in (4.5.5) have expectation 0 and variance 1.

24* Complete the following sketch of the proof of the Central Limit Theorem from Sect. 4.5. Start with a simplifying observation (based on Problem 23) that it is sufficient to consider random quantities $X_n, n = 1, 2, \ldots,$ with expectations equal to 0 and variances 1.

(a) Define $\mathcal{F}_X(u)$ as the inverse Fourier transform of the distribution of X:

$$\mathcal{F}_X(u) = \mathbf{E}e^{juX} = \int_{-\infty}^{\infty} e^{jux}\, dF_X(x).$$

Find $\mathcal{F}'_X(0)$ and $\mathcal{F}''_X(0)$. In statistical literature $\mathcal{F}_X(u)$ is called the characteristic function of the random quantity X. Essentially, it completely determines the probability distribution of X via the Fourier transform (inverse of the inverse Fourier transform).

(b) Calculate $\mathcal{F}_X(u)$ for the Gaussian $N(0, 1)$ random quantity. Note the fact that its functional shape is the same as that of the $N(0, 1)$ p.d.f. This fact is the crucial reason for the validity of CLT.

(c) Prove that, for independent random quantities X and Y,

$$\mathcal{F}_{X+Y}(u) = \mathcal{F}_X(u) \cdot \mathcal{F}_Y(u).$$

(d) Utilizing (c), calculate

$$\mathcal{F}_{\sqrt{n}(\bar{X}-\mu_X)/\mathrm{Std(X)}}(u).$$

Then find its limit as $n \to \infty$. Compare it with the characteristic of the Gaussian $N(0, 1)$ random quantity. (Hint: it is easier to work here with the logarithm of the above transform.)

25 Use the above introduced characteristic function technique to prove that the sum of two independent Gaussian random quantities is again a Gaussian random quantity.

26 What is the probability P that a randomly selected chord is shorter than the side S of an equilateral triangle inscribed in the circle? Here are two, seemingly reasonable solutions[20]:

(a) A chord is determined by its two endpoints. Fix one of them to be A. For the chord to be shorter than the side S, the other endpoint must be chosen on either the arc AB or on the arc CA, and each of them is subtended by an angle of 120°. Thus, $P = 2/3$.

(b) A chord is completely determined by its center. For the chord to be shorter than the side S, the center must lie outside the circle of radius equal to the half of the radius of the original circle and the same center. Hence, the probability P equals the ratio of the annular area between two circles and the area of the original circle, which is 3/4.

These two solutions are different. How is that possible?

27 Derive formulas for the c.d.f. $F_Y(y)$, and the p.d.f. $f_Y(y)$, of a transformation $Y = g(X)$ of a random quantity X, in terms of its c.d.f. $F_X(x)$, and p.d.f. $f_X(x)$, in case when the transforming function $y = g(x)$ is *monotonically decreasing*. Follow the line of reasoning used to derive the analogous formulas (4.1.11) and (4.1.12) for *monotonically increasing* transformations. How would you extend these formulas to transformations that are monotonically increasing on some intervals and decreasing on their complement?

28 Consider the Cauchy random quantity X defined in Remark 4.1.2. Plot its c.d.f., and then plot $X = X(\omega)$ as a function on the unit interval. Calculate the probability that X takes values between -3 and $+3$. Compare it with the similar probability for the standard Gaussian random quantity. Find and plot its p.d.f. Compare the rate

[20]For more information, see M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity and Chaotic behavior in Engineering and Science*, Birkhauser-Boston 1998, Example 5.1.1.

of decay at $+\infty$ of the Cauchy p.d.f. with that of the $N(0, 1)$ p.d.f. Show that the expectation of the Cauchy random quantity is undefined and its variance is infinite.

29 Show that the variance estimator S_n^2 introduced in (4.6.2) is unbiased, that is, $\mathbf{E}S_n^2 = \mathrm{Var}(X)$. Also, see problem 4.7.6.

Part II

Chapter 5
Stationary Signals

Abstract In this chapter we introduce basic concepts necessary to study the time-dependent dynamics of random phenomena. The latter will be modeled as a family of random quantities indexed by a parameter, interpreted in this book as time. The parameter may be either continuous or discrete. Depending on the context, and tradition followed by different authors, such families are called *random signals, stochastic processes*, or (random) *time series*. The emphasis here is on random dynamics which is *stationary*, that is, governed by an underlying statistical mechanisms that do not change in time, although, of course, particular realizations of such families will be functions that vary with time. Think here about a random signal produced by the proverbial repeated coin tossing; the outcomes vary while the fundamental mechanics remains the same.

5.1 Stationarity and Autocovariance Functions

A *random* (or *stochastic*) *signal* is a time-dependent family of real-valued[1] random quantities $X(t)$. Depending on the context, one can consider random signals on the positive time line, $t \geq 0$, on the whole time line, $-\infty < t < \infty$, or on a finite time interval, $t_0 \leq t \leq t_1$. Also, it is useful to be able to consider random vector signals and signals with discrete time $t = \ldots, -2, -1, 0, 1, 2, \ldots$

In this book we will restrict our attention to signals that are statistically stationary which means that at least some of their statistical characteristics do not change in time. Several choices are possible here:

First-Order Strictly Stationary Signals In this case the c.d.f., $F_{X(t)}(x) = \mathbf{P}(X(t) \leq x)$, does not change in time (is time-shift invariant), that is,

$$F_{X(t)}(x) = F_{X(t+\tau)}(x), \quad \text{for all} \quad t, \tau, x \tag{5.1.1}$$

[1] At the end of this section we will show how the concepts discussed below should be adjusted if one considers the complex-valued stochastic signals.

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_5

Second-Order Strictly Stationary Signals In this case the joint c.d.f.,

$$F_{(X(t_1),X(t_2))}(x_1, x_2)) = \mathbf{P}(X(t_1) \leq x_1, X(t_2) \leq x_2,)$$

does not change in time, that is

$$F_{(X(t_1),X(t_2))}(x_1, x_2) = F_{(X(t_1+\tau),X(t_2+\tau))}(x_1, x_2), \quad \text{for all} \quad t_1, t_2, \tau, x_1, x_2. \tag{5.1.2}$$

In a similar fashion one can define the n-th *order strict stationarity* of random signal $X(t)$, as the time-shift invariance of the n-th order joint c.d.f., that is the requirement that

$$F_{(X(t_1),\ldots,X(t_n))}(x_1, \ldots, x_n) = F_{(X(t_1+\tau),\ldots,X(t_n+\tau))}(x_1, \ldots, x_n), \tag{5.1.3}$$

for all $t_1, \ldots, t_n, \tau, x_1, \ldots, x_n$.

Finally, a random signal $X(t)$ is said to be *strictly stationary* if, for each $n = 1, 2, \ldots$, it is n-th order strictly stationary.

Obviously, as n increases, verifying the n-th order stationarity gets more and more difficult, not to mention practical difficulties with checking the full strict stationarity. For this reason, a more modest concept of *second-order weakly stationary signals* is useful. In this case the invariance property is demanded only of the moments of the signal up to order two. More precisely, we have the following fundamental definition:

Definition 5.1.1 A signal $X(t)$ is said to be second-order *weakly stationary* if its expectations and covariances are time-shift invariant, that is, if for all t, τ,

$$\mu_X(t) \equiv \mathbf{E}[X(t)] = \mathbf{E}[X(t+\tau)] \equiv \mu_X(t+\tau), \tag{5.1.4}$$

and, for all t_1, t_2, τ, *autocovariance function (ACvF)*

$$\begin{aligned} &\gamma_X(t_1, t_1+\tau) \equiv \mathrm{Cov}(X(t_1), X(t_1+\tau)) \\ &= \mathrm{Cov}(X(t_2), X(t_2+\tau)) \equiv \gamma_X(t_2, t_2+\tau), \end{aligned} \tag{5.1.5}$$

where, as in Chap. 3, for a random vector (X, Y), the covariance

$$\mathrm{Cov}(X, Y) = \mathbf{E}(X - \mu_X)(Y - \mu_Y).$$

It is a consequence of the above two conditions that, for any second-order weakly stationary signal,

$$\mu_X(t) = \mu_X = constant, \tag{5.1.6}$$

and the autocovariance function depends only on the *time lag* τ, and can be written as a function of a single variable:

$$\gamma_X(t, t+\tau) = \gamma_X(0, \tau) = \gamma_X(\tau) \tag{5.1.7}$$

or, equivalently,

$$\gamma_X(s, t) \equiv \gamma_X(0, t-s) = \gamma_X(t-s). \tag{5.1.8}$$

Note that the variance of the stationary signal is also independent of time and is equal to the value of ACvF at $\tau = 0$. Indeed,

$$\mathrm{Var}(X(t)) = \mathrm{Cov}(X(t), X(t)) = \gamma_X(t, t) = \gamma_X(0) = \sigma_X^2 = constant. \tag{5.1.9}$$

In the remainder of these lecture notes we will restrict our attention to second-order weakly stationary signals $X(t)$ which we will simply call *stationary signals*. We will analyze them assuming only the knowledge of their mean value μ_X and their autocovariance function $\gamma_X(t)$.

The following properties of the autocovariance function follow directly from its definition and the Schwartz Inequality (see Sect. 4.7):

$$\gamma_X(-\tau) = \gamma_X(\tau), \tag{5.1.10}$$

and

$$|\gamma_X(\tau)| \leq \gamma_X(0) = \sigma_X^2. \tag{5.1.11}$$

In other words the covariance function is even and its absolute value is dominated by its value at $\tau = 0$, where it is simply equal to the signal's variance.

Remark 5.1.1 (Autocovariance Function (ACvF) vs. Autocorrelation Function (ACF)) You may remember that in Chap. 4 (see (4.3.19)) we have defined the correlation coefficient of random quantities X and Y as normalized covariance, that is the covariance of X and Y, divided by the product of standard deviations of X and Y. Thus, for weakly stationary signals the *autocorrelation function* is also dependent only on the time lag and is expressed by the formula

$$\rho_X(\tau) = \frac{\mathrm{Cov}(X(t), X(t+\tau))}{\mathrm{Std}(X(t))\mathrm{Std}(X(t+\tau))} = \frac{\gamma_X(\tau)}{\gamma_X(0)}.$$

So, in view of (5.1.11), the autocorrelation function always takes values between -1 and $+1$. However, in this book we will employ only the autocovariance function as it also contains information about the variance of the signal (as its value at $\tau = 0$) which, as we will see later on, represents the mean power of the signal. However, in the signal processing literature one often finds the autocovariance

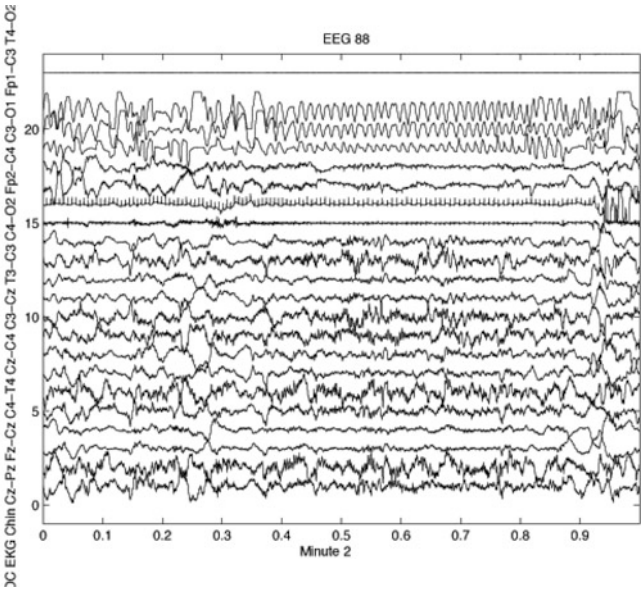


Fig. 5.1 A sample of a 21-channel recording of the sleep electroencephalogram (EEG) of a neonate. The duration of this multidimensional random signal is 60 s and the sampling rate is 64 Hz (From A. Piryatinska's Ph.D. Dissertation, Department of Statistics, Case Western Reserve University, 2004)

function $\gamma_X(\tau)$, called *autocorrelation function* without normalizing it. So, when consulting a particular book or article one has to make sure what definition of ACvF is employed.

The reminder of this section is devoted to a series of examples of stationary data. The first, real-life example shown in Fig. 5.1, displays a sample of a 21-channel recording of the sleep electroencephalogram (EEG) of a neonate. The duration of this multidimensional random signal is 1 min and the sampling rate is 64 Hz. This particular EEG was taken during the so-called mixed frequency sleep stage, and in addition to the EEG also shows related signals such as electrocardiogram (EKG), breathing signal, and eye muscle contraction signal. Signal's components seem stationary for some channels while other channels seem to violate the stationarity property. This can be due to some artifacts in the recordings caused, for example, by the physical movements of the infant or by the onset of a different sleep stage (active, passive, rapid eye movement (REM), etc.) The study of EEG signals provides important information on the state of the brain's neural network and, in the case of infants, can be used to assess the maturity level of their brain. In Sect. 5.2 we will provide a method to estimate the autocovariance function for such real-life data.

Examples 5.1.1–5.1.6 provide various mathematical models of stationary signals. In those cases, the autocovariance functions can be explicitly calculated.

Example 5.1.1 (A Random Harmonic Oscillation) Consider a signal which is a simple harmonic oscillation with nonrandom frequency $f_0 = 1/P$ but random amplitude A such that the second moment $\mathbf{E}A^2 < \infty$, and random phase Θ uniformly distributed over the period and independent of A. In other words,

$$X(t) = A\cos(2\pi f_0(t+\Theta)).$$

The signal is stationary because its mean value

$$\mathbf{E}X(t) = \mathbf{E}A\cos 2\pi f_0(t+\Theta) = \mathbf{E}A\cdot\int_0^P \cos 2\pi f_0(t+\theta)\,\frac{d\theta}{P} = \mathbf{E}A\cdot 0 = 0,$$

and its autocovariance

$$\gamma_X(t,t+\tau) = \mathbf{E}X(t)X(t+\tau) = \mathbf{E}[A\cos 2\pi f_0(t+\Theta)\cdot A\cos 2\pi f_0(t+\tau+\Theta)]$$

$$= \mathbf{E}A^2\cdot\int_0^P \cos 2\pi f_0(t+\theta)\cdot\cos 2\pi f_0(t+\tau+\theta)\,\frac{d\theta}{P}$$

$$= EA^2\frac{1}{2}\Big(\int_0^P \cos 2\pi f_0(t+t+\tau+2\theta)\,\frac{d\theta}{P} + \int_0^P \cos 2\pi f_0(\tau)\,\frac{d\theta}{P}\Big),$$

$$= \frac{\mathbf{E}A^2}{2}\cos 2\pi f_0(\tau),$$

where we used Table 1.1, and the independence of the amplitude A and the phase Θ to split the expectations of the product into the product of the expectations. As a result we see that the autocovariance $\gamma_X(t,t+\tau)$ is just a function of the time lag τ which means the signal is stationary. Thus, the ACvF

$$\gamma_X(\tau) = \frac{\mathbf{E}A^2}{2}\cos(2\pi f_0\tau).$$

Example 5.1.2 (Superposition of Random Harmonic Oscillations) In this example we consider a signal which is a sum of simple harmonic oscillations with frequencies $kf_0, k = 1,2,\ldots,N,$, random amplitudes $A_k, k = 1,2,\ldots,N$, such that $\mathbf{E}A_k^2 < \infty$, and random phases $\Theta_k, k = 1,2,\ldots,N$, uniformly distributed over the corresponding periods. All of the above random quantities are assumed to be independent of each other. In other words,

$$X(t) = \sum_{k=1}^{N} A_k\cos(2\pi k f_0(t+\Theta_k)).$$

In this case one can verify (see Problems and Exercises) that the signal is again stationary and the covariance function is of the form

$$\gamma_X(\tau) = \frac{1}{2}\sum_{k=1}^{N} \mathbf{E}A_k^2 \cos(2\pi k f_0 \tau).$$

Example 5.1.3 (Discrete-Time White Noise) In this example the time is discrete, that is, $t = n = \ldots, -2, -1, 0, 1, 2, \ldots$, and the random signal $W(n)$, has mean zero, and values at different times that are independent (uncorrelated would suffice) and identically distributed; we will denote their common variance by σ_W^2. In other words,

$$\mu_W = 0,$$

and

$$\gamma_W(n, n+\tau) = \mathbf{E}(W(n)W(n+\tau)) = \begin{cases} \sigma_W^2, & \text{if } \tau = 0, \\ 0, & \text{if } \tau \neq 0. \end{cases}$$

Note that the above defined signal is stationary because its autocovariance is indeed a function of only the time lag and can be written in the form

$$\gamma_W(n, n+\tau) = \sigma_W^2 \delta(\tau),$$

where

$$\delta(\tau) = \begin{cases} 1, & \text{if } \tau = 0; \\ 0, & \text{if } \tau \neq 0, \end{cases}$$

is the discrete-time version of Dirac delta-function which is usually called *Kronecker delta*. This kind of signal is called *discrete-time white noise* and it has mean zero, and autocovariance function,

$$\gamma_W(\tau) = \sigma_W^2 \delta(\tau).$$

Observe that in the definition of the white noise we did not specify the distribution of the random quantities $W(n)$. So, in principle, the white noise can have an arbitrary distribution as long as its variance is finite. In practice, the distribution in the white noise model to be employed must be determined from the detailed analysis of the physical phenomenon under consideration (or experimentation and estimation). Figure 5.1 shows a sample discrete-time white noise random signal $W(n), n = 1, 2, \ldots, 50$, with W_n's all distributed uniformly on the interval $[-1/2 + 1/2]$. Hence $\mathbf{E}W_n = 0$, and $\sigma_W^2 = 1/12$.

By the *standard white noise* we will always mean the white noise with variance $\sigma_W^2 = 1$. Thus we can standardize any white noise $W(n)$ by dividing all of its values by its standard deviation σ_W. So, in Example 5.1.1, the white noise $W(n)$ is not standard, but the white noise $W(n)/\sqrt{12}$ is.

Example 5.1.4 (Moving Average of the White Noise) The moving average signal $X(n)$ is obtained from the white noise $W(n)$ with variance σ_W^2 by the "windowing" procedure. The windowing procedure mixes values of the white noise, $W(n)$, $W(n-1), \ldots, W(n-q)$, in the time window of fixed width $q+1$, extending into the past, giving values with different time lags different weights, say, $b_0, b_1, \ldots, b_q$. More precisely,

$$X(n) = b_0 W(n) + b_1 W(n-1) + \ldots + b_q W(n-q).$$

You can interpret the moving average signal as a discrete-time convolution of the white noise with the windowing weight sequence. One immediately obtains that $\mu_X = 0$. Since, for independent random quantities, the variance of the sum is equal to the sum of the variances, the variance

$$\sigma_X^2 = \sigma_W^2 \sum_{i=0}^{q} b_i^2.$$

Calculation of the autocovariance function is little more complicated (see Problems and Exercises) and for now we will carry it out only in the case of the window of width 2, when

$$X(n) = b_0 W(n) + b_1 W(n-1).$$

Then

$$\gamma_X(n, n+\tau) = \mathbf{E} X(n) X(n+\tau)$$

$$= \mathbf{E}\Big(\Big(b_0 W(n) + b_1 W(n-1)\Big)\Big(b_0 W(n+\tau) + b_1 W(n+\tau-1)\Big)\Big)$$

$$= b_0^2 \mathbf{E}(W(n)W(n+\tau)) + b_0 b_1 \mathbf{E}\Big(W(n-1)W(n+\tau)\Big)$$

$$+ b_0 b_1 \mathbf{E}\Big(W(n)W(n+\tau-1)) + b_1^2 \mathbf{E}(W(n-1)W(n+\tau-1))$$

$$= \begin{cases} (b_0^2 + b_1^2)\sigma_W^2, & \text{if } \tau = 0; \\ b_0 b_1 \sigma_W^2, & \text{if } \tau = 1; \\ b_0 b_1 \sigma_W^2, & \text{if } \tau = -1; \\ 0, & \text{if } |\tau| > 1. \end{cases}$$

Since $\gamma_X(n, n+\tau)$ depends only on the time lag τ the moving average signal is stationary. For the sample white noise signal from Fig. 5.1, the moving average signal $X(n) = 2W(n) + 5W(n-1)$ is shown in Fig. 5.2, and its corresponding autocovariance, function

$$\gamma_X(\tau) = \begin{cases} 29/12, & \text{if } \tau = 0; \\ 10/12, & \text{if } \tau = \pm 1; \\ 0, & \text{if } \tau = \pm 2, \pm 3, \ldots, \end{cases}$$

is shown in Fig. 5.3. Compare Figs. 5.2 and 5.3 and note that the moving average operation smoothed out the original white noise signal.

The method of determining ACvF for a moving average signal from Example 5.1.4 can be streamlined using the fact that the ACvF of a standard white noise is the Kronecker delta. This "Kronecker delta calculus" makes it also easy to obtain the ACvF of an arbitrary infinite moving average of the white noise of the form

$$X(n) = \sum_{k=-\infty}^{\infty} b_k W(n-k), \tag{5.1.12}$$

where $W(n)$ is the standard white noise. Since $\mathbf{E}W(n)W(n+\tau) = \delta(\tau)$ which is 0 if $\tau \neq 0$, and 1 if $\tau = 0$, we have automatically that

$$\gamma_X(\tau) = \mathbf{E}\left(\sum_{k=-\infty}^{\infty} b_k W(n-k) \cdot \sum_{l=-\infty}^{\infty} b_l W(n+\tau-l) \right)$$

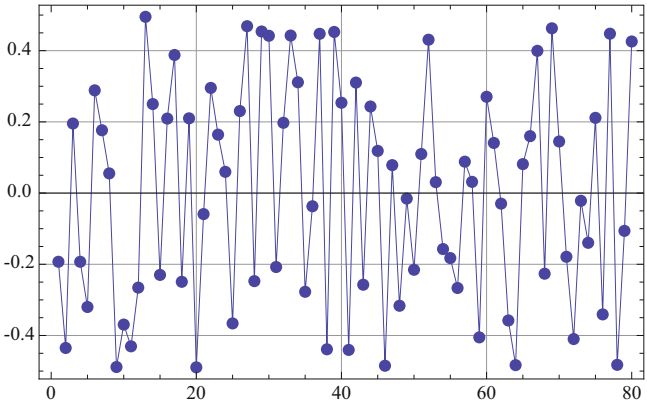


Fig. 5.2 A sample discrete-time white noise random signal $W(n)$, $n = 1, 2, \ldots, 50$, with uniform distribution on the interval $[-1/2, +1/2]$, so that $\sigma_W^2 = 1/12$. For the sake of the clarity of the picture, values of $W(n)$ for consecutive integers n were joined by straight-line segments

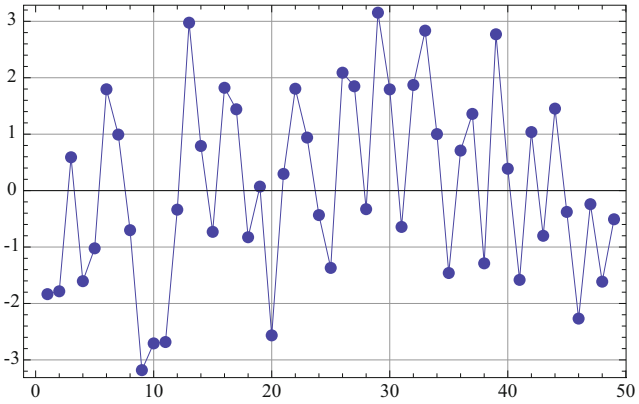


Fig. 5.3 Sample moving average signal $X(n) = 2W(n) + 5W(n-1)$ for the sample white noise shown in Fig. 5.1. Note that the moving average signal appears smoother than the original white noise. The constrained oscillations are a result of nontrivial, although short-term in this example, correlations

$$= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} b_k b_l \mathbf{E}\Big(W(n-k) \cdot W(n+\tau-l)\Big)$$

$$= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} b_k b_l \delta\Big((n+\tau-l) - (n-k)\Big) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} b_k b_l \delta\Big((\tau+k) - l\Big).$$

Since $\delta(\tau - (l-k)) = 1$, if, and only if $l = \tau + k$ (otherwise it is zero), the whole double summation over the whole (k, l) lattice reduces to the single summation on the "diagonal," $l = \tau + k$, and we get the final result

$$\gamma_X(\tau) = \sum_{k=-\infty}^{\infty} b_k b_{k+\tau}. \tag{5.1.13}$$

The variance of such a moving average signal is

$$\sigma_X^2 = \gamma_X(0) = \sum_{k=-\infty}^{\infty} b_k^2,$$

and to assure that it is finite the sequence of coefficients, $\ldots, b_{-1}, b_0, b_1, \ldots$, must be square summable, i.e., the condition $\sum_{k=-\infty}^{\infty} b_k^2 < \infty$ must be satisfied.

Example 5.1.5 (Random Switching Signal) Consider a continuous-time signal $X(t)$ switching back and forth between values $+1$ and -1 at random times. More precisely, the initial value of the signal, $X(0)$, is a random quantity with

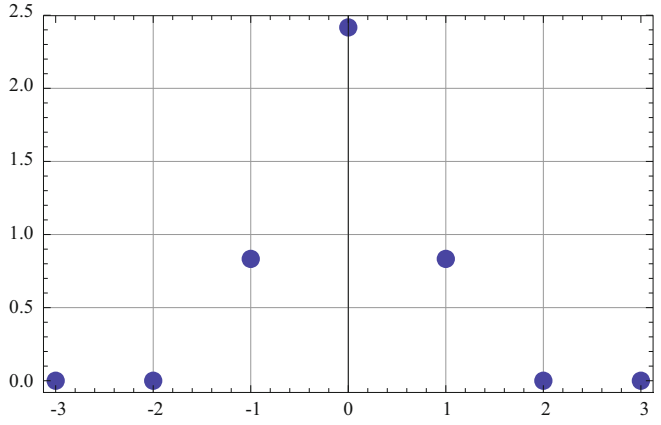


Fig. 5.4 Autocovariance function for the moving average signal $X(n) = 2W(n) + 5W(n-1)$. Note that the values of the signal separated by more that one time unit are uncorrelated

the symmetric Bernoulli distribution, that is $\mathbf{P}(X(0) = \pm 1) = 1/2$, and the interswitching times form a sequence $T_1, T_2, \ldots$, of independent random quantities with the identical standard exponential c.d.f.s:

$$\mathbf{P}(T_i \leq t) = 1 - e^{-t}, \qquad t > 0,$$

of mean one. The initial random value $X(0)$ is assumed to be independent of interswitching times T_i. A typical sample of such a signal is shown in Fig. 5.4.

Calculation of the mean and the autocovariance function of the switching signal depends on the knowledge of the fact that such a random signal can be written in the form

$$X(t) = X(0) \cdot (-1)^{N(t)},$$

where $N(t)$ is the (nonstationary) random signal counting the number of switches up to time t; in particular, $N(0) = 0$. One can prove[2] that $N(t)$ has increments in disjoint time intervals that are statistically independent, with the distributions thereof depending only on the interval's length. More strikingly, these increments must have the Poisson probability distribution with mean equal to the interval's length, that is

$$\mathbf{P}\Big(N(t+\tau) - N(t) = k\Big) = \mathbf{P}\Big(T_1 + \ldots + T_k \leq \tau < T_1 + \ldots + T_{k+1}\Big) = e^{-\tau} \cdot \frac{\tau^k}{k!},$$

[2]See, for example, O. Kallenberg, *Foundations of Modern Probability*, Springer-Verlag 1997.

for any $t, \tau \geq 0$, and $k = 0, 1, 2, \ldots$ Indeed,

$$\mathbf{P}(N(t) \geq k) = \mathbf{P}(T_1 + \ldots + T_k \leq t) = \int_0^t e^{-s} \frac{s^{k-1}}{(k-1)!}\, ds = 1 - e^{-t} \sum_{l=0}^{k} \frac{t^l}{l!},$$

because the p.d.f. of the sum of k independent standard exponential random quantities is

$$f_{T_1+\ldots+T_k}(s) = e^{-s} \frac{s^{k-1}}{(k-1)!}, \qquad s \geq 0,$$

see Problem 5.3.8; the above integral was evaluated by repeated integration by parts.

Armed with this information we can now easily complete calculations of the mean and autocovariance function of the switching signal:

$$\mu_X(t) = \mathbf{E}X(t) = \mathbf{E}X(0) \cdot \mathbf{E}(-1)^{N(t)} = 0,$$

and, for $\tau > 0$,

$$\gamma_X(t, t+\tau) = \mathbf{E}[X(t)X(t+\tau)] = \mathbf{E}X^2(0) \cdot \mathbf{E}\Big[(-1)^{N(t)}(-1)^{N(t+\tau)}\Big]$$

$$= 1 \cdot \mathbf{E}\Big[(-1)^{2N(t)}(-1)^{N(t+\tau)-N(t)}\Big] = \mathbf{E}(-1)^{N(t+\tau)-N(t)}$$

$$\sum_{k=0}^{\infty} (-1)^k \cdot \frac{e^{-\tau}\tau^k}{k!} = e^{-2\tau}.$$

Therefore, the random switching signal $X(t)$ is stationary and, because of the symmetry property of all autocovariance functions, its ACvF

$$\gamma_X(\tau) = e^{-2|\tau|}.$$

Remark 5.1.2 (Transition from a Switching Signal to the Bernoulli White Noise in Continuous Time) Now, let us make the switching model more flexible by permitting the exponential interswitching times $T_1, T_2, \ldots$, to have mean (expected value) $\mu > 0$. That means that the common p.d.f. of T_k's is $f_T(t) = e^{-t/\mu}/\mu$, $t \geq 0$. Recall that in Example 5.1.5, we simply assumed that $\mu = 1$. The corresponding counting, Poisson signal, $N_\mu(t)$, now has the distribution

$$\mathbf{P}(N_\mu(t) = k) = e^{-t/\mu} \frac{(t/\mu)^k}{k!}, \qquad k = 0, 1, 2, \ldots,$$

with expectation $\mathbf{E}N_\mu(t) = t/\mu$. Define the rescaled switching signal

$$X_\mu(t) = \frac{X(0)}{\sqrt{\mu}} \cdot (-1)^{N_\mu(t)},$$

with $X(0)$ independent of $N_\mu(t)$, and $\mathbf{P}(X(0) = \pm 1) = 1/2$, so that the signal $X_\mu(t)$ now switches between the values $+1/\sqrt{\mu}$ and $-1/\sqrt{\mu}$. Repeating the calculation from Example 5.1.5 in the present, general case we obtain the following expression for its ACvF:

$$\gamma_{X_\mu}(t, t+\tau) = \mathbf{E}[X_\mu(t)X_\mu(t+\tau)] = \frac{\mathbf{E}X^2(0)}{\mu} \cdot \mathbf{E}\Big[(-1)^{N_\mu(t)}(-1)^{N_\mu(t+\tau)}\Big]$$

$$= \frac{1}{\mu}\mathbf{E}(-1)^{N_\mu(t+\tau)-N_\mu(t)} = \frac{1}{\mu}\sum_{k=0}^{\infty}(-1)^k \cdot \frac{e^{-\tau/\mu}(\tau/\mu)^k}{k!} = \frac{1}{\mu}e^{-2\tau/\mu},$$

for $\tau \geq 0$. So, the random switching signal $X_\mu(t)$ is stationary and its autocovariance function

$$\gamma_{X_\mu}(\tau) = \frac{1}{\mu}e^{-2|\tau|/\mu}.$$

Now, if we let $\mathbf{E}T = \mu \to 0$, that is if we permit the switching signal to switch more and more often, as the size of the switches increase, then its ACvF converges to the Dirac-delta impulse $\delta(t)$, see Fig. 2.9. So, we can think about the limit of the switching signals, with $\mu \to 0$, as a continuous-time white noise; it switches between $+\infty$ and $-\infty$ "infinitely often" in any finite time interval. Indeed, for any t, t_0, the expected number of switches in the time interval $[t_0, t_0 + t]$,

$$\mathbf{E}(N_\mu(t+t_0) - N_\mu(t_0)) = \frac{t}{\mu} \to \infty, \qquad \text{as} \qquad \mu \to 0.$$

Example 5.1.6 (Solution of a Stochastic Difference Equation) Consider a stochastic difference equation

$$X(n) = \alpha X(n-1) + \beta W(n), \qquad n = -2, -1, 0, 1, 2, \ldots,$$

where $W(n)$ is a standard discrete-time white noise with $\sigma_W^2 = 1$. Observe that the above system, rewritten in the form

$$\frac{X(n) - X(n-1)}{\Delta n} = (\alpha - 1)X(n-1) + \beta W(n), \qquad n = -2, -1, 0, 1, 2, \ldots,$$

can be viewed as a discrete-time version of the stochastic differential equation

$$dX(t) = (\alpha - 1)X(t)\,dt + \beta W(t)dt,$$

where $W(t)$ represents the continuous-time version of the white noise to be discussed in later chapters and mentioned in Remark 5.1.2.

The solution of the above stochastic difference equation can be found by recursion. So,

$$X(n) = \alpha(\alpha X(n-2) + \beta W(n-1)) + \beta W(n)$$

$$= \alpha^2 X(n-2) + \alpha\beta W(n-1) + \beta W(n) = \dots$$

$$= \alpha^l X(n-l) + \sum_{k=0}^{l-1} \alpha^k \beta W(n-k).$$

for any $l = 1, 2, \dots$. Assuming that $|\alpha| < 1$ and that $X(n-k)$ remain bounded, the first term $\alpha^k X(n-k) \to 0$ as $k \to \infty$. In that case the second term converges to the infinite sum and the solution is of the form

$$X(n) = \beta \sum_{k=0}^{\infty} \alpha^k W(n-k).$$

This is the special form of the general moving average signal appearing in (4.1.12), with the windowing sequence,

$$c_k = \begin{cases} \beta\alpha^k, & \text{for} k = 0, 1, 2, \dots; \\ 0, & \text{for} k = -1, -2, \dots. \end{cases}$$

Hence its autocovariance function, see ((4.1.12)–(4.1.13)); also, see Problem 4.3.4,

$$\gamma_X(\tau) = \sum_{k=-\infty}^{\infty} c_k c_{\tau+k} = \beta^2 \sum_{k=0}^{\infty} \alpha^k \alpha^{\tau+k} = \beta^2 \frac{\alpha^\tau}{1-\alpha^2},$$

for positive $\alpha < 1$.

Remark 5.1.3 (Autoregressive and ARMA Processes) There are two other classes of stationary time series which are very important, both theoretically and in applications. One is the general autoregressive process $X(n)$ satisfying the equation,

$$X(n) = c + \sum_{k=1}^{p} \alpha_k X(n-k) + \beta W(n),$$

where $|\alpha_k| \leq 1$, and $W(n)$ is the discrete time white noise.

Another example is the ARMA (Autoregressive Moving Average) process satisfying a more general equation,

$$X(n) = c + W(n) + \sum_{k=1}^{p} \alpha_k X(n-k) + \sum_{k=1}^{q} \beta_k W(n-k).$$

Example 5.1.7 (Using Moving Averages to Filter Noise out of Signal) Consider a signal of the form

$$X(n) = \sin(0.02n) + W(n),$$

where $W(n)$ is the white noise considered in Example 5.1.3, and let $Y(n)$ be a moving average of signal $X(n)$ with the windowing sequence $b_0 = b_1 = b_2 = b_3 = b_4 = 1/5$, that is,

$$Y(n) = \frac{1}{5}X(n) + \frac{1}{5}X(n-1) + \frac{1}{5}X(n-2) + \frac{1}{5}X(n-3) + \frac{1}{5}X(n-4).$$

The values of both signals $X(n)$ and $Y(n)$ for time instants $n = 1, 2, \ldots, 750$ are shown in Fig. 5.5. Clearly, the moving average operation filtered some of the white noise out of the original signal and the transformed signal appears smoother (Fig. 5.6).

Remark 5.1.4 (ACvF for Complex-Valued Signals) For complex values stationary signals $X(t)$ the definition of the autocovariance function has to be adjusted so that the value of ACvF at $t = 0$ remains the variance of the signal which must be a nonnegative number. That is why taking the expectation of the simple product of values of the signal separated by the time lag τ will not do; a square of a complex number is in general a complex number. For that reason, for complex-

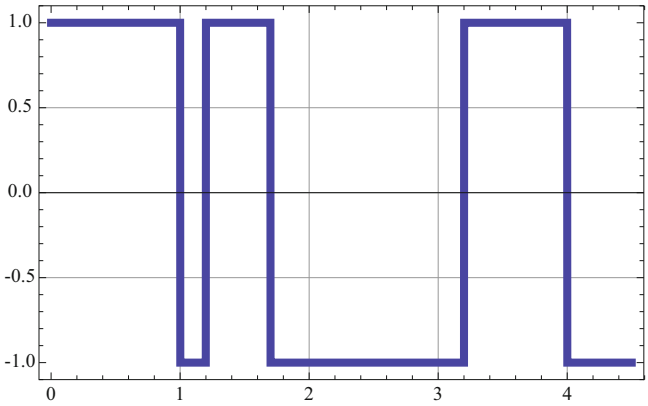


Fig. 5.5 A sample of the random switching signal from Example 5.1.4. The values are ± 1 and the initial value is $+1$. The interswitching times are independent and have an exponential c.d.f. of mean one

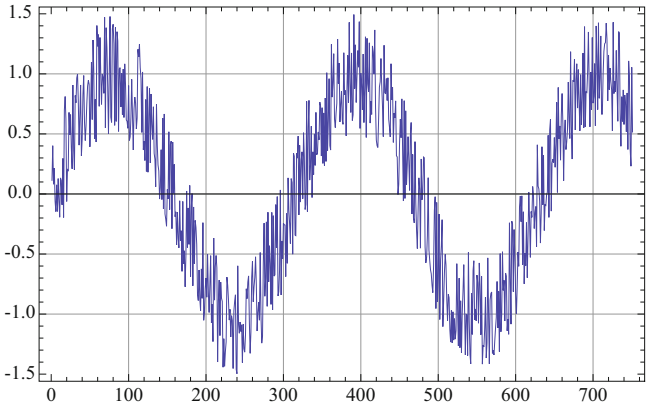


Fig. 5.6 (Top) Signal $X(n)$ from Example 5.1.6 containing a nonrandom harmonic component plus a random white noise. (Bottom) The same signal after a smoothing, moving average operation filtered out some of the white noise. The figure shows values of both signals for times $n = 1, 2, \ldots, 750$

valued stationary signals, the autocovariance function is defined, in the zero-mean case, by the formula

$$\gamma_X(\tau) = \mathbf{E}[X^*(t) \cdot X(t+\tau)], \tag{5.1.14}$$

where the asterisk denotes the complex conjugate. In this case, of course, the variance

$$\operatorname{Var} X(t) = \mathbf{E}[X^*(t) \cdot X(t)] = \mathbf{E}|X(t)|^2 = \gamma_X(0) \geq 0.$$

Note that in the complex-valued case, ACvF is not necessarily an even function of the time lag τ. However, we do have the equality,

$$\gamma_X(-\tau) = \mathbf{E}[X^*(t) \cdot X(t-\tau)] = \Big(\mathbf{E}[X^*(t-\tau) \cdot X(t)]\Big)^* = \gamma_X^*(\tau). \tag{5.1.15}$$

Example 5.1.8 (Simple Complex Random Harmonic Oscillation) Consider a complex-valued random signal represented by a simple complex exponential with a random, possibly complex-valued, amplitude A of zero mean, $\mathbf{E}A = 0$, and finite variance $\sigma_A^2 = \mathbf{E}|A|^2$:

$$X(t) = A \cdot e^{j2\pi f_0 t}.$$

Then, clearly, $\mathbf{E}\,X(t) = \mathbf{E}\,A \cdot e^{j2\pi f_0 t} = 0$, and

$$\gamma_X(\tau) = \mathbf{E}[X^*(t) \cdot X(t+\tau)] = \mathbf{E}|A|^2 \cdot e^{-j2\pi f_0 t} \cdot e^{j2\pi f_0 (t+\tau)} = \sigma_A^2 \cdot e^{j2\pi f_0 \tau}.$$

This result is analogous to the result for the simple random real-valued oscillation introduced at the beginning of this section. However, in the complex-valued case no random phase is needed to produce a stationary signal.

Example 5.1.9 (Superposition of Simple Complex-Valued Random Harmonic Oscillations) As in the real-valued case in Example 5.1.2 we can consider a superposition of simple complex-valued random harmonic oscillations. Let $A_1, A_2, \ldots, A_n$ be a sequence of independent (or just uncorrelated, possibly, complex-valued) random amplitudes with $\mathbf{E}A_k = 0$, and finite variance $\sigma A_k^2 = \mathbf{E}|A_k|^2$. Set

$$X(t) = \sum_{k=1}^{n} A_k \cdot e^{j2\pi f_k t},$$

where $f_1, f_2, \ldots, f_n$, is a sequence of different frequencies. Then, again,

$$\mathbf{E}\, X(t) = \mathbf{E} \sum_{k=1}^{n} A_k \cdot e^{j2\pi f_k t} = \sum_{k=1}^{n} \mathbf{E}(A_k) \cdot e^{j2\pi f_k t} = 0,$$

and

$$\gamma_X(\tau) = \mathbf{E}[X^*(t) \cdot X(t+\tau)] = \mathbf{E}\left(\sum_{k=1}^{n} A_k^* \cdot e^{-j2\pi f_k t} \cdot \sum_{l=1}^{n} A_l \cdot e^{j2\pi f_l (t+\tau)} \right)$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{n} \mathbf{E}(A_k^* A_l) \cdot e^{-j2\pi (f_k - f_l)t} \cdot e^{j2\pi f_l \tau} = \sum_{k=1}^{n} \mathbf{E}|A_k|^2 \cdot e^{j2\pi f_k \tau},$$

because, for different k, l, the covariance $\mathbf{E}(A_k^* A_l) = \mathbf{E}(A_k^*)\mathbf{E}(A_l) = 0$.

5.2 Estimating the Mean and the Autocovariance Function, Ergodic Signals

If one can obtain multiple independent samples of the same random stationary signal, then the estimation of its parameters, the mean value and the autocovariance function, can be based on procedures described in Sect. 4.6. However, very often, the only available information is a single but, perhaps, long (timewise) sample of the signal; think here about the historical temperature records at a given location, Dow-Jones stock market index daily quotations over the past 10 years, or measurements of the sun spot activity over a period of time; these measurements cannot be independently repeated. Estimation of the mean and the autocovariance function of a stationary signal $X(t)$ based on its single sample is a delicate matter because the

standard Law of Large Numbers and the Central Limit Theorem cannot be applied. So one has to proceed with caution, as we now illustrate.

Estimation of the Mean μ_X If a stationary signal $X(t)$ is sampled with the sampling interval T, that is, the known values are

$$X(0), X(T), X(2T), \ldots, X(NT), \ldots,$$

then the obvious candidate for an estimator $\hat{\mu}_X$ of the signal's mean μ_X is

$$\hat{\mu}_X(N) = \frac{1}{N} \sum_{i=0}^{N-1} X(iT).$$

This estimator is easily seen to be unbiased as

$$\mathbf{E}[\hat{\mu}_X(N)] = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{E}[X(iT)] = \mu_X. \tag{5.2.1}$$

To check whether the estimator $\hat{\mu}_X(N)$ converges to μ_X as the observation interval $NT \to \infty$, that is to check the estimator's consistency, we will take a look at the estimation error in the form of the mean-square distance (variance) between $\hat{\mu}_X(N)$ and μ_X,

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_X(N)) &= \mathbf{E}[(\hat{\mu}_X - \mu_X)^2] \\
&= \frac{1}{N^2} \mathbf{E}\Big[\sum_{i=0}^{N-1} (X(iT) - \mu_X) \sum_{k=0}^{N-1} (X(kT) - \mu_X) \Big] \\
&= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} \gamma_X(iT, kT) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} \gamma_X((i-k)T) \\
&= \frac{\sigma_X^2}{N} + \frac{2}{N} \sum_{k=1}^{N-1} (1 - \frac{k}{N}) \gamma_X(kT).
\end{aligned} \tag{5.2.2}$$

So the error of replacing the true value μ_X by the estimator $\hat{\mu}_X$ will converge to zero, as $N \to \infty$, only if the sum in (5.2.2) increases slower[3] than N , i.e.,

[3]Here we use Landau's asymptotic notation: we write that $f(x) = o(g(x))$, as $x \to x_0$, and say that $f(x)$ is little "oh" of $g(x)$ at x_0, if $\lim_{x \to x_0} f(x)/g(x) = 0$.

$$\sum_{k=0}^{N-1}(1-\frac{k}{N})\gamma_X(kT) = o(N), \qquad \text{as} \qquad N \to \infty. \tag{5.2.3}$$

So, for example, if the covariance function $\gamma_X(\tau)$ vanishes outside a finite interval, as was the case for finite moving averages in Example 5.1.2, then $\hat{\mu}_X$ is a consistent estimator for μ_X.

Example 5.2.1 (Consistency of the Estimator $\hat{\mu}_X$ for Solutions of Discrete-Time Stochastic Difference Equations) Consider the solution $X(n)$ of the stochastic difference equation from Example 5.1.5. Its autocovariance function was found to be of the form

$$\gamma_X(\tau) = \beta^2 \frac{|\alpha|^\tau}{1-\alpha^2}, \qquad |\alpha| < 1.$$

Since it decays exponentially as $\tau \to \infty$, the sum in (5.2.2) converges and condition (5.2.3) is satisfied. The mean-square error of replacing μ_X by the estimator $\hat{\mu}_X$ can now be controlled:

$$\mathrm{Var}(\hat{\mu}_X(N)) = \mathbf{E}[(\hat{\mu}_X - \mu_X)^2] = \frac{\gamma_X(0)}{N} + \frac{2}{N}\sum_{k=0}^{N-1}(1-\frac{k}{N})\beta^2\frac{|\alpha|^k}{1-\alpha^2}$$

$$\leq \frac{\beta^2}{N(1-\alpha^2)}\left(1+2\sum_{k=0}^{N-1}|\alpha|^k\right) = \frac{\beta^2(3-|\alpha|-2|\alpha|^N)}{N(1-\alpha^2)(1-\alpha)}.$$

Estimation of the Autocovariance Function $\gamma_X(\tau)$. For simplicity's sake assume that $\mu_X = 0$, the sampling interval $T = 1$, the signal is real-valued, and that observations, $X(0), \ldots, X(N)$, are given. The natural candidate for an estimator of the autocovariance function $\gamma_X(\tau) = \mathbf{E}X(0)X(\tau)$ is the time average:

$$\hat{\gamma}_X(\tau; N) = \frac{1}{N-\tau}\sum_{k=0}^{N-\tau-1} X(k)X(k+\tau). \tag{5.2.4}$$

It is an unbiased estimator since, for each fixed time lag, τ,

$$\mathbf{E}[\hat{\gamma}_X(\tau, N)] = \frac{1}{N-\tau}\mathbf{E}\Big[\sum_{k=0}^{N-\tau-1} X(k)X(k+\tau)\Big]$$

$$= \frac{1}{N-\tau}\sum_{k=0}^{N-\tau-1}\gamma_X(\tau) = \gamma_X(\tau).$$

One can also prove that if $\gamma_X(\tau) \to 0$ sufficiently fast,[4] as $n \to \infty$, and if $\gamma_X(0) = \sigma_X^2 < \infty$, then the mean-square distance from $\hat{\gamma}_X(\tau; N)$ to $\gamma_X(\tau)$ decreases to 0 as $N \to \infty$. In other words, the estimator (5.2.4) is consistent.

Remark 5.2.1 (Ergodicity) If the estimator $\hat{\mu}_X$ is unbiased and consistent, that is

$$\mathbf{E}\,\hat{\mu}_X(N) = \mu_X, \quad \text{and} \quad \mathrm{Var}(\hat{\mu}_X(N)) \to 0,$$

as $N \to \infty$, then one often says that the signal is *ergodic in the mean*. Note that, in general, this does not imply that for every sample path of the random signal the estimator converges to the estimated parameter. To guarantee that, for a general test function g, the time averages

$$\frac{g(X(1)) + g(X(2)) + \cdots + g(X(N)}{N},$$

converge to $\mathbf{E}g(X(1))$, as $N \to \infty$, for (almost) every sample path of the random signal, stronger ergodicity and stricter stationarity assumptions are needed. A.I. Khinchin proved[5] (in the context of statistical mechanics) that a decay of the autocorrelation function to zero is a sufficient condition for ergodicity. A more detailed analysis of the ergodic behavior of stationary time series can be found in the above quoted books by M. Denker and W.A. Woyczyński, and by P.J. Brockwell and R.A. Davis.

Remark 5.2.2 (Confidence Intervals) Under fairly weak assumptions one can show that the asymptotic distributions ($N \to \infty$) of the suitably rescaled estimators $\hat{\mu}_X(N)$, $\hat{\gamma}_X(\tau; N)$, are asymptotically normal. Thus the confidence intervals for them can be constructed following the ideas discussed in Sect. 4.6.

Example 5.2.2 (Estimated Autocorrelation Functions of EEG Signals) Figure 5.7 shows two samples of the central channel recording for a full-term neonate EEG (see Fig. 5.1 for a sample of the full 21-channel EEG). The duration of each of the samples was 3 min, and the signals were sampled at 64 Hz. The data in the top picture were recorded during the quiet sleep stage, and in the bottom picture—during the active sleep stage.

The estimated autocorrelation functions (ACFs) (not ACvFs!) for both signals were then calculated using formula (5.2.4) and are shown in Fig. 5.8. The example is taken from A. Piryatinska's Case Ph.D. Dissertation mentioned already in Sect. 5.1.

Note that the ACF of the active sleep signal decays much slower than the ACF of the quiet sleep, indicating the longer-range dependence structure of former. Information of the rate of decay in EEG ACFs can then be used to automatically

[4]For a thorough exposition of these issues see, for example, P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York 1991.

[5]See A.I. Khinchin, *Mathematical Foundation of Statistical Mechanics*, Dover Publications, Inc., New York, 1949, p. 68.

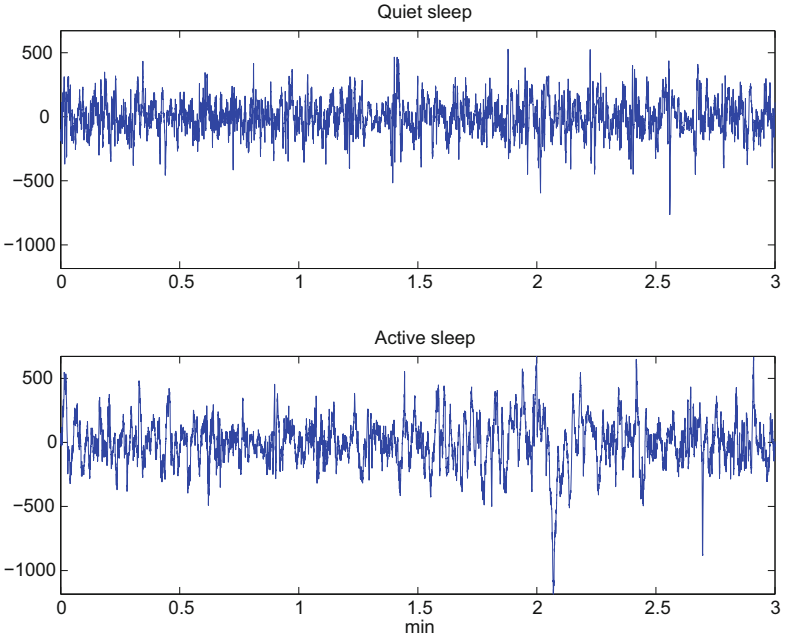


Fig. 5.7 (Top) Three-minute recording of the central channel EEG for an infant in a quiet sleep stage. (Bottom) Analogous recording for an active sleep stage

classify stationary segments of the EEG signals as those corresponding to different sleep stages recognized by pediatric neurologists.

5.3 Problems and Exercises

1* Consider a random signal

$$X(t) = \sum_{k=0}^{n} A_k \cos\Big(2\pi k f_0 (t + \Theta_k)\Big),$$

where $A_0, \Theta_0, \ldots, A_n, \Theta_n$ are independent random variables of finite variance, and $\Theta_0, \ldots, \Theta_n$ are independent, independent of As and uniformly distributed on the time interval $[0, P = 1/f_0]$. Is this signal stationary? Find its mean and autocovariance functions.

2* Consider a random signal

$$X(t) = A_1 \cos 2\pi f_0 (t + \Theta_0),$$

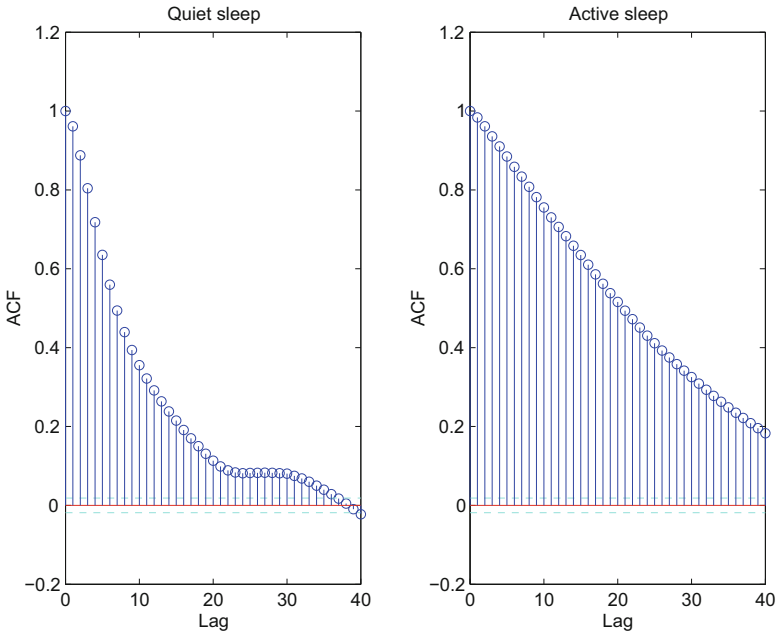


Fig. 5.8 (Left) Estimated autocovariance function (ACF) for the quiet sleep EEG signal from Fig. 5.7. (Right) Analogous estimated ACF for the active sleep stage

where A_1, Θ_0, are independent random variables, and Θ_0 is uniformly distributed on the time interval $[0, P/3 = 1/(3f_0)]$. Is this signal stationary? Is the signal $Y(t) = X(t) - \mathbf{E}X(t)$ stationary? Find its mean and autocovariance functions.

3 Find the mean and autocovariance functions of the discrete-time signal

$$Y(n) = 3W(n) + 2W(n-1) - W(n-2),$$

where $W(n), n = \ldots, -2, -1, 0, 1, 2, \ldots$, is the discrete-time white noise with $\sigma_W^2 = 4$, that is

$$EW(n) = 0$$

and

$$\mathbf{E}(W(k)W(n)) = \begin{cases} 4, & \text{if} n-k=0; \\ 0, & \text{if} n-k \neq 0. \end{cases}$$

4 Consider a general complex-valued moving average signal

$$X(n) = \sum_{k=-\infty}^{\infty} c_k W_{n-k},$$

where c_k is a complex-valued "windowing" sequence. Determine a condition on the windowing sequence that would guarantee that $X(n)$ has finite variance. $W(n)$ is the standard white noise signal with mean zero, and $\gamma_W(n) = \delta(n)$.

5 *Simulation of a discrete-time white noise with an arbitrary probability distribution.* Formula (4.1.11), $F_Y(y) = F_X(g^{-1}(y))$, describes the c.d.f. $F_Y(y)$ of the random quantity $Y = g(X)$ in terms of the c.d.f. $F_X(x)$ of the random quantity X, and a strictly increasing function $g(x)$. It also permits construction of an algorithm to produce random samples from any given probability distribution provided a random sample uniformly distributed on the interval [0, 1] is given. The latter can be obtained by using the random number generator in any computing platform, see Problem 1.4.14.

Let U be a uniformly distributed on [0,1] random quantity U with the c.d.f.

$$F_U(u) = u, \qquad 0 \le u \le 1. \tag{5.3.1}$$

Then, for a given c.d.f. $F_Z(z)$, the random quantity $Z = F_Z^{-1}(U)$, where $F_Z^{-1}(u)$ is the function inverse to $F_Z(z)$ (that is, a solution of the equation $u = F_Z(F_Z^{-1}(u))$) has the c.d.f. $F_Z(z)$. Indeed, a simple calculation, using (4.3.1), shows that

$$\mathbf{P}(F_Z^{-1}(U) \le z) = \mathbf{P}(U \le F_Z(z)) = F_Z(z),$$

because $0 \le F_Z(z) \le 1$. So, for example, if the desired c.d.f. is exponential, with $F_Z(z) = 1 - e^{-z}, z \ge 0$, then $F_Z^{-1}(u) = -\ln(1-u), 0 \le u \le 1$, and the random quantity $Z = -\ln(1-U)$ has the above exponential c.d.f.

The general simulation algorithm is thus as follows:

(i) Choose the sample size N, and produce a random sample, $u_1, u_2, \ldots, u_N$, uniformly distributed on [0,1];
(ii) Calculate the inverse function $F_Z^{-1}(u)$;
(iii) Substitute the random sample, $u_1, u_2, \ldots, u_N$, into $F_Z^{-1}(u)$ to obtain the random sample

$$z_1 = F_Z^{-1}(u_1), z_2 = F_Z^{-1}(u_2), \ldots, z_N = F_Z^{-1}(u_N),$$

which has the desired c.d.f. $F_Z(z)$.

Use the above algorithm, and Problem 1.4.15, to produce and plot examples of the white noise $W(n)$ with:

(a) The standard Gaussian $N(0, 1)$ p.d.f.
(b) The double exponential p.d.f. $f_W(w) = e^{-|w|}/2$. Be careful as its c.d.f. has a different analytic expression for positive and negative w's.
(c) The p.d.f., $f_W(w) = \sqrt{2}(\pi(1 + w^4))^{-1}$. Check that the variance is finite in this case. Start with a calculation of the corresponding c.d.f.s; a symbolic manipulation platform such as *Mathematica* is going to be great help here. Check the result graphically by plotting the histograms of the random samples against the theoretical p.d.f.s.

6 *Simulations of stationary random signals.* Using the algorithm from the above Problem 5, produce simulations of stationary signals from Examples 5.1.4 and 5.1.6, using both uniformly distributed white noise and the white noises constructed in parts (a), (b), and (c) of the above problem. Experiment with these simulations by varying parameters in the above models, and changing the length of the sample of the produced discrete-time random signals.

7 Using the procedures described in Sect. 5.2, estimate the means and the autocovariance functions (ACvF) for sample signals obtained in simulations in Problem 6. Then, compare graphically the estimated and the theoretical ACvFs.

8* Show that if $X_1, X_2, \ldots, X_n$ are independent, exponentially distributed random quantities with identical p.d.f.s e^{-x}, $x \geq 0$, then their sum $Y_n = X_1 + X_2 + \ldots + X_n$ has the p.d.f. $e^{-y}y^{n-1}/(n-1)!$, $y \geq 0$. Use the technique of characteristic functions (Fourier transforms) from Chap. 3. The random quantity Y_n is said to have the Gamma probability distribution with parameter n. Thus the Gamma distribution with parameter one is just the standard exponential distribution, see Example 5.1.4. Produce plots of Gamma p.d.f.s with parameters $n = 2, 5, 20$, and 50. Comment on what you observe as n increases.

Chapter 6
Power Spectra of Stationary Signals

Abstract The Fourier transform $X(f)$ of the sample paths of a stationary, real-valued random signal $X(t)$ does not exist in the usual sense and analysis of the spectral contents of such signals requires a different, more subtle approach which has to rely on the concept of the *mean power* of the random signal. Only then we can investigate how it is distributed over different frequencies. The question is, of course, of fundamental importance in practical applications as real-life signal processing devices such as measuring instruments, amplifiers, and antennas transmit different frequencies with different attenuation.

6.1 Mean Power of a Stationary Signal

For stationary signals with periodic sample paths, like the superpositions of simple harmonic oscillations with random amplitudes discussed in Examples 5.1.2 and 5.1.9, the concept of the mean power is a straightforward adaptation of the power concept for periodic nonrandom signals:

$$\mathbf{E}(\mathbf{PW}_X) = \mathbf{E}\left(\frac{1}{P}\int_0^P |X(t)|^2\,dt\right) = \frac{1}{P}\int_0^P \mathbf{E}|X(t)|^2\,dt = \sigma_X^2.$$

Note that $\mathbf{PW}_X$ itself is here a random quantity. Hence, in particular, in Example 5.1.9, where

$$X(t) = \sum_{k=1}^{n} A_k \cdot e^{j2\pi(kf_0)t},$$

with $P = 1/f_0$, we have

$$\mathbf{E}(\mathbf{PW}_X) = \sigma_X^2 = \gamma_X(0) = \sum_{k=1}^{n} \mathbf{E}|A_k|^2,$$

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_6

and the last expression provides a clear description of how mean power is distributed over different component frequencies of the signal's sample paths; the *power spectrum* in this case is discrete and the mean power carried by the frequency f_k is equal to $\mathbf{E}|A_k|^2$.

For general stationary signals the situation is more complicated. Mean energy $\mathbf{E}(\mathbf{EN}_X)$ of a stationary signal $X(t)$ over the whole time line, that is the expected value of energy, is infinite. Indeed, using the linearity property of expectations we can interchange the order of taking the mean and the integration to obtain that

$$\mathbf{E}(\mathbf{EN}_X) = \mathbf{E}\int_{-\infty}^{\infty} X^2(t)\,dt = \int_{-\infty}^{\infty} \mathbf{E}(X^2(t))\,dt = \int_{-\infty}^{\infty} \sigma_X^2\,dt = \infty. \tag{6.1.1}$$

However, the mean power $\mathbf{E}(\mathbf{PW}_X)$ of a stationary signal, taken as a limit of mean power over finite but expanding time intervals, is always finite since

$$\mathbf{E}(\mathbf{PW}_X) = \mathbf{E}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} X^2(t)\,dt = \sigma_X^2 < \infty. \tag{6.1.2}$$

To find the distribution of mean power $\mathbf{E}(\mathbf{PW}_X)$ over different frequencies f we will consider a windowed signal,

$$X_T(t) = \begin{cases} X(t), & \text{for} |t| \le T; \\ 0, & \text{otherwise,} \end{cases} \tag{6.1.3}$$

that is, the original signal restricted to the time window, $-T \le t \le T$, of duration $2T$. Then, with the well-defined Fourier transform of the windowed signal defined by the equality,

$$X_T(f) = \int_{-\infty}^{\infty} X_T(t)e^{-j2\pi ft}\,dt = \int_{-T}^{T} X(t)e^{-j2\pi ft}\,dt,$$

we can express the mean power of the original signal by the formula,

$$\mathbf{E}[\mathbf{PW}_X] = \mathbf{E}\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} X^2(t)\,dt\right]$$

$$= \mathbf{E}\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty} X_T^2(t)\,dt\right] = \mathbf{E}\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty} |X_T(f)|^2\,df\right]$$

$$= \int_{-\infty}^{\infty}\lim_{T\to\infty}\frac{\mathbf{E}|X_T(f)|^2}{2T}\,df,$$

where the Parseval Equality (see Sect. 2.4) was used in the second line of the above calculation. Denoting

$$S_X(f) := \lim_{T\to\infty} \frac{\mathbf{E}|X_T(f)|^2}{2T}, \tag{6.1.4}$$

the mean power has the representation

$$\mathbf{E}(\mathbf{PW}_X) = \sigma_X^2 = \int_{-\infty}^{\infty} S_X(f)\, df. \tag{6.1.5}$$

The function $S_X(f)$ is called the *power spectral density* or, simply, *power spectrum* of the stationary signal $X(t)$. It shows how the mean power $\mathbf{PW}_X$ of the random stationary signal $X(t)$ is distributed over different frequencies $f, -\infty < f < \infty$. The mean power concentrated in a frequency band, $f_1 < f < f_2$, is then given by the integral

$$\mathbf{PW}_X[f_1, f_2] = \int_{f_1}^{f_2} S_X(f)\, df.$$

6.2 Power Spectrum and Autocovariance Function

What makes the power spectrum $S_X(f)$ a practical tool in the analysis of random stationary signals is the fact that it is simply the Fourier transform of the signal's autocovariance function $\gamma_X(t)$. In other words,

$$S_X(f) = \int_{-\infty}^{\infty} \gamma_X(t) e^{-j2\pi ft} dt. \tag{6.2.1}$$

This fundamental property can be easily verified by direct calculation. Indeed,

$$\begin{aligned}
S_X(f) &= \lim_{T\to\infty} \frac{\mathbf{E}|X_T(f)|^2}{2T} = \lim_{T\to\infty} \frac{\mathbf{E}(X_T^*(f)X_T(f)}{2T} \\
&= \lim_{T\to\infty} \frac{1}{2T} \mathbf{E}\left[\int_{-T}^{T} X^*(t) e^{2\pi jft} dt \int_{-T}^{T} X(s) e^{-2\pi jfs} ds\right] \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T}\int_{-T}^{T} \mathbf{E}[X^*(t)X(s)] e^{-2\pi jf(s-t)} dt\, ds \\
&= \lim_{T\to\infty} \frac{1}{2T} \left[\int_{-T}^{T} \left(\int_{s-T}^{s+T} \gamma_X(\tau) e^{-2\pi jf\tau}\, d\tau\right) ds\right] \\
&= \int_{-\infty}^{\infty} \gamma_X(\tau) e^{-2\pi jf\tau} d\tau.
\end{aligned}$$

Given the properties of the Fourier transform, we also immediately obtain that the autocovariance $\gamma_X(\tau)$ of signal $X(t)$ is the inverse Fourier transform of the power

spectrum $S_X(f)$:

$$\gamma_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau}\, df. \tag{6.2.2}$$

Remark 6.2.1 (What Kind of Functions Can Serve as Autocovariance Functions of Stationary Signals?) Although any integrable nonnegative function,

$$S(f) \geq 0, \int_{-\infty}^{\infty} S(f)\, df < \infty,$$

can serve as a power spectrum of some stationary signal, the above formula (6.2.2) shows that for $\gamma(t)$ to be an autocovariance function of a stationary process it must be the inverse Fourier transform of a nonnegative integrable function $S(f)$. This turns out to be a very restrictive condition. In particular, it forces $\gamma(t)$ to satisfy the following *positive-definiteness condition*:

> For any positive integer N, any real number $t_1, \ldots, t_N$, and any complex numbers $z_1, \ldots, z_N$, the quadratic form
>
> $$\sum_{n=1}^{N}\sum_{k=1}^{N} \gamma(t_n - t_k) z_n z_k^* \geq 0.$$

Indeed, since $S(f) \geq 0$,

$$\sum_{n=1}^{N}\sum_{k=1}^{N} \gamma(t_n - t_k) z_n z_k^* = \sum_{n=1}^{N}\sum_{k=1}^{N} \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f(t_n - t_k)}\, df\, z_n z_k^*.$$

$$= \int_{-\infty}^{\infty} S_X(f) \sum_{n=1}^{N}\sum_{k=1}^{N} \left(z_n e^{j2\pi f t_n}\right) \cdot \left(z_k e^{j2\pi f t_k}\right)^*\, df.$$

$$= \int_{-\infty}^{\infty} S_X(f) \left|\sum_{n=1}^{N} z_n e^{j2\pi f t_n}\right|^2\, df \geq 0.$$

Actually the positive-definiteness condition is *necessary and sufficient* of a function to be an ACvF. This result is known as the *Bochner Theorem*. The practical lesson is that one cannot pick examples of ACvF 's off the top of ones head. There are numerous criteria guaranteeing that a given function actually is positive definite. For example, one can prove that if $\gamma(\tau)$ is even, and decreasing and convex on the positive half-line (like $\gamma(t) = e^{-|\tau|}$), then it is positive definite, see bibliography on Fourier analysis provided at the end of this book.

Estimation of the Power Spectrum $S_X(f)$ For simplicity's sake assume that the signal is real-valued, and that the observations $X(0), \ldots, X(N)$ are made at discrete

sampling times, $t = 0, 1, 2, \ldots, N$. To estimate the spectrum the natural way to proceed is to replace the theoretical ACvF $\gamma_X(\tau)$ in (6.2.1) by the estimated ACvF $\hat{\gamma}_X(\tau; N)$ given by the formula (5.2.4) and replacing the integral by the finite sum. This yields the estimator:

$$\hat{S}_X(f; N) = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}_X(|\tau|; N) e^{-j2\pi f\tau}.$$

A direct discretization of the defining formula (6.1.4) immediately gives another estimator for the power spectrum,

$$I_N(f) := \frac{1}{N} \left| \sum_{n=1}^{N} X(n) e^{-j2\pi fn} \right|^2. \tag{6.2.3}$$

For large N, $\hat{S}_X(f; N) \approx I_N(f)$ (see Sect. 10.2), and the random quantity $I_N(f)$ is usually called the *periodogram* of the sampled signal $X(t)$ based on a sample of size N.

Observe that if we have a concrete, discrete-time sample,

$$X(1) = x_1, \ldots, X(N) = x_N,$$

of the signal $X(t)$, then the (nonrandom) sum inside the modulus of the periodogram formula,

$$\sum_{n=1}^{N} x_n e^{-j2\pi fn},$$

is a finite Fourier (complex-valued) trigonometric polynomial with coefficients $x_1, \ldots, x_N$. It is a periodic function of f with period $P = 1$, so the periodogram $I_N(f)$ needs to be studied only for f in the interval $[0, 1]$ (or any other interval of length 1, such as, e.g., $[-1/2, +1/2]$). In view of the Parseval formula (2.1.12) for Fourier series,

$$\int_0^1 I_N(f)\, df = \int_0^1 \frac{1}{N} \left| \sum_{n=1}^{N} x_n e^{-j2\pi fn} \right|^2 df = \frac{1}{N} \sum_{n=1}^{N} x_n^2.$$

The expression on the right is, of course, the average power (the energy per unit time) of the sample signal $x_1, \ldots, x_N$,. Thus the above formula shows that, indeed, the periodogram $I_N(f)$ gives the correct distribution of the average power of the sample signal over the frequencies $f \in [0, 1]$.

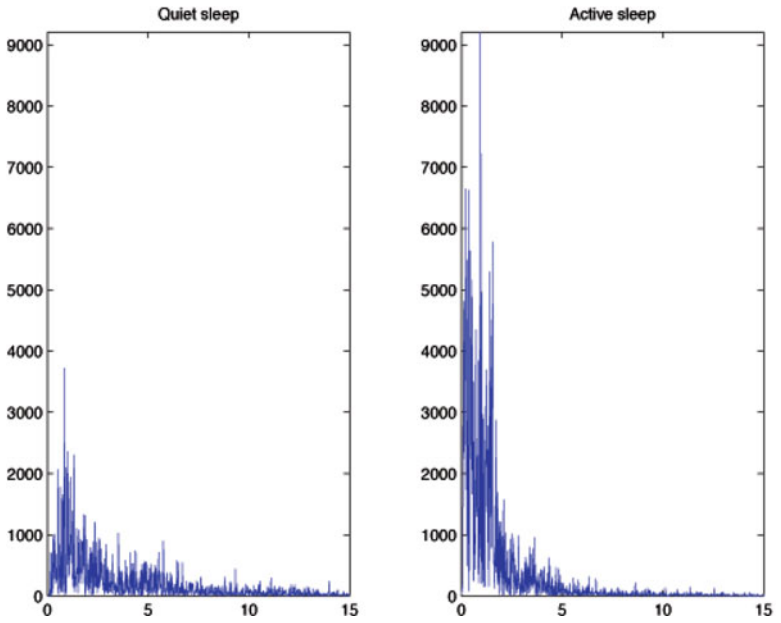


Fig. 6.1 *Left:* Periodogram of the neonatal quiet sleep EEG signal from Fig. 5.7. *Right:* Analogous periodogram for the active sleep stage (From A. Piryatinska's 2004 Case Ph.D. Dissertation.)

Let us return now to two samples of neonatal sleep signals displayed in Fig. 5.7. Their estimated autocovariance functions were shown in Fig. 5.8. Their periodograms have been calculated using formula (6.2.3) and are reproduced in Fig. 6.1. Since the signal was recorded at the sampling frequency of 64 Hz, and the duration of each recording was 3 min, the total number of sample points is $N = 192$. The reader will notice that the periodogram is quite noisy and, perhaps, should be smoothed out to better reflect the true spectrum of the random signal. Nevertheless, a comparison of these rough spectra for quiet sleep and active sleep segment clearly shows that the active sleep signal shows bigger concentration of the spectrum at low frequencies than the quite sleep signal.

Example 6.2.1 (Simple Random Harmonic Oscillation) In this case the random signal is of the form

$$X(t) = A\cos(2\pi f_0(t + \Theta)),$$

where the random amplitude A has zero mean, $\mathbf{E}A = 0$, and finite variance $\mathbf{E}A^2 < \infty$. The random phase Θ is independent of A and uniformly distributed on the interval $[0, P]$ with $P = 1/f_0$. In Chap. 5 we have calculated that the autocovariance function for this signal is

$$\gamma_X(\tau) = \frac{\mathbf{E}|A|^2}{2}\cos(2\pi f_0\tau).$$

Hence the power spectrum of the simple random harmonic oscillation with fundamental frequency f_0 is

$$S_X(f) = \int_{-\infty}^{\infty} \gamma_X(\tau)e^{-2\pi jf\tau}d\tau$$

$$= \int_{-\infty}^{\infty} \frac{\mathbf{E}|A|^2}{2}\frac{e^{j2\pi f_0\tau} + e^{-2\pi jf_0\tau}}{2}e^{-j2\pi f\tau}\,d\tau$$

$$= \frac{\mathbf{E}|A|^2}{4}\Big(\delta(f - f_0) + \delta(f + f_0)\Big),$$

because the inverse Fourier transform of $\delta(f - f_0)$ is

$$\int_{-\infty}^{\infty} \delta(f - f_0)e^{2\pi jf\tau}\,df = e^{2\pi jf_0\tau}.$$

Example 6.2.2 (Superposition of Random Harmonic Oscillations (Random Periodic Signal)) The signal is of the form

$$X(t) = \sum_{k=1}^{N} A_k\cos(2\pi kf_0(t + \Theta_k)),$$

where the zero-mean amplitudes $A_1, \ldots, A_N$, and phases $\Theta_1, \ldots, \Theta_N$, are all independent random quantities and $\Theta_1, \ldots, \Theta_N$, are uniformly distributed on the interval $[0, P]$, $P = 1/f_0$. The autocovariance function of this signal is

$$\gamma_X(\tau) = \sum_{k=1}^{N} \frac{\mathbf{E}|A_k|^2}{2}\cos(2\pi kf_0\tau),$$

and, arguing as in Example 6.2.1, the power spectrum is a linear combination of the Dirac deltas:

$$S_X(f) = \frac{1}{4}\sum_{k=1}^{N} \mathbf{E}|A_k|^2\Big(\delta(f - kf_0) + \delta(f + kf_0)\Big).$$

Thus in this case the power spectrum is concentrated on discrete frequencies $\pm f_0, \pm 2f_0, \ldots, \pm Nf_0$.

Example 6.2.3 (Band-Limited Noise) A stationary signal $X(t)$ is said to be a *band-limited noise* if its spectrum,

$$S_X(f) = \begin{cases} \mathcal{N}_0, & \text{for} - f_{\max} < f < f_{\max}; \\ 0, & \text{elsewhere.} \end{cases}$$

In other words, for a band-limited noise the mean power is distributed uniformly over the frequency band $[-f_{\max}, +f_{\max}]$. The mean power of the band-limited white noise,

$$\mathbf{PW}_X = \int_{-\infty}^{\infty} S_X(f)\, df = \int_{-f_{\max}}^{f_{\max}} \mathcal{N}_0\, df = 2 f_{\max} \mathcal{N}_0,$$

is finite. The autocovariance function of the band-limited white noise can be easily calculated by taking the inverse Fourier transform. Thus we obtain

$$\gamma_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau}\, df = \mathcal{N}_0 \int_{-f_{\max}}^{f_{\max}} e^{j2\pi f\tau}\, df$$

$$= \frac{\mathcal{N}_0}{j2\pi\tau}\left(e^{j2\pi f_{\max}\tau} - e^{-j2\pi f_{\max}\tau}\right) = \frac{\mathcal{N}_0}{\pi\tau}\sin(2\pi f_{\max}\tau).$$

Figure 6.2 shows both the power spectrum of a band-limited white noise and its autocovariance function, for $f_{\max} = 1$ and $\mathcal{N}_0 = 1$. Observe that, not surprisingly, as the bandwidth $2f_{\max}$ expands to infinity the autocovariance function approaches the Dirac delta, the autocovariance function of the ideal white noise which will be discussed in the next example. Note that the maximum value of the autocovariance function $\gamma_X(\tau)$ is attained at $\tau = 0$ and is equal to the mean power $\mathbf{PW}_X = 2f_{\max}$ which diverges to $+\infty$ as the bandwidth increases. However, $\int_{-\infty}^{\infty} \gamma_X(\tau)\, d\tau = S_X(0) = 1$, and the value of the power spectrum at zero frequency is independent of the bandwidth and remains constant.

Example 6.2.4 (The Continuous-Time White Noise Signal) By a standard white noise signal we mean a stationary signal $W(t)$ with a totally flat power spectrum over the whole frequency range,

$$S_W(f) = 1, \qquad -\infty < f < \infty.$$

We can think about it as a limit, for $f_{\max} \to \infty$, of the band limited noise described in Example 6.2.3, but, clearly, the white noise signal is not realizable physically since its mean power is infinite:

$$\mathbf{E}(\mathbf{PW}_W) = \int_{-\infty}^{\infty} 1\, df = \infty.$$

However it is a very useful abstraction. The Fourier transform of its autocovariance function $\gamma_W(\tau)$ must satisfy the equation

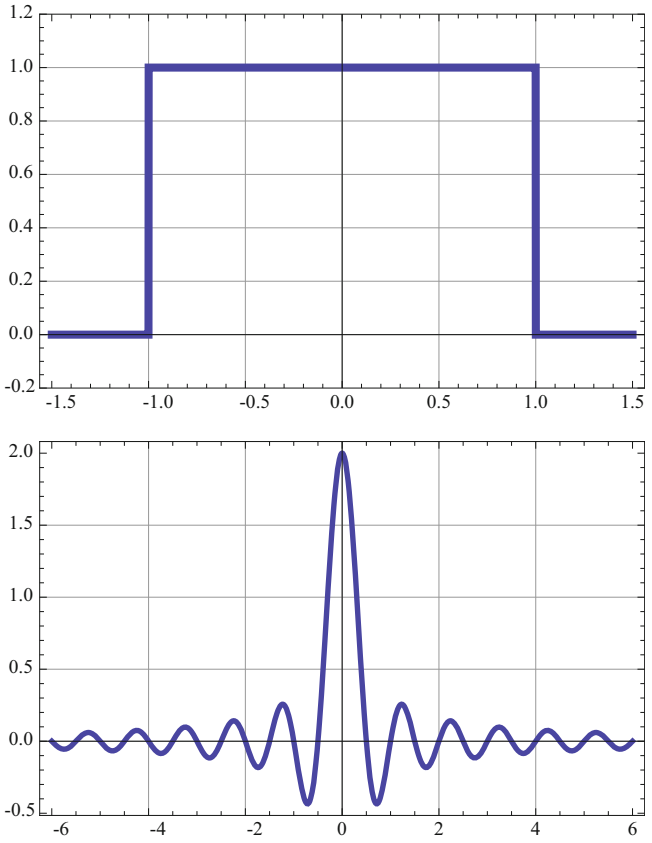


Fig. 6.2 *Top:* Power spectrum of the band-limited white noise $X(t)$ from Example 6.2.3. The bandwidth is $2f_{\max}$ and mean power $\mathbf{PW}_X = 2$. *Bottom:* Autocovariance function of the above band-limited noise. Observe that as the bandwidth expands to infinity the autocovariance function approaches the Dirac delta, the autocovariance function of the ideal white noise

$$\int_{-\infty}^{\infty} \gamma_W(\tau) e^{-j2\pi f\tau}\, d\tau \equiv 1$$

for all $-\infty < f < \infty$, which implies that

$$\gamma_W(\tau) = \delta(\tau).$$

Loosely speaking the above formula can be interpreted as follows: we can say that, for $t \neq s$, the white noise has "values," $X(t)$ and $X(s)$, that are uncorrelated and, for $t = s$, the covariance between $X(t)$ and $X(s)$ is infinite. This autocovariance function is thus not a true function but its shape is not surprising if you compare it to the shape of the autocovariance function for the discrete-time white noise discussed

in Chap. 5. Because of the form of its autocovariance function the white noise is sometimes called a *delta-correlated signal*.

If a random signal $W(t)$ has the spectrum $S_W(f) \equiv \mathcal{N}, > 0$, then we shall call $W(t)$ a white noise of amplitude $\mathcal{N}$.

Example 6.2.5 (Random Switching Signal) The random switching signal $X(t)$ discussed in Chap. 5 has the autocovariance function

$$\gamma_X(\tau) = e^{-2|\tau|}.$$

Thus its power spectral density can be directly calculated by taking the Fourier transform of the autocovariance function:

$$S_X(f) = \int_{-\infty}^{\infty} e^{-2|t|} e^{-j2\pi f t} dt = \int_0^{\infty} e^{-(2+j2\pi f)t} dt + \int_{-\infty}^{0} e^{-(-2+j2\pi f)t} dt$$

$$= \frac{1}{2}\frac{1}{1+j\pi f} + \frac{1}{2}\frac{1}{1-j\pi f} = \frac{1}{1+(\pi f)^2}.$$

Observe that the autocovariance function decays here exponentially as the time lag increases while the power spectrum decays only like the inverse square of the frequency when the latter goes to infinity. The situation is pictured in Fig. 6.3.

At this point is worth recalling Remark 5.1.2, where we made the following observation: If instead of the above standard switching signal X, with standard, mean-one exponential inter-switching times, one considers a more general switching signal X_μ, with μ (on the average) switches per unit time, then as $\mu \to 0$, then its ACvF,

$$\gamma_{X_\mu}(\tau) = \frac{1}{\mu} e^{-2|\tau|/\mu} \to \delta(\tau).$$

Thus, the standard white noise can also be seen as a limit of the general switching signals for which the switching rate and the amplitude of the switches become larger and larger.

6.3 Power Spectra of Interpolated Digital Signals

A random signal sampled at discrete sampling time interval T_s, that is with sampling frequency $f_s = 1/T_s$, produces a sequence of random quantities

$$\ldots, X(-2T_s),\ X(-T_s),\ X(0);\ X(T_s),\ X(2T_s), \ldots \tag{6.3.1}$$

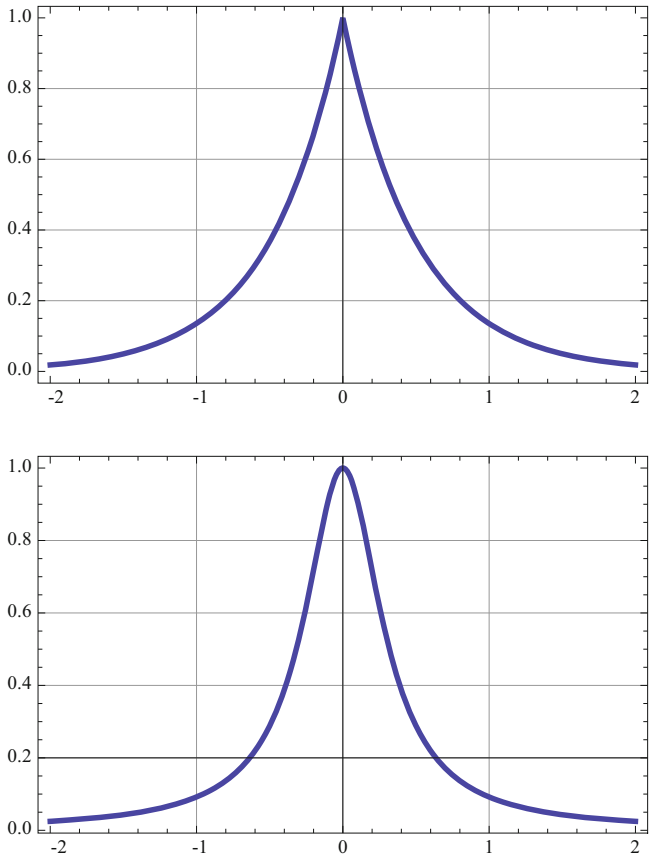


Fig. 6.3 *Top:* Autocovariance function of the random switching signal from Example 6.2.5. *Bottom:* The corresponding power spectrum

To fill in the gaps in the signal produced by discrete sampling at times nT_s we shall interpolate the discrete signal[1] by extending its definition to other times t via the formula

$$X(t) = X(nT_s), \quad \text{for} \quad nT_s \leq t < (n+1)T_s, \tag{6.3.2}$$

and $n = \ldots, -2, -1, 0, 1, 2, \ldots$. Having extended the definition of the signal to continuous time we can obtain its power spectrum following the method developed in Sect. 5.1. In the present case the windowed signal is of the form

[1]The material of this section should be compared with the analysis of the discrete and the fast Fourier transforms carried out in Sect. 2.7 for nonrandom, deterministic signals.

$$X_N(t) = \begin{cases} X(t), & \text{for} - NT_s \le t < NT_s; \\ 0, & \text{elsewhere,} \end{cases}$$

with the window size being $2NT_S$.

Now the mean power

$$\mathbf{E}(\mathbf{PW}_X) = \mathbf{E} \lim_{N\to\infty} \frac{1}{2NT_s} \sum_{n=-N}^{N-1} X^2(nT_s)T_s$$

$$= \mathbf{E} \lim_{N\to\infty} \frac{1}{2NT_s} \int_{-\infty}^{\infty} |X_N(f)|^2 df = \int_{-\infty}^{\infty} \lim_{N\to\infty} \frac{\mathbf{E}|X_N(f)|^2}{2NT_s} df \tag{6.3.3}$$

$$= \int_{-\infty}^{\infty} S(f)\, df,$$

with the power spectral density

$$S(f) = \lim_{N\to\infty} \frac{\mathbf{E}|X_N(f)|^2}{2NT_s}, \tag{6.3.4}$$

and the equality in (6.3.3) resulting from the Parseval formula.

In the next step we evaluate the Fourier transform $X_N(f)$ of the windowed interpolated signal which is needed in formula (6.3.4):

$$X_N(f) = \int_{-\infty}^{\infty} X_N(t) e^{-j2\pi ft}\, dt = \sum_{n=-N}^{N-1} \int_{nT_s}^{(n+1)T_s} X(t) e^{-j2\pi ft}\, dt$$

$$= \frac{1}{-j2\pi f} \sum_{n=-N}^{N-1} X(nT_s) \left(e^{-j2\pi f(n+1)T_s} - e^{-j2\pi fnT_s} \right)$$

$$= \frac{1 - e^{-j2\pi fT_s}}{j2\pi f} \sum_{n=-N}^{N-1} X(nT_s) e^{-j2\pi . fnT_s}.$$

Substituting this result into (6.3.4) we get the following structure of the power spectrum of $X(t)$:

$$S(f) = \lim_{N\to\infty} \frac{|1 - e^{-j2\pi fT_s}|^2}{4\pi^2 f^2} \cdot \frac{\mathbf{E}|\sum_{n=-N}^{N-1} X(nT_s) e^{-2\pi jfnT_s}|^2}{2NT_s}$$

$$= \frac{1-\cos 2\pi f T_s}{2\pi^2 f^2} \lim_{N\to\infty} \sum_{k=-N}^{N-1} \sum_{n=-N}^{N-1} \gamma_X((n-k)T_s) e^{-2\pi j(n-k) f T_s} \frac{1}{2NT_s}.$$

Changing the second summation variable by substitution $n = m + k$ we get

$$S(f) = \frac{1-\cos 2\pi f T_s}{2\pi^2 f^2} \lim_{N\to\infty} \sum_{k=-N}^{N-1} \sum_{m=-N-k}^{N-1-k} \gamma_X(mT_s) e^{-j2\pi m f T_s} \frac{1}{2NT_s}$$

$$= \frac{1-\cos 2\pi f T_s}{2\pi^2 f^2 T_s^2} \cdot \sum_{m=-\infty}^{\infty} \gamma_X(mT_s) e^{-j2\pi m f T_s} T_s.$$

Hence, the power spectrum can be written as a product

$$S(f) = S_1(f) S_2(f), \tag{6.3.5}$$

where the factor

$$S_1(f) = \frac{1-\cos 2\pi f T_s}{2\pi^2 f^2 T_s^2} \tag{6.3.6}$$

decays to 0 at infinite frequencies ($f \to \pm\infty$) and is independent of the statistical properties of the signal (that is, of the autocovariance function $\gamma_X(nT_s)$). The second factor

$$S_2(f) = \sum_{m=-\infty}^{\infty} \gamma_X(mT_s) e^{-j2\pi m f T_s} T_s \tag{6.3.7}$$

is a periodic function with period $f_s = 1/T_s$, represented by the Fourier series with coefficients given by the discrete-time autocovariance function of the discretely sampled signal.

So, if instead of the original power spectrum we consider the ratio $S(f)/S_1(f)$, then we obtain clean relationships paralleling the symmetry of formulas for continuous-time signals:

$$\frac{S(f)}{S_1(f)} = \sum_{m=-\infty}^{\infty} \gamma_X(mT_s) e^{-j2\pi m f T_s} T_s \tag{6.3.8}$$

and

$$\gamma_X(mT_s) = \int_{-f_s/2}^{f_s/2} \frac{S(f)}{S_1(f)} e^{j2\pi m f T_s}\, df. \tag{6.3.9}$$

Remark 6.3.1 It is clear that all the relevant information about the spectrum of the signal sampled with the sampling interval T_s is contained in the frequency interval $(-f_s/2, +f_s/2)$. Power assigned to higher frequencies appearing in the side "lobes" of the spectrum (see Fig. 6.4) is simply an artifact of the interpolation. Should we select a different interpolation scheme, the factor $S_1(f)$ responsible for the decay of the "lobes" would look differently.

Example 6.3.1 (Interpolated Moving Average of the Discrete-Time White Noise) Let the sampling interval $T_s = 1$, and let $W(n)$ be a discrete-time white noise signal ($EW(n) = 0$, $\gamma_W(\tau) = \delta(\tau)/2$). For the moving average signal

$$Y(n) = \frac{1}{2}W(n) + \frac{1}{2}W(n-1),$$

we have calculated in Chap. 5 that

$$\gamma_Y(0) = 1/4, \quad \gamma_Y(\pm 1) = 1/8, \quad \gamma_Y(\tau) = 0, \quad \text{for } |\tau| \geq 2.$$

So, the periodic $S_2(f)$ factor of the power spectrum of the interpolated $Y(n)$ is of the form

$$S_2(f) = \frac{1}{8}e^{j2\pi f\cdot 1} + \frac{1}{4} + \frac{1}{8}e^{-j2\pi f\cdot 1} = \frac{1}{4}(1 + \cos 2\pi f),$$

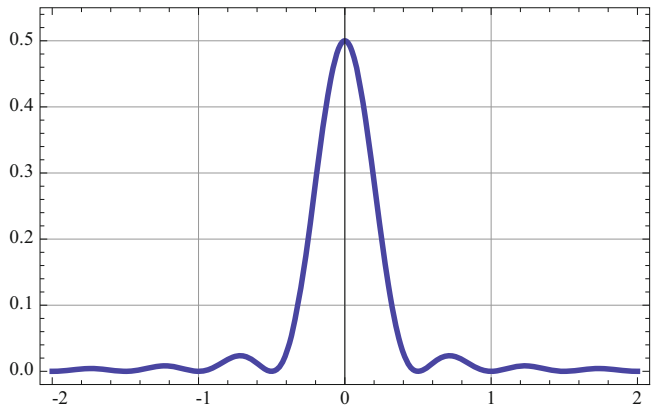


Fig. 6.4 Power spectrum of the interpolated moving average of the discrete-time white noise signal. The sampling rate is $f_s = 1/T_s = 1$, and the relevant spectrum is concentrated in the interval $(-f_s/2, +f_s/2)$. The side "lobes" are an artifact of the interpolation scheme

and the power spectral density itself of the interpolated Y is

$$S_Y(f) = S_1(f)S_2(f) = \frac{1-\cos 2\pi f}{2\pi^2 f^2} \cdot \frac{1}{4}(1+\cos 2\pi f) = \frac{1}{2}\left[\frac{\sin(2\pi f)}{2\pi f}\right]^2.$$

6.4 Problems and Exercises

1 Consider the first-order moving average signal

$$Y(n) = 4W(n) - 6W(n-1) + 3W(n-2),$$

where $W(n)$ is the standard discrete-time white noise signal with $\sigma_W^2 = 1$.

(a) Simulate long samples of this signal using both uniformly distributed (symmetric) and standard Gaussian white noises, and estimate its power spectrum via the periodogram formula (6.2.3). Plot it. Then smooth it out by taking its convolution with a Gaussian kernel. Plot it again.

(b) Calculate and plot the power spectrum density of Y via "interpolation" formula in Example 6.3.1. Compare this plot with the plots obtained in Part (a).

2 With $W(n)$ being the discrete-time white noise signal with $\sigma_W^2 = 5$ (either uniformly distributed and symmetric, or Gaussian) simulate long samples of the signal

$$Y(n) = W(n) + 0.5W(n-1) - 0.3W(n-2).$$

Derive and plot the power spectrum density of Y both via the periodogram formula (6.1.3) and the "interpolated" formula in Example 6.3.1. Follow the plan described in Problem 1.

3 For a given window of size q, find the power spectrum density of a general moving average signal

$$Y(n) = b_0 W(n) + b_1 W(n-1) + \cdots + b_q W(n-q),$$

where $W(n)$ is the discrete-time white noise with $\sigma_W^2 = 1$ (see Example 6.3.1).

4 *Discrete sampling with linear interpolation.* Consider a signal X sampled at sampling interval T_s. Its interpolation to continuous time signal is given by the following formula:

$$X(t) = \sum_{m=-\infty}^{\infty} X(mT_s)\Lambda(t - mT_s),$$

where the interpolating kernel

$$\Lambda(t) = \begin{cases} 1 - t/T_s, & \text{for} 0 < t < T_s; \\ 1 + t/T_s, & \text{for} -T_s < t < 0; \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Plot the kernel $\Lambda(t)$ and the interpolated $X(t)$ for an example of the sampled signal selected by you. Explain the interpolation effect.
(b) Demonstrate that the Fourier transform of the interpolated signal is of the form

$$X_N(f) = \sum_{m=-N}^{N} X(mT_s) e^{-2\pi j m T_s f} \Lambda(f),$$

where $\Lambda(f)$ is the Fourier transform of the kernel $\Lambda(t)$. Produce a plot of $\Lambda(f)$.
(c) Verify that the power spectrum density for the interpolated signal $X(t)$ is

$$S(f) = \lim_{N\to\infty} \frac{E|X_N(f)|^2}{(2N+1)T_s} = \Lambda^2(f) \frac{1}{T_s} \sum_{m=-\infty}^{\infty} \gamma_X(mT_s) e^{-2\pi j m f T_s}.$$

5* A stationary signal $X(t)$ has the autocovariance function

$$\gamma_X(\tau) = 16e^{-5|\tau|} \cos 20\pi\tau + 8 \cos 10\pi\tau.$$

(a) Find the variance of this signal.
(b) Find the power spectrum density of this signal.
(c) Find the value of the spectral density at zero frequency.

6 A stationary signal $X(t)$ has the spectral density of the form

$$S_X(f) = \begin{cases} 5, & \text{for} \frac{10}{2\pi} \le |f| \le \frac{20}{2\pi}; \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the mean power of X.
(b) Find the autocovariance function of X.
(c) Find the value of the autocovariance at $\tau = 0$.

7 A stationary signal $X(t)$ has the spectral density of the form

$$S_X(f) = \frac{9}{(2\pi f)^2 + 64}.$$

At what frequency does the spectral density fall to one half of its maximal value (this value is called the *half-power bandwidth*)?

(a) Write an expression for the spectral density of a bandlimited white noise Y that has the same value at zero frequency and the same mean power as X. What is its bandwidth? It is called the *equivalent noise bandwidth* of X. Compare it with the half-power bandwidth.
(b) Find the autocovariance function of signal X.
(c) Find the autocovariance function of signal Y.
(d) Compare the values of these two autocovariance functions at $\tau = 0$.

8

(a) Consider a solution of the stochastic differential equation described in Example 5.1.6. Take $\alpha = 0.7$, $\beta = 1$, and assume that the white noise $W(n)$ is Gaussian, N(0,1). Produce pictures of five different trajectories of length 100 of this solution truncating the infinite series representing the solution to the first ten terms.
(b) Use the above generated sample signals to estimate their mean and ACvF. Plot the ACvF and compare it graphically with the theoretically derived ACvF. For better comparison, smooth out the empirical ACvF s by taking their convolution with a "nice" kernel, cf. the "moving average" technique applied in Chap. 5 to random signals themselves.
(c) Use the periodogram formula from Sect. 6.2 to estimate the power spectra of the above sample signals. Smooth them out. Compare them graphically with the theoretical power spectrum of the same signal.

9* Verify positive-definiteness (see Remark 6.2.1) of autocovariance functions of stationary signals directly from their definition,

$$\gamma_X(\tau) = \mathbf{E}\Big[(X(t) - \mathbf{E}(X(t)))^* \cdot (X(t+\tau) - \mathbf{E}(X(t)))\Big].$$

Is stationarity condition necessary for positive-definiteness of the covariance function of $X(t)$?

Chapter 7
Transmission of Stationary Signals Through Linear Systems

Abstract Signals produced in nature are almost never experienced in their original form. Usually, we have access to them after they pass through various sensing and/or transmission devices such as voltmeter for electric signals, an ear (Fig. 7.1) for acoustic signals, an eye for visual signals, a fiber optic cable for wide-band internet signals, etc. All of them impose restrictions on the signal being transmitted by attenuating different frequency components of the signal to a different degree. This process is generally called *filtering* and the devices that change signal's spectrum are traditionally called *filters*.

The typical examples here are the so-called *band-pass filters* which permit transmission of the components of the signal only in a certain frequency band, attenuating the frequencies in that band in a uniform fashion, but totally "killing" the frequencies outside this band. Figure 7.1 shows results of filtering a portion of the EEG signal from Fig. 5.1 through four band-pass filters with frequency bands (top to bottom) 0.5–3.5 Hz, 4–7.5 Hz, 8–12.5 Hz, and 13–17 Hz. In neurological literature the contents of the EEG signal within these frequency bands are traditionally called Delta, Theta, Alpha, and Beta waves, respectively.

In this chapter we study how statistical characteristics of random stationary signals are affected by transmission through linear filters. The linearity assumption means that we suppose that there is a linear relationship between the signals on the input and on the output of the filter. In real life it is not always the case, but the study of nonlinear filters is much more difficult than the linear theory presented below, and beyond the scope of this book.

7.1 Time Domain Analysis

In this section we conduct the time-domain analysis of transmission of random signals through a linear system shown schematically below:

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_7

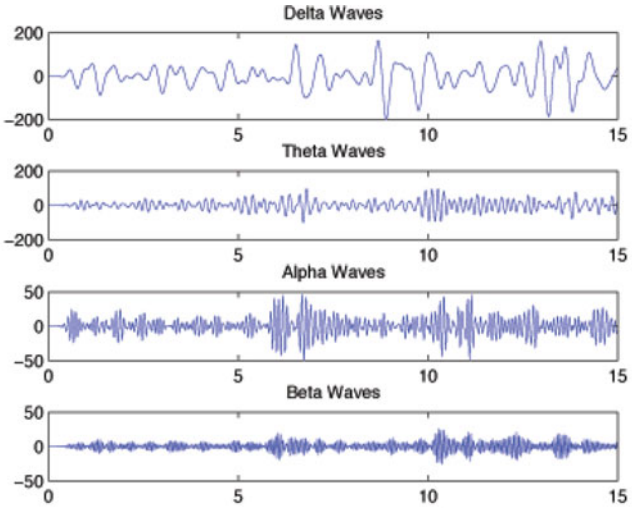


Fig. 7.1 A portion of the EEG signal from Fig. 5.1 filtered through four band-pass filters with frequency bands (top to bottom) 0.5–3.5 Hz, 4–7.5 Hz, 8–12.5 Hz, and 13–17 Hz, respectively

$$X(t) \longrightarrow \boxed{h(t)} \longrightarrow Y(t)$$

The input signal $X(t)$ is assumed to be (real-valued) random and stationary, with mean $\mu_X = \mathbf{E}X(t)$, and autocovariance function

$$\gamma_X(\tau) = \mathbf{E}(X(t) - \mu_X)(X(t+\tau) - \mu_X).$$

The system is identified by a "structure" function $h(t)$, and the output signal $Y(t)$ is defined as the continuous-time moving average (convolution):

$$Y(t) = \int_{-\infty}^{\infty} X(s)h(t-s)\,ds = \int_{-\infty}^{\infty} X(t-s)h(s)\,ds. \tag{7.1.1}$$

Note that in the case of a nonrandom Dirac-delta impulse input $\delta(t)$ the nonrandom output signal is

$$y(t) = \int_{-\infty}^{\infty} \delta(s)h(t-s)\,ds = h(t-0) = h(t).$$

For this reason the system-identifying time-domain "structure" function $h(t)$ is usually called the *impulse response function*.

The mean value of the output signal is easily calculated in terms of the input signal and of the impulse response function:

$$\mathbf{E}Y(t) = \int_{-\infty}^{\infty} \mathbf{E}[X(t-s)]h(s)\,ds = \mu_X \int_{-\infty}^{\infty} h(s)\,ds. \tag{7.1.2}$$

The above formula makes sense only if the last integral is well defined. For this reason we will always assume that the system is *realizable*, that is,

$$\int_{-\infty}^{\infty} |h(s)|\,ds < \infty. \tag{7.1.3}$$

In view of (7.1.2), for realizable systems, if the input signal has zero mean, then the output signal has also zero mean:

$$\mu_X = 0 \qquad \Longrightarrow \qquad \mu_Y = 0.$$

In this situation, from now onwards, we will restrict our attention only to zero mean signals.

The calculation of the autocovariance function of the output signal $Y(t)$ is a little bit more involved. Replacing the product of the integrals by the double integral we obtain

$$\begin{aligned}\gamma_Y(\tau) &= \mathbf{E}(Y(t)Y(t+\tau)) \\ &= \mathbf{E}\Big[\int_{-\infty}^{\infty} X(t-s)h(s)\,ds \int_{-\infty}^{\infty} X(t+\tau-u)h(u)\,du\Big] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{E}\Big[X(t-s)X(t+\tau-u)\Big]h(s)h(u)\,ds\,du.\end{aligned}$$

Then, in view of the stationarity assumption,

$$\mathbf{E}\Big[X(t-s)X(t+\tau-u)\Big] = \mathbf{E}\Big[X(-s)X(\tau-u)\Big] = \gamma_X(\tau-u+s),$$

so that, finally,

$$\gamma_Y(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \gamma_X(\tau-u+s)h(s)h(u)\,ds\,du. \tag{7.1.4}$$

A system is said to be *causal* if the current values of the output depend only on the past and present values of the input. This property can be equivalently stated as the requirement that the impulse response function,

$$h(t) = 0, \qquad \text{for} \qquad t \le 0. \tag{7.1.5}$$

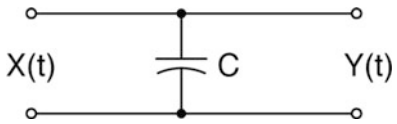


Fig. 7.2 A standard integrating circuit. The voltage $Y(t)$ on the output is the integral of the current $X(t)$ on the input

In other words, the moving average is performed only over the past. This condition, in particular, implies that the second output integral in (7.1.1) is restricted to the positive half-line,

$$Y(t) = \int_0^\infty X(t-s)h(s)\,ds, \tag{7.1.6}$$

and formula (6.1.4) for the autocovariance function takes the form

$$\gamma_Y(\tau) = \int_0^\infty \int_0^\infty \gamma_X(\tau - u + s)h(s)h(u)\,ds\,du. \tag{7.1.7}$$

In what follows in this chapter we will consider only causal filters.

Example 7.1.1 (An Integrating Circuit) A standard integrating circuit with a single capacitor is shown in Fig. 7.2.

The impulse response function for this system is the Unit Step Function $u(t)$ multiplied by $1/C$, where the constant C represents capacitance of the capacitor:

$$h(s) = \frac{1}{C}u(s) = \begin{cases} 0, & \text{for } s < 0; \\ 1/C, & \text{for } s \geq 0. \end{cases}$$

The output

$$Y(t) = \frac{1}{C}\int_{-\infty}^{\infty} X(s)u(t-s)ds = \frac{1}{C}\int_{-\infty}^{t} X(s)\,ds.$$

Obviously, this system, although causal, is not realizable over the whole time line since

$$\int_{-\infty}^{\infty} |h(t)|\,dt = \int_0^\infty \frac{1}{C}\,dt = \infty.$$

To avoid this difficulty we need to restrict the integrating circuit to a finite time interval and assume that the adjusted impulse response function is of the form

$$h(s) = \begin{cases} 0, & \text{for } s < 0; \\ 1/C, & \text{for } 0 \leq s \leq T; \\ 0, & \text{for } s > T. \end{cases} \tag{7.1.8}$$

In this situation the system is realizable and the output is

$$Y(t) = \int_{-\infty}^{\infty} X(s)h(t-s)ds = \frac{1}{C}\int_{t-T}^{t} X(s)ds.$$

The autocovariance function is equal to

$$\gamma_Y(\tau) = \int_0^T \int_0^T \gamma_X(\tau - u + s)\, h(s)\, h(u)\, ds\, du$$

$$= \frac{1}{C^2}\int_0^T \int_0^T \gamma_X(u - (\tau + s))\, ds\, du, \tag{7.1.9}$$

because, for real-valued signals, the autocovariance function is even, $\gamma_X(-\tau) = \gamma_X(\tau)$.

Therefore, if the input signal is the standard white noise $X(t) = W(t)$ with the autocovariance $\gamma_W(t) = \delta(t)$, and $C = 1$, then for $\tau \geq 0$, the output autocovariance function

$$\gamma_Y(\tau) = \int_0^T \int_0^T \delta(u - (\tau + s))\, du\, ds = \int_0^T \zeta(s)\, ds,$$

where

$$\zeta(s) = \begin{cases} 0, & \text{for } \tau + s < 0; \\ 1/2, & \text{for } \tau + s = 0; \\ 1, & \text{for } 0 < \tau + s < T; \\ 1/2, & \text{for } \tau + s = T; \\ 0, & \text{for } \tau + s > T. \end{cases}$$

Hence, (Figs. 7.3 and 7.4)

$$\gamma_Y(\tau) = \begin{cases} 0, & \text{for } \tau < -T; \\ T - |\tau|, & \text{for } -T \leq \tau \leq T; \\ 0, & \text{for } \tau > T. \end{cases} \tag{7.1.10}$$

If the input signal $X(t)$ is a simple random harmonic oscillation with the autocovariance function $\gamma_X(\tau) = \cos\tau$, and, again, $C = 1$, then the output autocovariance is

$$\gamma_Y(\tau) = \int_0^T \int_0^T \cos(\tau - u + s)\, ds\, du = 2\cos\tau(1 - \cos T). \tag{7.1.11}$$

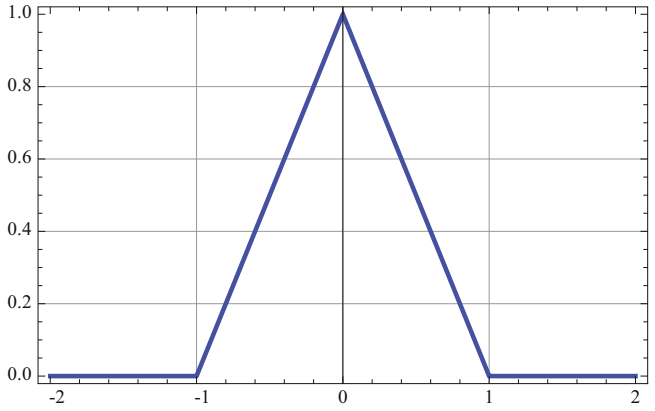


Fig. 7.3 The output autocovariance function $\gamma_Y(\tau)$ (7.1.10) of the integrating system (7.1.8) with $T = 1$, in the case of the standard white noise input $X(t) = W(t)$

As simple as the formula (7.1.9) for the output autocovariance function seems to be, the analytic evaluation of the double convolution may get tedious very quickly. Consider, for example, an input signal $X(t)$ with the autocovariance function

$$\gamma_X(\tau) = \frac{1}{1+\tau^2}, \tag{7.1.12}$$

which corresponds to the exponentially decaying power spectrum (see Sect. 6.4).

In this case,

$$\begin{aligned}\gamma_Y(\tau) &= \int_0^T \int_0^T \frac{1}{1+(\tau-u+s)^2}\, ds\, du \\ &= \frac{1}{2}\Big(2(T-\tau)\arctan(T-\tau) - 2\tau \arctan\tau \\ &\quad - \log(1+(T-\tau)^2) + \log(1+\tau^2)\Big) \\ &\quad + \frac{1}{2}\Big(-2\tau\arctan(\tau) + 2(\tau+T)\arctan(\tau+T) \\ &\quad + \log(1+\tau^2) + \log(1+T^2+2T\tau+\tau^2)\Big).\end{aligned} \tag{7.1.13}$$

So, even for a relatively simple autocovariance function of the input, the output autocovariance may be quite complex. And yes, you guessed right, to avoid the tedium of the paper-and-pencil calculations we have obtained the above formula using *Mathematica*. Figure 7.5 traces graphically the dependence of $\gamma_Y(\tau)$ on T.

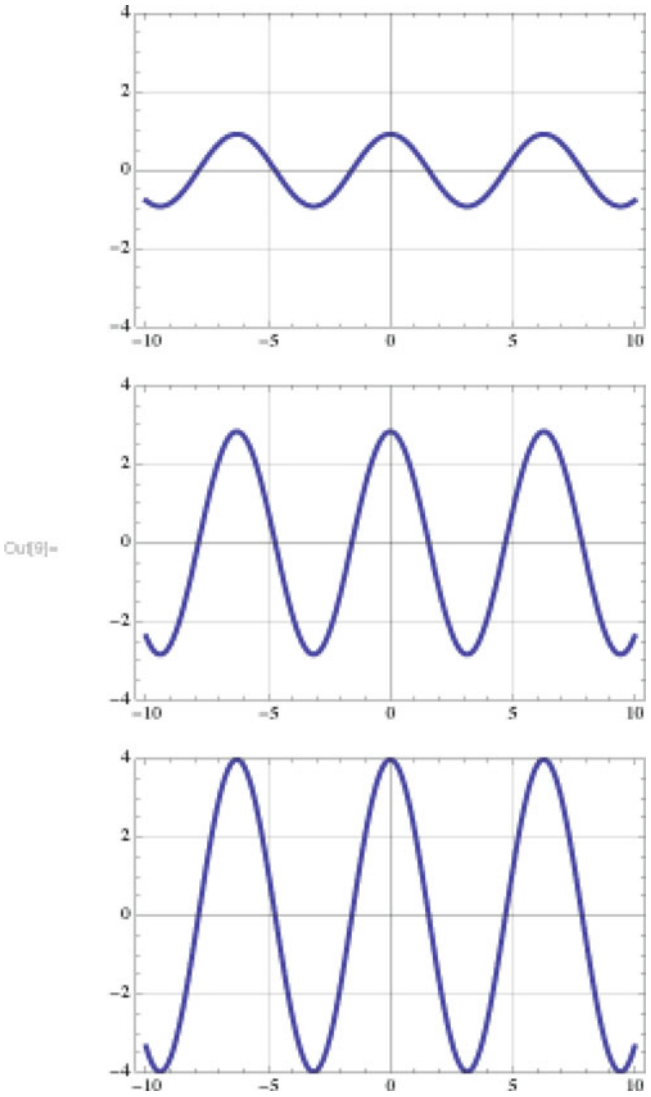


Fig. 7.4 The output autocovariance functions $\gamma_Y(\tau)$ (7.1.11) of the integrating system (7.1.8) with $T = 1, 2$, and 3 (top to bottom), in the case simple random harmonic oscillation input with $\gamma_X(\tau) = \cos\tau$. Note the increasing amplitude of $\gamma_Y(\tau)$ as T increases

Example 7.1.2 (An RC-Filter) A standard RC-filter is shown in Fig. 7.6.

The impulse response function of this circuit is of the form

$$h(t) = \frac{1}{RC} \exp\left(-\frac{t}{RC}\right) \cdot u(t), \tag{7.1.14}$$

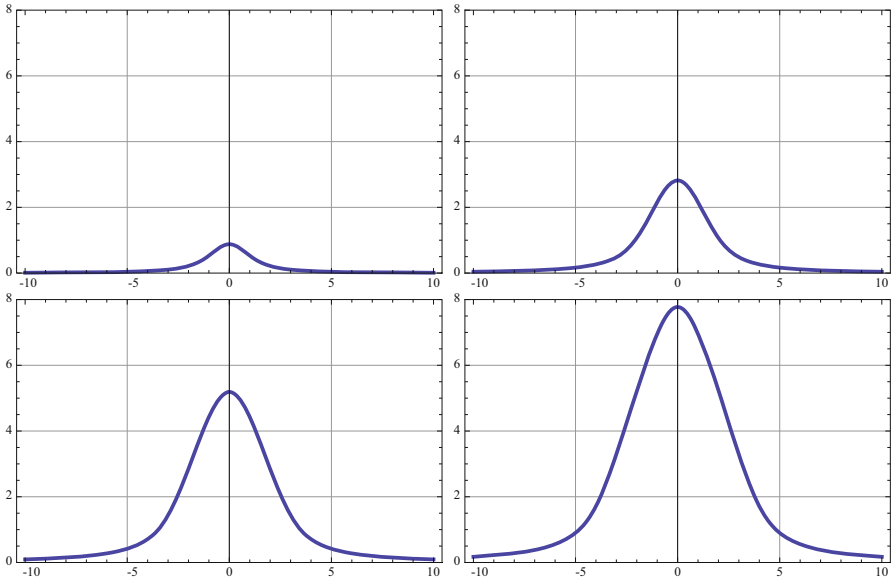


Fig. 7.5 The output autocovariance functions $\gamma_Y(\tau)$ (7.1.13) of the integrating system (7.1.8) with $T = 1, 2, 3$, and 4, in the case of input with $\gamma_X(\tau) = 1/(1+\tau^2)$. Note the growing maximum and spread of $\gamma_Y(\tau)$ as T increases

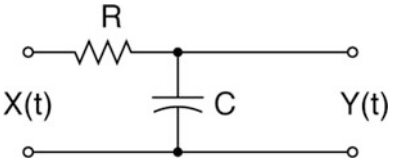


Fig. 7.6 A standard RC-filter with the impulse response function $h(t) = (1/RC)\exp(-t/RC)\cdot u(t)$

where $u(t)$ is the usual Unit Step Function, R is the electrical resistance, and C is the capacitance. The product RC represents the so-called time constant of the circuit.

In the case of the white noise input signal with $\gamma_X(\tau) = \delta(\tau)$, the output autocovariance function, for $\tau > 0$, is

$$\gamma_Y(\tau) = \int_0^\infty \int_0^\infty \delta(u-(s+\tau))h(u)h(s)\,du\,ds = \int_0^\infty h(s+\tau)h(s)\,ds$$

$$= \int_0^\infty \frac{1}{RC}e^{-\frac{s+\tau}{RC}}\cdot\frac{1}{RC}e^{-\frac{s}{RC}}\,ds = \frac{1}{2RC}e^{-\frac{\tau}{RC}}.$$

So

$$\gamma_Y(\tau) = \frac{1}{2RC}\exp\left(-\frac{|\tau|}{RC}\right). \tag{7.1.15}$$

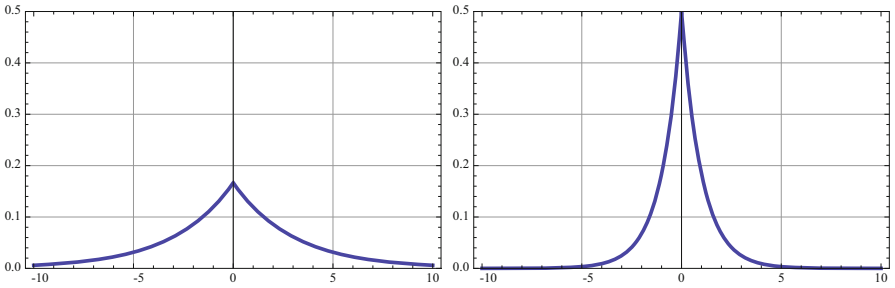


Fig. 7.7 The output autocovariance function $\gamma_Y(\tau)$ for the RC-filter (7.1.14) with a standard white noise input with $\gamma_X(\tau) = \delta(\tau)$. The figure on the left shows the case of a small time constant, $RC = 1$, and the one on the right, the case of a larger time constant, $RC = 3$. Note the difference in the maximum and the spread of $\gamma_Y(\tau)$ in these two cases

The shape of the output autocovariance function for small and large values of the RC constant is shown in Fig. 7.7.

Remark 7.1.1 (Ornstein-Uhlenbeck Stationary Signals) You may have noticed that the ACvF appearing in (7.1.15) has the same exponential shape as that of the switching signal considered in Sect. 5.1, and also, in discrete time, that of the solution of a stochastic difference equation considered in the same section. However, if the input white noise in the above example has a Gaussian distribution, then the output is also Gaussian (obviously, not a switching signal which takes only two values). A Gaussian stationary signal with the exponential ACvF (7.1.15) is traditionally called the Ornstein-Uhlenbeck signal (process) and it appears as a model in numerous physical and engineering problems, see Chap. 9 for a detailed discussion of Gaussian stationary signals.

For the simple random harmonic oscillation with autocovariance $\gamma_X(\tau) = \cos\tau$ as the input, the output autocovariance is

$$\gamma_Y(\tau) = \int_0^\infty \int_0^\infty \cos(\tau - u + s) \frac{1}{RC} \exp\left(\frac{-s}{RC}\right) \frac{1}{RC} \exp\left(\frac{-u}{RC}\right) ds\, du$$

$$= \frac{\cos\tau}{1 + (RC)^2}.$$

But a slightly more complex input autocovariance function,

$$\gamma_X(\tau) = e^{-2|\tau|},$$

corresponding to the switching input signal produces the output autocovariance function of the form,

$$\gamma_Y(\tau) = \frac{1}{(RC)^2}\int_0^\infty\int_0^\infty e^{-|\tau-u+s|}e^{-(s+u)/(RC)}\,ds\,du \tag{7.1.16}$$

$$= \frac{1}{(RC)^2}\Bigg[\int_0^\tau\int_0^\infty e^{-(\tau-u+s)}e^{-(s+u)/(RC)}\,ds\,du$$

$$+\int_\tau^\infty\left(\int_0^{u-\tau} e^{\tau-u+s}e^{-(s+u)/(RC)}\,ds + \int_{u-\tau}^\infty e^{-(\tau-u+s)}e^{-(s+u)/(RC)}\,ds\right)du\Bigg],$$

which, although doable (see Problems and Exercises), is not fun to evaluate.

7.2 Frequency Domain Analysis and System's Bandwidth

Examples provided in the preceding section demonstrated analytic difficulties related to the time domain analysis of random stationary signals transmitted through linear systems. In many cases analysis becomes much simpler if it is carried out in the frequency domain. For this purpose let us consider the Fourier transform $H(f)$ of the system's impulse response function $h(t)$:

$$H(f) = \int_{-\infty}^\infty h(t)e^{-2\pi jft}\,dt, \tag{7.2.1}$$

which traditionally is called the system's *transfer function*.

Now the task is to calculate the power spectrum,

$$S_Y(f) = \int_{-\infty}^\infty \gamma_Y(\tau)e^{-2\pi jf\tau}\,d\tau, \tag{7.2.2}$$

of the output signal given the power spectrum,

$$S_X(f) = \int_{-\infty}^\infty \gamma_X(\tau)e^{-2\pi jf\tau}\,d\tau,$$

of the input signal. Since the output autocovariance function $\gamma_Y(t)$ has been calculated in Sect. 6.1, substituting the expression obtained in (7.1.4) into (7.2.1) we get

$$S_Y(f) = \int_{-\infty}^\infty\left(\int_{-\infty}^\infty\int_{-\infty}^\infty \gamma_X(\tau-s+u)h(s)h(u)\,ds\,du\right)e^{-2\pi jf\tau}\,d\tau$$

$$= \int_{-\infty}^\infty\int_{-\infty}^\infty\left(\int_{-\infty}^\infty \gamma_X(\tau-s+u)e^{-2\pi jf(\tau-s+u)}\,d\tau\right)h(s)e^{-2\pi jfs}ds\cdot h(u)e^{2\pi jfu}du.$$

Making the substitution $\tau - s + u = w$ in the inner integral we arrive at the final formula

$$S_Y(f) = S_X(f) \cdot H(f) \cdot H^*(f) = S_X(f) \cdot |H(f)|^2. \tag{7.2.3}$$

So the output power spectrum is obtained simply by multiplying the input power spectrum by a fixed factor $|H(f)|^2$ which is called the system's *power transfer function*.

The appearance of the power transfer function, $|H(f)|^2$, in formula (7.2.3) suggests introduction of the concept of the system's bandwidth. As in the case of signals (see Sect. 6.4) several choices are possible.

The *equivalent-noise bandwidth* BW_n is defined as the cutoff frequency f_{max} of the limited-band white noise with the amplitude equal to the value of the system's power transfer function at 0 and the mean power equal to the integral of the system's power transfer function, that is,

$$2BW_n|H(0)|^2 = \int_{-\infty}^{\infty} |H(f)|^2 \, df,$$

which gives

$$BW_n = \frac{1}{2|H(0)|^2} \int_{-\infty}^{\infty} |H(f)|^2 \, df. \tag{7.2.4}$$

The *half-power bandwidth* $BW_{1/2}$ is defined as the frequency where the system's power transfer function declines to one half of its maximum value which is always equal to $|H(0)|^2$. Thus it is obtained by solving, for an unknown $BW_{1/2}$, the equation,

$$|H(BW_{1/2})|^2 = \frac{1}{2}|H(0)|^2. \tag{7.2.5}$$

Obviously, the above bandwidth concepts make best sense for low-pass filters, that is in the case when the system's power transfer function has a distinctive maximum at 0, dominating its values elsewhere. But for other systems, such as bandpass filters, similar bandwidth definitions can be easily devised.

Example 7.2.1 (An RC-Filter) Recall that in this case the impulse response function is given by

$$h(t) = \frac{1}{RC} e^{-\frac{t}{RC}} \cdot u(t).$$

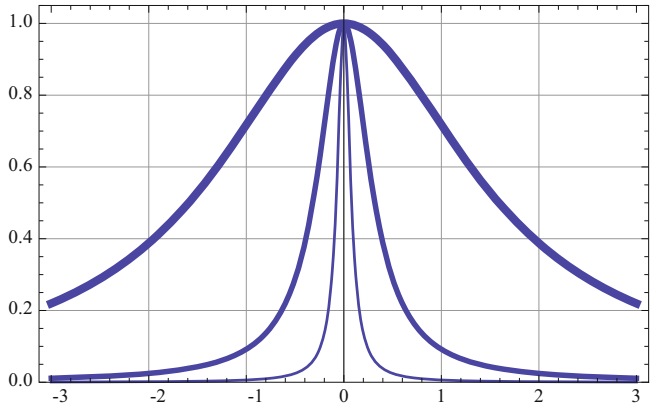


Fig. 7.8 Power transfer functions $|H(f)|^2 = 1/(1 + (2\pi RCf)^2)$ for the RC-filter with the RC constants 0.1 (thick line), 0.5 (medium line), and 2.0 (thin line). The half-power bandwidths $BW_{1/2}$ are, respectively, 1.6, 0.32, and 0.08

So, the transfer function is

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-2\pi jft}\,dt = \int_{0}^{\infty} \frac{1}{RC} e^{-\frac{t}{RC}} e^{-2\pi jft}\,dt = \frac{1}{1 + 2\pi jRCf}$$

and, consequently, the power transfer function is (Fig. 7.8)

$$|H(f)|^2 = \frac{1}{1 + 2\pi jRCf} \cdot \frac{1}{1 - 2\pi jRCf} = \frac{1}{1 + (2\pi RCf)^2}. \tag{7.2.6}$$

The half-power bandwidth of the RC-filter is easily computable from the equation,

$$\frac{1}{1 + (2\pi RC(BW_{1/2}))^2} = \frac{1}{2},$$

which gives

$$BW_{1/2} = \frac{1}{2\pi RC}.$$

The bandwidth decreases hyperbolically with the increase of the RC constant.

The output power spectra for an RC-filter are thus easily evaluated. In the case of the standard white noise input with $S_X(f) \equiv 1$, the output power spectrum is

$$S_Y(f) = \frac{1}{1 + (2\pi RCf)^2}.$$

If the input signal is a random oscillation with the power spectrum,

$$S_X(f) = \frac{A_0^2}{2}\Big(\delta(f - f_0) + \delta(f + f_0)\Big),$$

then the output power spectrum is

$$S_Y(f) = \frac{A_0^2}{2}\Big(\delta(f - f_0) + \delta(f + f_0)\Big) \cdot \frac{1}{1 + (2\pi RCf)^2}.$$

If the input is a switching signal with the power spectrum,

$$S_X(f) = \frac{1}{1 + (af)^2},$$

then the output power spectrum is

$$S_Y(f) = \frac{1}{1 + (af)^2} \cdot \frac{1}{1 + (2\pi RCf)^2}.$$

Example 7.2.2 (Bandwidth of the Finite-Time Integrating Circuit) Let us calculate the bandwidths BW_n and $BW_{1/2}$ for the finite time integrator with the impulse response function

$$h(t) = \begin{cases} 1, & \text{for } 0 \le t \le T; \\ 0, & \text{elsewhere.} \end{cases}$$

In this case the transfer function

$$H(f) = \int_0^T e^{-2\pi jft}\,dt = \frac{1}{2\pi jf}\Big(1 - e^{-2\pi jfT}\Big),$$

so that the power transfer function

$$|H(f)|^2 = \frac{(1 - e^{-2\pi jfT})(1 - e^{2\pi jfT})}{(2\pi f)^2} = \frac{2(1 - \cos 2\pi fT)}{(2\pi f)^2}. \tag{7.2.7}$$

Finding directly the integral of the power transfer function is a little tedious but, fortunately, by Parseval's formula,

$$\int_{-\infty}^{\infty} |H(f)|^2\,df = \int_{-\infty}^{\infty} h^2(t)\,dt = \int_0^T dt = T,$$

and

$$H(0) = \int_0^T h(t)\,dt = T.$$

Thus the equivalent-noise bandwidth (7.2.4) is

$$BW_n = \frac{1}{2T^2} \cdot T = \frac{1}{2T}.$$

Finding the half-power bandwidth requires solving Eq. (7.2.5)

$$\frac{2(1 - \cos 2\pi (BW_{1/2})T)}{(2\pi (BW_{1/2}))^2} = \frac{T^2}{2},$$

which can be done only numerically. Indeed, a quick graphical analysis (see Fig. 7.9), for $T = 1$, gives the half-power bandwidth $BW_{1/2} = 0.443$, slightly less than the corresponding equivalent-noise bandwidth $BW_{\text{eqn}} = 0.500$.

7.3 Digital Signal, Discrete Time Sampling

In this section we will take a look at transmission of random stationary signals through linear systems when the signals are sampled at discrete times with the sampling interval T_s. The system can be schematically represented as follows:

$$X(nT_s) \longrightarrow \boxed{h(nT_s)} \longrightarrow Y(nT_s).$$

The input signal now forms a stationary random sequence,

$$X(nT_s), \quad n = \ldots -1, 0, 1, \ldots, \tag{7.3.1}$$

and the output signal,

$$Y(nT_s), \quad n = \ldots -1, 0, 1, \ldots, \tag{7.3.2}$$

is produced by discrete-time convolution of the input signal $X(nT_s)$ with the discrete time impulse response sequence $h(nTs)$:

$$Y(nT_s) = \sum_{i=-\infty}^{\infty} X(iT_s)h(nT_s - iT_s)T_s. \tag{7.3.3}$$

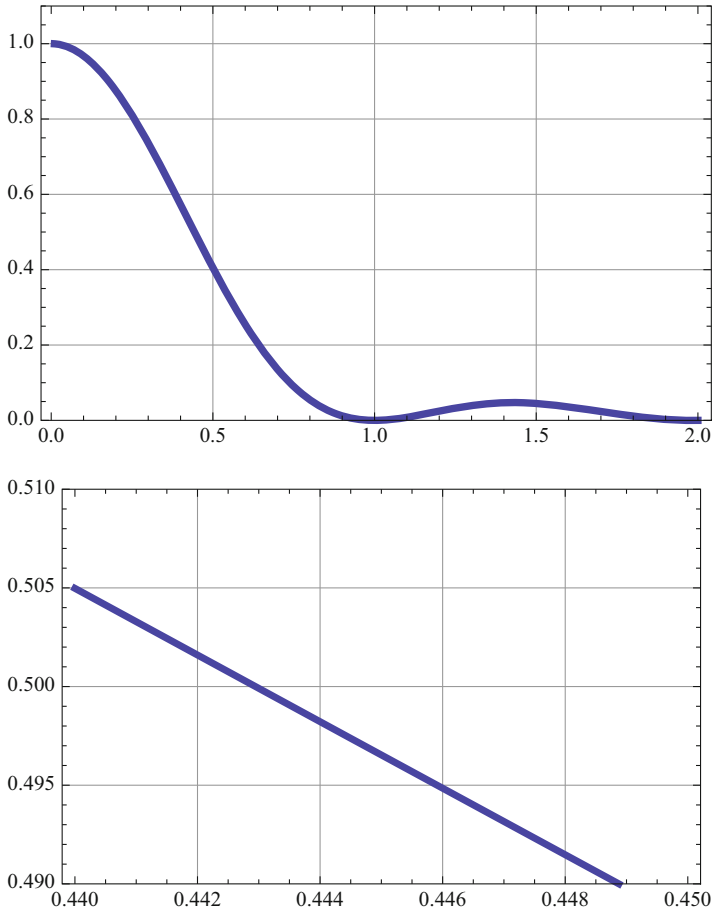


Fig. 7.9 Top: Power transfer function (7.2.7) of the finite time integrating circuit with $T = 1$. Bottom: Magnified portion of the power transfer function for f between 0.44 and 0.45. This graphical analysis gives the half-power bandwidth $BW_{1/2} = 0.443$

In the discrete-time case the realizability condition is

$$\sum_{n=-\infty}^{\infty} |h(nT_s)| < \infty,$$

and the causality condition means that

$$h(nT_s) = 0, \text{ for } n < 0.$$

With discrete-time inputs and outputs the autocovariance functions are just discrete sequences and are defined by the formulas,

$$\gamma_X(kT_s) = \mathbf{E}(X(nT_s)X(nT_s + kT_s)), \quad \gamma_Y(kT_s) = E(Y(nT_s)Y(nT_s + kT_s)).$$

Then a direct application of (7.3.3) yields the following formula for the output autocovariance sequence as a function of the input autocovariance sequence and the impulse response sequence:

$$\gamma_Y(kT_s) = \sum_{l=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \gamma_X(kT_s - lT_s + iT_s)h(lT_s)h(iT_s)T_s^2. \tag{7.3.4}$$

To move into the frequency domain one can either directly apply the discrete or fast Fourier transforms or, as in Sect. 6.3, use the straight continuous-time Fourier transform technique assuming that both the signal and the impulse response function have been interpolated by constants between sampling points. We will follow the latter approach. So, using the formula (6.3.5), we get

$$S_X(f) = S_1(f) \cdot S_{2,X}(f), \tag{7.3.5}$$

with

$$S_{2,X}(f) = \sum_{m=-\infty}^{\infty} \gamma_X(mT_s)e^{-j2\pi mfT_s}T_s,$$

and

$$S_Y(f) = S_1(f) \cdot S_{2,Y}(f), \tag{7.3.6}$$

with

$$S_{2,Y}(f) = \sum_{m=-\infty}^{\infty} \gamma_Y(mT_s)e^{-j2\pi mfT_s}T_s,$$

and

$$S_1(f) = \frac{1 - \cos 2\pi fT_s}{2\pi^2 f^2 T_s^2}.$$

Remember that all the relevant information about the discrete sampled signal is contained in the frequency interval $(-f_s/2, f_s/2)$ (see Remark 6.3.1). The transfer function of this system is

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}\,dt = \sum_{k=-\infty}^{\infty} h(kT_s) \int_{kT_s}^{(k+1)T_s} e^{-j2\pi ft}\,dt$$

$$= \frac{1 - e^{j2\pi fT_s}}{-j2\pi fT_s} \sum_{k=-\infty}^{\infty} h(kT_s)e^{-j2\pi fkT_s}T_s, \tag{7.3.7}$$

so that the power transfer function,

$$|H(f)|^2 = \frac{1 - \cos 2\pi fT_s}{2\pi^2 f^2 T_s^2} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(kT_s)h(nT_s)e^{-j2\pi f(k-n)T_s}T_s^2. \tag{7.3.8}$$

Again, all the relevant information about the discrete power transfer function is contained in the frequency interval $(-f_s/2,\, f_s/2)$ (see Remark 6.3.1).

Finally, since we already know from Sect. 7.2 that

$$S_Y(f) = |H(f)|^2 S_X(f),$$

we also get from ((7.3.5)–(7.3.6)) that

$$S_{2,Y}(f) = |H(f)|^2 S_{2,X}(f). \tag{7.3.9}$$

or, equivalently,

$$\sum_{m=-\infty}^{\infty} \gamma_Y(mT_s)e^{-j2\pi mfT_s}T_s = |H(f)|^2 \cdot \sum_{m=-\infty}^{\infty} \gamma_X(mT_s)e^{-j2\pi mfT_s}T_s. \tag{7.3.10}$$

Example 7.3.1 (Autoregressive Moving Average Signal (ARMA)) We now take the sampling period $T_s = 1$ and the output $Y(n)$ determined from the input $X(n)$ via the autoregressive moving average scheme with parameters p and q (in brief, $ARMA(p, q)$):

$$Y(n) = \sum_{l=0}^{q} b(l)X(n-l) - \sum_{l=1}^{p} a(l)Y(n-l). \tag{7.3.11}$$

Defining $a(0) = 1$, we can then write

$$\sum_{l=0}^{p} a(l)Y(n-l) = \sum_{l=0}^{q} b(l)X(n-l).$$

Since the Fourier transform of the convolution is a product of Fourier transforms, we have

$$X(f)\sum_{l=0}^{q} b(l)e^{-2\pi jflT} = Y(f)\sum_{l=0}^{p} a(l)e^{-2\pi jflT},$$

so the transfer function

$$H(f) = \frac{Y(f)}{X(f)} = \frac{\sum_{l=0}^{q} b(l)e^{-2\pi jflT}}{\sum_{l=0}^{p} a(l)e^{-2\pi jflT}}. \tag{7.3.12}$$

Example 7.3.2 (A Solution of Stochastic Difference Equation) This example was considered in Chap. 5 but let us observe that it is a special case of Example 7.3.1, with parameters $p = 1$, $q = 0$, and the input signal being the standard discrete white noise $W(n)$ with $\sigma_W^2 = 1$. In other words,

$$Y(n) = -a_1 Y(n-1) + b_0 W(n).$$

In view of (7.3.12), the power transfer function

$$|H(f)|^2 = \frac{b_0}{1 + a_1 e^{-2\pi jf}} \cdot \frac{b_0}{1 + a_1 s e^{2\pi jf}} = \frac{b_0^2}{1 + a_1^2 + 2a_1 \cos 2\pi f},$$

with, again, all the relevant information contained in the frequency interval $-1/2 < f < 1/2$.

Given that the input is the standard white noise, we have that

$$S_Y(f) = |H(f)|^2 \cdot 1 = \frac{b_0^2}{1 + a_1^2 + 2a_1 \cos 2\pi f}. \tag{7.3.13}$$

One way to find the output autocovariance sequence $\gamma_Y(n)$ would be to take into account the relationship (7.3.10) and expand (7.3.13) into the Fourier series; its coefficients will form the desired autocovariance sequence. This procedure is straightforward and requires only an application of the formula for the sum of a geometric series (see Sect. 6.4).

However, we would like to explore here a different route and employ a recursive procedure to find the output autocovariance sequence. First, observe that

$$\gamma_Y(k) = \mathbf{E}(Y(n)Y(n+k))$$

$$= \mathbf{E}(-a_1 Y(n-1) + b_0 X(n)) \cdot (-a_1 Y(n+k-1) + b_0 X(n+k))$$

$$= a_1^2 \mathbf{E}(Y(n-1)Y(n+k-1)) - a_1 b_0 \mathbf{E}(Y(n-1)X(n+k))$$

$$-a_1 b_0 \mathbf{E}(X(n)Y(n+k-1)) + b_0^2 \mathbf{E}(X(n)X(n+k))$$

$$= a_1^2 \gamma_Y(k) - a_1 b_0 \gamma_{XY}(k-1) + b_0^2 \gamma_X(k),$$

where

$$\gamma_{XY}(k) = \mathbf{E}(X(n)Y(n+k))$$

is the crosscovariance sequence of signals $X(n)$ and $Y(n)$. So,

$$\gamma_Y(k) = \frac{b_0}{1-a_1^2}\Big(- a_1 \gamma_{XY}(k-1) + b_0 \gamma_X(k)\Big).$$

For $k = 0$,

$$\gamma_Y(0) = \sigma_Y^2 = \frac{b_0}{1-a_1^2}\Big(- a_1 \mathbf{E}(X(n)Y(n-1)) + b_0 \gamma_X(0)\Big)$$

$$= \frac{b_0^2}{1-a_1^2}\gamma_X(0) = \frac{b_0^2}{1-a_1^2}.$$

For $k = 1$,

$$\gamma_Y(1) = \frac{b_0}{1-a_1^2}\Big(- a_1 \gamma_{XY}(0) + b_0 \gamma_X(1)\Big) = \frac{b_0(-a_1)}{1-a_1^2}\mathbf{E}(X(0)Y(0))$$

$$= \frac{b_0(-a_1)}{1-a_1^2}\mathbf{E}\Big(X(0)\Big(a_1 Y(-1) + b_0 X(0)\Big)\Big) = \frac{b_0^2(-a_1)}{1-a_1^2}.$$

For a general $k > 1$,

$$\gamma_Y(k) = \frac{b_0}{1-a_1^2}\Big(- a_1 \gamma_{XY}(k-1) + b_0 \gamma_X(k)\Big),$$

and, as above,

$$\gamma_{XY}(k-1) = \mathbf{E}\Big(X(0)Y(k-1)\Big)$$

$$= \mathbf{E}\Big(X(0)(-a_1 Y(k-2) + b_0 X(k-1))\Big)$$

$$= (-a_1)\mathbf{E}\Big(X(0)Y(k-2)\Big)$$

$$= (-a_1)\gamma_{XY}(k-2) = \cdots = (-a_1)^{k-1}\gamma_{XY}(0) = b(0)(-a_1)^{k-1}.$$

Since the autocovariance sequence must be an even function of variable k, we finally get, for any $k = \ldots, -2, -1, 0, 1, 2, \ldots,$

$$\gamma_Y(k) = \frac{b_0^2}{1 - a_1^2}(-a_1)^{|k|},$$

thus recovering the result from Chap. 5.

7.4 Problems and Exercises

In the next three exercises also try solving the problem by finding first the autocovariance function of the output to see how hard the problem is in the time-domain framework.

1* The impulse response function of a linear system is $h(t) = 1 - t$, for $0 \le t \le 1$, and 0 elsewhere.

(a) Produce a graph of $h(t)$.
(b) Assume that the input is the standard white noise. Find the autocovariance function of the output.
(c) Find the power transfer function of the system, its equivalent-noise bandwidth, and half-power bandwidth.
(d) Assume that the input has the autocovariance function $\gamma_X(t) = 3/(1 + 4t^2)$. Find the power spectrum of the output signal.
(e) Assume that the input has the autocovariance function $\gamma_X(t) = \exp(-4|t|)$. Find the power spectrum of the output signal.
(f) Assume that the input has the autocovariance function $\gamma_X(t) = 1-|t|$ for $|t| < 1$ and 0 elsewhere. Find the power spectrum of the output signal.

2 The impulse response function of a linear system is $h(t) = e^{-2t}$, for $0 \le t \le 2$, and 0 elsewhere.

(a) Produce a graph of $h(t)$.
(b) Assume that the input is the standard white noise. Find the autocovariance function of the output.
(c) Find the power transfer function of the system, its equivalent-noise bandwidth, and half-power bandwidth.
(d) Assume that the input has the autocovariance function $\gamma_X(t) = 3/(1 + 4t^2)$. Find the power spectrum of the output signal.
(e) Assume that the input has the autocovariance function $\gamma_X(t) = \exp(-4|t|)$. Find the power spectrum of the output signal.
(f) Assume that the input has the autocovariance function $\gamma_X(t) = 1-|t|$ for $|t| < 1$ and 0 elsewhere. Find the power spectrum of the output signal.

3 The impulse response function of a linear system is $h(t) = e^{-0.05t}$, for $t \geq 10$, and 0 elsewhere.

(a) Produce a graph of $h(t)$.
(b) Assume that the input is the standard white noise. Find the autocovariance function of the output.
(c) Find the power transfer function of the system, its equivalent-noise bandwidth, and half-power bandwidth.
(d) Assume that the input has the autocovariance function $\gamma_X(t) = 3/(1 + 4t^2)$. Find the power spectrum of the output signal.
(e) Assume that the input has the autocovariance function $\gamma_X(t) = \exp(-4|t|)$. Find the power spectrum of the output signal.
(f) Assume that the input has the autocovariance function $\gamma_X(t) = 1-|t|$ for $|t| < 1$ and 0 elsewhere. Find the power spectrum of the output signal.

4 For a pair of random signals $X(t)$ and $Y(t)$, the *crosscovariance* γ_{XY} is defined as follows:

$$\gamma_{XY}(t, s) = \mathbf{E}((X(t) - \mu_X(t))(Y(s) - \mu_Y(s)).$$

Random signals $X(t)$ and $Y(t)$ are said to be *jointly stationary* if they are stationary and their crosscovariance satisfies the condition,

$$\gamma_{XY}(t, t + \tau) = \gamma_{XY}(\tau).$$

Consider random signals

$$X(t) = a\cos(2\pi(f_0 t + \Theta)), \qquad Y(t) = b\sin(2\pi(f_0 t + \Theta)),$$

where a and b are nonrandom constants and Θ is uniformly distributed on $[0, 1]$. Find the crosscovariance function for X and Y. Are these signals jointly stationary?

5* Consider the circuit shown in (Fig. 7.10).

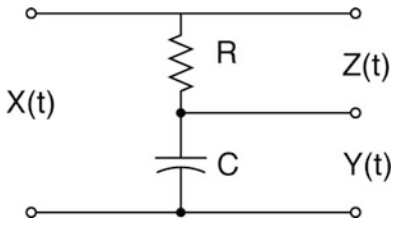
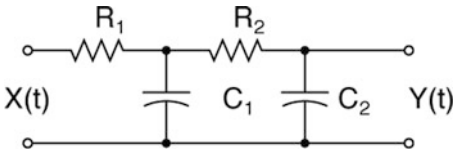


Fig. 7.10

Fig. 7.11



Assume that the input is the standard white noise.

(a) Find the power spectra $S_Y(f)$ and $S_Z(f)$ of the outputs $Y(t)$ and $Z(t)$.
(b) Find the crosscovariance,

$$\gamma_{YZ}(\tau) = \mathbf{E}\Big(Z(t)Y(t+\tau)\Big),$$

between those two outputs.

6 Find the output autocovariance sequence for the discrete-time system representing a stochastic difference equation described in Example 7.3.2. Use the Fourier series expansion of formula (7.3.12).

7 Consider the circuit shown in Fig. 7.11.

(a) Assume that the input is the standard white noise. Find the power spectrum $S_Y(f)$ and the autocovariance function $\gamma_Y(\tau)$ of the output $Y(t)$. *Hint:* Think about the above circuit as two simple RC filters in series.
(b) Find the half-power and equivalent-noise bandwidth for the system shown in Fig. 7.11 in the case when $R_1 = R_2$ and $C_1 = C_2$.

8 Show that a *continuum limit of RC filters in series* has a Gaussian p.d.f.-like power transfer function. Then prove that a white noise transmitted through such a filter yields a stationary signal on the output with Gaussian p.d.f.-like autocovariance function. More precisely, consider n rescaled RC filters in series, each with the time constant equal to $RC/\sqrt{n}$. Calculate its power transfer function, and take $n \to \infty$, to obtain the sought power transfer function of the form

$$|H(f)|^2 = e^{-(2\pi RCf)^2}.$$

Hint Use the basic calculus fact that $(1 + 1/x)^x \to e$, as $x \to \infty$. Then use the inverse Fourier transform to calculate the desired ACvF of the output. Note the fundamental fact: The (inverse) Fourier transform of a Gaussian p.d.f.-like function is also a Gaussian p.d.f.-like function.

Part III

Chapter 8
Optimization of Signal-to-Noise Ratio in Linear Systems

Abstract Useful, deterministic signals passing through various transmission devices often acquire extraneous random components due to, say, thermal noise in conducting materials, radio clutter or *aurora borealis* magnetic field fluctuations in the atmosphere, or deliberate jamming in warfare. If there exists some prior information about the nature of the original useful signal and the contaminating random noise it is possible to devise algorithms to improve the relative power of the useful component of the signal or, in other words, to increase the *signal-to-noise ratio* of the signal, by passing it through a filter designed for the purpose. In this short chapter we give a few examples of such designs just to show how the previously introduced techniques of analysis of random signals can be applied in this context.

8.1 Parametric Optimization for a Fixed Filter Structure

The general problem of optimization (maximization) of the signal-to-noise ratio in a linear system schematically pictured below

$$x(t) + N(t) \longrightarrow \boxed{\text{h(t)}} \longrightarrow y(t) + M(t)$$

can be formulated as follows: Consider a linear filter (system) characterized by its impulse response function $h(t)$ with the input signal $X(t)$ of the form

$$X(t) = x(t) + N(t), \tag{8.1.1}$$

where $x(t)$ is a deterministic "useful" signal, and $N(t)$ is a random stationary "noise" signal with zero-mean and autocovariance function $\gamma_N(t)$. Given the linearity of the system, the output signal $Y(t)$ is of the form

$$Y(t) = y(t) + M(t), \tag{8.1.2}$$

where the deterministic "useful" output component

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_8

$$y(t) = \int_{-\infty}^{\infty} x(s)h(t-s)ds, \tag{8.1.3}$$

and the "noise" output is a stationary zero-mean signal with the autocovariance function,

$$\gamma_M(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \gamma_N(\tau - s + u)h(s)h(u)\, ds\, du.$$

The task is as follows: given the shape of the input signal, design the structure of the filter which would maximize the signal-to-noise power ratio on the output. More precisely, we need to find an impulse response function $h(t)$ such that, for a given detection time t, the signal-to-noise ratio

$$\mathcal{S}/\mathcal{N} = \frac{\mathbf{PW}_y(t)}{\mathbf{E}(\mathbf{PW}_M)} \tag{8.1.4}$$

is maximized over all possible impulse response functions; in brief, we want to find $h(t)$ for which

$$\mathcal{S}/\mathcal{N} = \max.$$

Here, $\mathbf{PW}_y(t) = y^2(t)$ is the instantaneous power of the output signal, and $\mathbf{E}(\mathbf{PW}_M) = \gamma_M(0) = \sigma_M^2$ is the mean power of the output noise. Hence, the optimization problem is to find $h(t)$, and also detection time t_0, such that

$$\mathcal{S}/\mathcal{N} = \frac{y^2(t_0)}{\gamma_M(0)} = \frac{y^2(t_0)}{\sigma_M^2} = \max. \tag{8.1.5}$$

In the present section we will take a look at a relatively simple situation when the general structure of the filter is essentially fixed and only certain parameters, including the detection time t_0, need to be optimized.

To show the essence of our approach we will just consider the RC-filter with the impulse response function

$$h(t) = be^{-bt} \cdot u(t), \tag{8.1.6}$$

with a single parameter $b = 1/RC$ to be determined in addition to the optimal detection time t_0.

Suppose that the "useful" input signal we are trying to detect on the output is a rectangular impulse

$$x(t) = \begin{cases} A, & \text{for } 0 \le t \le T; \\ 0, & \text{elsewhere,} \end{cases} \tag{8.1.7}$$

and that the input noise is a white noise of "amplitude" N_0, with the autocovariance $\gamma_N(t) = N_0\delta(t)$.

The deterministic "useful" output signal

$$y(t) = \int_{-\infty}^{\infty} x(s)h(-(s-t))ds$$

$$= \begin{cases} \int_0^t Abe^{-b(t-s)}\,ds & \text{for } 0 < t < T; \\ \int_0^T Abe^{-b(t-s)}\,ds & \text{for } t \ge T, \end{cases}$$

$$= \begin{cases} A(1-e^{-bt}), & \text{for } 0 < t \le T; \\ A(1-e^{-bT})e^{-b(t-T)}, & \text{for } t \ge T. \end{cases} \tag{8.1.8}$$

It is pictured in Fig. 8.1 below.

Clearly, the maximum of the output signal is attained at $t_0 = T$. On the other hand, as calculated in Chap. 6, the autocovariance function of the output noise,

$$\gamma_M(\tau) = N_0\frac{b}{2}e^{-b\tau},$$

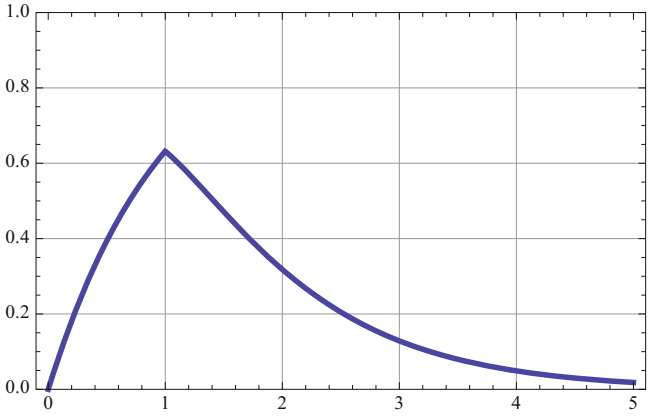


Fig. 8.1 Response $y(t)$ (8.1.8) of the RC-filter (8.1.6) to the rectangular input signal $x(t)$ (8.1.7). The parameter values are $T = 1$, $A = 1$, and $b = 1/RC = 1$. The maximum is clearly attained for $t_0 = T$

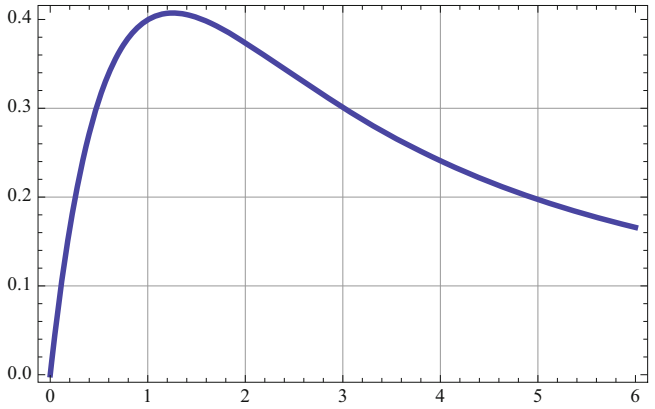


Fig. 8.2 Graph of the factor $(1 - e^{-z})^2/z$ in the formula (8.1.9) for the signal-to-noise ratio $\mathcal{S}/\mathcal{N}(z)$

so that, at already optimized detection time $t_0 = T$,

$$\frac{\mathcal{S}}{\mathcal{N}} = \frac{y^2(T)}{\gamma_M(0)} = \frac{A^2T[1 - e^{-bT}]^2}{bN_0/2}.$$

To simplify our calculations we will substitute $z = bT$. Now, our final task is to find the maximum of the function

$$\frac{\mathcal{S}}{\mathcal{N}}(z) = \frac{2A^2T}{N_0} \cdot \frac{(1 - e^{-z})^2}{z} \tag{8.1.9}$$

of one variable z. Function $\mathcal{S}/\mathcal{N}(z)$, although simple-looking, is a little tricky and we will start exploration of its maximum by graphing it, see Fig. 8.2. To find the location of the maximum we calculate the derivative and try to solve the equation

$$\frac{d}{dz}\frac{(1 - e^{-z})^2}{z} = \frac{2(1 - e^{-z})e^{-z}z - (1 - e^{-z})^2}{z^2} = 0.$$

Although the above equation can be easily simplified to the equation

$$e^z - 1 - 2z = 0,$$

the latter cannot be solved explicitly. So, as usual, as the first step we explore the solution graphically, see Fig. 8.3. The nontrivial zero is approximately at $z_{\max} = 1.25$, which gives $b_{\max} = 1.25/T$ so that the optimal RC constant

$$RC_{\max} \approx \frac{1}{b_{\max}} = \frac{T}{1.25} = 0.8T. \tag{8.1.10}$$

Note that $RC_{\max}$ is independent of the "amplitude," N_0, of the input noise.

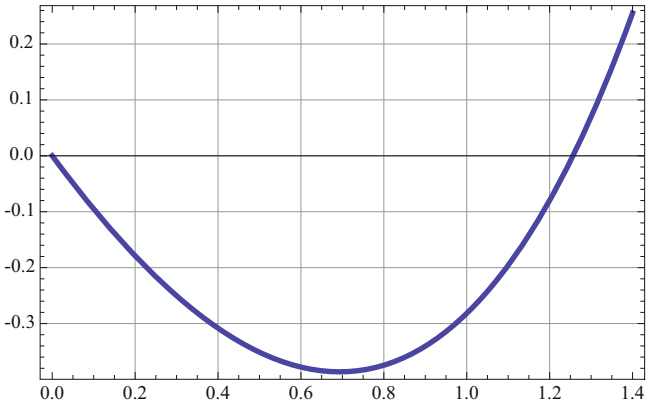


Fig. 8.3 A plot of function $e^z - 1 - 2z = 0$. The nontrivial zero is approximately at $z_{\max} = 1.25$

Evaluated at the optimal values of parameters, t_0 and b, the maximum available signal-to-noise ratio is

$$\frac{\mathcal{S}}{\mathcal{N}}_{\max} \approx \frac{y^2(T)}{b_{\max} N_0/2} = \frac{2A^2T[1 - e^{-b_{\max}T}]^2}{b_{\max} N_0} = 0.81 \cdot \frac{A^2T}{N_0}. \tag{8.1.11}$$

It is proportional to signal's duration T, and to the square of its amplitude A, but inversely proportional to the "amplitude," N_0, of the noise.

8.2 Filter Structure Matched to Input Signal

In this section we will solve a more ambitious problem of designing the structure of the filter to maximize the signal-to-noise ratio on the output rather than just optimizing filter parameters. To be more precise, the task at hand is to find an impulse response function $h(t)$, and the detection time t_0, such that

$$\mathcal{S}/\mathcal{N} = \frac{y^2(t_0)}{\sigma_M^2} = \max, \tag{8.2.1}$$

for a given deterministic (nonrandom) input signal $x(t)$ transmitted in the presence of the white noise input $N(t)$ with autocovariance function $\gamma_N(t) = N_0\delta(t)$, where, as before, $x(t) = 0$, for $t \leq 0$, and

$$y(t) = \int_0^\infty x(t - s)h(s)\, ds. \tag{8.2.2}$$

For the output noise,

$$\sigma_M^2 = \gamma_M(0) = N_0 \int_0^\infty \Big(\int_0^\infty \delta(u-s)h(u)\,du \Big) h(s)\,ds = N_0 \int_0^\infty h^2(s)\,ds. \tag{8.2.3}$$

In this situation

$$\mathcal{S}/\mathcal{N} = \frac{y^2(t_0)}{\sigma_M^2} = \frac{(\int_0^\infty x(t_0-s)h(s)\,ds)^2}{N_0 \int_0^\infty h^2(s)ds}. \tag{8.2.4}$$

In view of the Cauchy-Schwartz Inequality,

$$\mathcal{S}/\mathcal{N} \le \frac{\int_0^\infty x^2(t_0-s)\,ds \cdot \int_0^\infty h^2(s)\,ds}{N_0 \int_0^\infty h^2(s)ds} = \frac{1}{N_0} \int_0^\infty x^2(t_0-s)\,ds, \tag{8.2.5}$$

with the equality, that is the maximum for $\mathcal{S}/\mathcal{N}$, achieved when the two factors, $h(s)$ and $x(t_0-s)$, in the scalar product in the numerator of (8.2.4) are linearly dependent. In other words, for any constant c, the impulse response function

$$h(s) = cx(t_0-s)u(s) = cx(-(s-t_0))u(s) \tag{8.2.6}$$

gives the optimal structure of the filter and maximizes the $\mathcal{S}/\mathcal{N}$ ratio. This, so-called, *matching filter* has the impulse response function equal to the input signal $x(t)$ run backwards in time, then shifted to the right by t_0, and, finally, cut off at 0.

With the selection of the matching filter, in view of (8.2.4), the maximal value of the $\mathcal{S}/\mathcal{N}$ ratio is

$$\mathcal{S}/\mathcal{N}_{\max} = \frac{(\int_0^\infty x(t_0-s)cx(t_0-s)u(s)\,ds)^2}{N_0 \int_0^\infty (cx(t_0-s)u(s))^2\,ds} = \frac{\int_0^\infty x^2(t_0-s)\,ds}{N_0}. \tag{8.2.7}$$

Example 8.2.1 (Matching Filter for a Rectangular Input Signal) Consider a rectangular input signal of the form

$$x(t) = \begin{cases} A, & \text{for } 0 < t < T; \\ 0, & \text{elsewhere,} \end{cases}$$

transmitted in the presence of an additive white noise with autocovariance function $\gamma_N(t) = N_0\delta(t)$. According to formula (8.2.6), its matching filter at detection time t_0, is

$$h(t) = \begin{cases} A, & \text{for } 0 < t < t_0; \\ 0, & \text{elsewhere,} \end{cases}$$

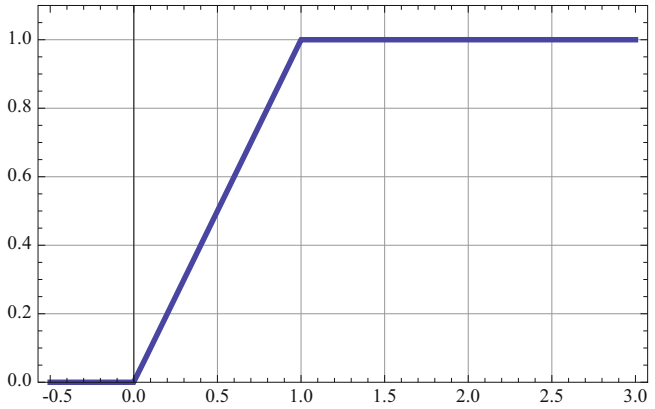


Fig. 8.4 The dependence of the optimal signal-to-noise ratio on the detection time t_0 for the matching filter from Example 8.2.1. The input signal is the sum of a rectangular signal of amplitude $A = 1$, and duration $T = 1$, and the white noise with autocovariance function $\gamma_N(t) = \delta(t)$

if $0 \leq t_0 \leq T$, and

$$h(t) = \begin{cases} A, & \text{for } t_0 - T < t < t_0; \\ 0, & \text{elsewhere,} \end{cases}$$

if $t_0 > T$. So the $\mathcal{S}/\mathcal{N}_{\max}$, as a function of the detection time t_0, is

$$\mathcal{S}/\mathcal{N}_{\max}(t_0) = \begin{cases} A^2 t_0/N_0, & \text{for } 0 < t_0 < T; \\ A^2 T/N_0, & \text{for } t_0 > T. \end{cases}$$

Clearly, the earliest detection time t_0 to maximize $\mathcal{S}/\mathcal{N}_{\max}(t_0)$ is $t_0 = T$ (see Fig. 8.4).

At the optimal detection time $t_0 = T$, or any later detection time,

$$\mathcal{S}/\mathcal{N}_{\max} = \frac{A^2 T}{N_0}. \tag{8.2.8}$$

This result should be compared with the maximum signal-to-noise ratio $0.81A^2T/N_0$ (see (8.1.11)) obtained in Sect. 8.1 by optimally tuning the RC-filter: the best matching filter gives about a 25% gain in the signal-to-noise ratio over the best RC-filter.

It is also instructive to trace the behavior of the deterministic part $y(t)$ of the output signal for the matching filter as a function of detection time t_0. The formula (8.2.2) applied to the matching filter immediately gives that, for $0 < t_0 < T$,

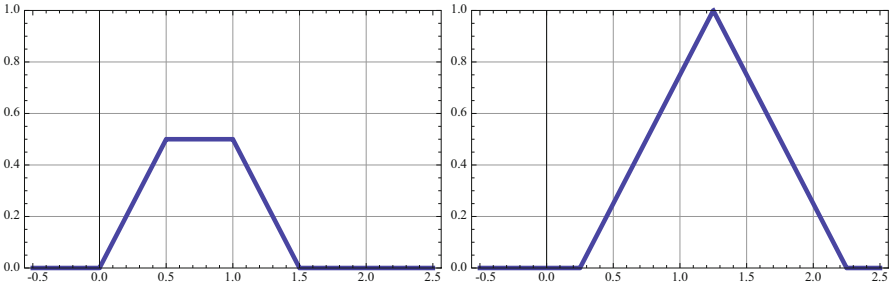


Fig. 8.5 The response $y(t)$ of the matching filter for the rectangular input signal with amplitude $A = 1$ and duration $T = 1$ (see Example 8.2.1). *Left:* For detection time $t_0 = 0.5 < T = 1$. *Right:* For detection time $t_0 = 1.25 > T = 1$

$$y(t) = \begin{cases} A^2 t, & \text{for } 0 < t < t_0; \\ A^2 t_0, & \text{for } t_0 < t < T; \\ -A^2(t - (t_0 + T)), & \text{for } T < t < t_0 + T; \\ 0, & \text{elsewhere,} \end{cases} \tag{8.2.9}$$

and, for $t_0 \geq T$,

$$y(t) = \begin{cases} A^2(t - (t_0 - T)), & \text{for } t_0 - T < t < t_0; \\ -A^2(t - (t_0 + T)), & \text{for } t_0 < t < t_0 + T; \\ 0, & \text{elsewhere.} \end{cases} \tag{8.2.10}$$

These two output signals are pictured in Fig. 8.5.

8.3 The Wiener Filter

Acausal Filter Given stationary random signals $X(t)$ and $Y(t)$ the problem is to find a (not necessarily causal) impulse response function $h(t)$ such that the mean-square distance between $Y(t)$ and the output signal,

$$Y_h(t) = \int_{-\infty}^{\infty} X(t - s)h(s)\, ds,$$

is smallest possible. In other words, we need $h(t)$ minimizing the error quantity

$$\mathbf{E}\Big(Y(t) - Y_h(t)\Big)^2.$$

In the space of all finite variance (always zero-mean) random quantities equipped with the covariance as the scalar product, the best approximation $Y_h(t)$ of a random quantity $Y(t)$ by elements of the linear subspace $\mathcal{X}$ spanned by linear combinations of values of $X(t-s)$, $-\infty < s < \infty$, is given by the orthogonal projection of $X(t)$ on $\mathcal{X}$.[1] That means that the difference $Y(t) - Y_h(t)$ must be orthogonal to all $X(t-s)$, $-\infty < s < \infty$, or more formally,

$$\mathbf{E}\Big((Y(t) - Y_h(t)) \cdot X(t-s)\Big)$$

$$= \mathbf{E}\Big(Y(t) \cdot X(t-s)\Big) - \mathbf{E}\Big(\int_{-\infty}^{\infty} X(t-u)h(u)\,du \cdot X(t-s)\Big)$$

$$= \gamma_{YX}(s) - \int_{-\infty}^{\infty} \gamma_X(s-u)h(u)\,du = 0,$$

for all s, $-\infty < s < \infty$. Hence, the optimal $h(t)$ can be found by solving, for each s, the integral equation

$$\gamma_{YX}(s) = \int_{-\infty}^{\infty} \gamma_X(s-u)h(u)\,du, \tag{8.3.1}$$

which involves only the autocovariance function $\gamma_X(s)$ and the crosscorrelation function $\gamma_{YX}(s)$. The solution is found readily in the frequency domain. Remembering that the Fourier transform of a convolution is the product of Fourier transforms, and denoting by $H(f)$ the transfer function (the Fourier transform of the impulse response function) of the optimal $h(t)$, Eq. (8.3.1) can be rewritten in the form

$$S_{YX}(f) = S_X(f) \cdot H(f),$$

which immediately gives the explicit formula for the transfer function of the optimal filter:

$$H(f) = \frac{S_{YX}(f)}{S_X(f)}. \tag{8.3.2}$$

The minimal error can then also be calculated explicitly:

$$\mathbf{E}\Big(Y(t) - Y_h(t)\Big)^2 = \gamma_Y(0) - \int_{-\infty}^{\infty} \gamma_{YX}(s)h(s)\,ds, \tag{8.3.3}$$

[1] This argument is analogous to the one encountered in Chap. 2, when we discussed the best approximation in power of deterministic periodic signals by their Fourier series.

or, in terms of the optimal transfer function, using the Parseval formula for the last integral, we have

$$\mathbf{E}\Big(Y(t) - Y_h(t)\Big)^2 = \int_{-\infty}^{\infty} \Big(S_Y(f) - S^*_{YX}(f)H(f)\Big)\, df. \tag{8.3.4}$$

Example 8.3.1 (Filtering White Noise Out of a Stationary Signal) Assume that signal $X(t)$ is the sum of a "useful" signal $Y(t)$ and noise $N(t)$, that is $X(t) = Y(t) + N(t)$, where $Y(t)$ has the power spectrum

$$S_Y(f) = \frac{1}{1+f^2},$$

and is uncorrelated with the white noise $N(t)$, which is assumed to have the power spectrum $S_N(f) \equiv 1$. Then,

$$S_{YX}(f) = S_Y(f) = \frac{1}{1+f^2}, \qquad \text{and} \qquad S_X(f) = S_Y(f) + S_N(f) = \frac{2+f^2}{1+f^2}.$$

The transfer function of the optimal filter is then

$$H(f) = \frac{S_{YX}(f)}{S_X(f)} = \frac{1}{2+f^2},$$

with the corresponding impulse response function

$$h(t) = \frac{1}{2\sqrt{2}} e^{-\sqrt{2}|t|},$$

and the error

$$\mathbf{E}\Big(Y(t)-Y_h(t)\Big)^2 = \int_{-\infty}^{\infty} \Big(\frac{1}{1+f^2} - \frac{1}{1+f^2}\cdot\frac{1}{2+f^2}\Big)\, df = \int_{-\infty}^{\infty} \frac{1}{2+f^2}\, df = \frac{\pi}{\sqrt{2}}.$$

Causal Filter For given stationary random signals $X(t)$ and $Y(t)$, the construction of the optimal *causal filter* requires finding a causal impulse response function $h(t) = 0$, for $t \le 0$, such that the error

$$\mathbf{E}\left(Y(t) - \int_0^{\infty} X(t-s)h(s)\, ds\right)^2$$

is minimal. In other words, we are trying to find the best mean-square approximation to $Y(t)$ by (continuous) linear combinations of the *past* values of $X(t)$. Using the same orthogonality argument we applied for the acausal optimal filter we obtain another integral equation for the optimal $h(t)$:

$$\gamma_{YX}(s) = \int_0^\infty \gamma_X(s-u)h(u)\,du,$$

this time valid only for all $s > 0$. This equation is traditionally called the *Wiener-Hopf equation*. It is clear that to solve the above equation via an integral transform method we have to replace the Fourier transform used in the acausal case by the Laplace transform. However, the details here are more involved and, for the solution, we refer the reader to the literature of the subject.[2]

8.4 Problems and Exercises

1 The triangular signal $x(t) = 0.01t$, for $0 < t < 0.01$, and 0 elsewhere, is combined with white noise having a flat power spectrum of $2\,V^2/Hz$. Find the value of the RC-constant such that the signal-to-noise ratio at the output of the RC-filter is maximal at $t = 0.01$ s.

2* A signal of the form $x(t) = 5e^{-(t+2)}u(t)$ is to be detected in the presence of white noise with a flat power spectrum of $0.25\ V^2/Hz$ using a matched filter.

(a) For $t_0 = 2$ find the value of the impulse response of the matched filter at $t = 0, 2, 4$.
(b) Find the maximum output signal-to-noise ratio that can be achieved if $t_0 = \infty$.
(c) Find the detection time t_0 that should be used to achieve output signal-to-noise ratio that is equal to 95% of the maximum signal-to-noise ratio discovered in part (b).
(d) The signal $x(t) = 5e^{-(t+2)}u(t)$ is combined with white noise having a power spectrum of $2\ V^2/Hz$. Find the value of RC such that the signal/noise at the output of the RC filter is maximal at $t = 0.01$ s.

3 Repeat the construction of the optimal filter from Example 8.3.1 in the case when the useful signal $Y(t)$ has a more general power spectrum

$$S_Y(f) = \frac{a}{b^2 + f^2},$$

and the uncorrelated white noise $N(t)$ has arbitrary power spectrum $S_N(f) \equiv \mathcal{N}$. Discuss the properties of this filter when the noise power is much bigger than the power of the useful signal, that is, when $\mathcal{N} \gg S_Y(f)$? Construct the optimal acausal filters for other selected spectra of $Y(t)$ and $N(t)$.

[2] Norbert Wiener's original *Extrapolation, Interpolation, and Smoothing of Stationary Time series*, MIT Press and J. Wiley, New York 1950, is still very readable, but also see Chap. 10 of A. Papoulis, *Signal Analysis*, McGraw-Hill, New York, 1977.

Chapter 9
Gaussian Signals, Covariance Matrices, and Sample Path Properties

Abstract In general, determination of the shape of the sample paths of a random signal $X(t)$ requires knowledge of n-D (or, in the terminology of signal processing, n-point) probabilities

$$\mathbf{P}\Big(a_1 < X(t_1) < b_1, \ldots, a_n < X(t_n) < b_n\Big),$$

for an arbitrary n, and arbitrary windows $a_1 < b_1, \ldots, a_n < b_n$. But, usually, this information cannot be recovered if the only signal characteristic known is the autocorrelation function. The latter depends on the 2-point distributions but does not uniquely determine them. However, in the case of Gaussian signals, the autocovariances determine not only 2-point probability distributions but also all the n-point probability distributions, so that complete information is available within the second-order theory. In particular, that means that you only have to estimate means and covariances to obtain the complete model. Also, in the Gaussian universe, the weak stationarity implies the strict stationarity as defined in Chap. 4. For the sake of simplicity all signals in this chapter are assumed to be real-valued. The chapter ends with a more subtle analysis of sample paths properties of stationary signals such as continuity and differentiability; in the Gaussian case these issues have fairly complete answers.

Of course, faced with real-world data the proposition that they are distributed according to a Gaussian distribution must be tested rigorously. Many such tests have been developed by the statisticians.[1] In other cases, one can make an argument in favor of such a hypothesis based on the Central Limit Theorem (4.5.5) and (4.5.6).

9.1 Linear Transformations of Random Vectors

In Chap. 3 we have calculated probability distributions of transformed random quantities. Repeating that procedure in the case of a linear transformation of the

[1] See, e.g., M. Denker and W.A. Woyczyński's book mentioned in previous chapters.

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_9

1D random quantity X given by the formula

$$Y = aX, \qquad a > 0, \tag{9.1.1}$$

we can obtain the cumulative distribution function (c.d.f.) $F_Y(y)$ of the random quantity Y in terms of the c.d.f. $F_X(x)$ of the random quantity X as follows:

$$F_Y(y) = P(Y \le y) = P(aX \le y) = P(X \le y/a) = F_X(y/a). \tag{9.1.2}$$

To obtain an analogous formula for the probability density functions (p.d.f.) it suffices to differentiate both sides of (9.1.2) to see that

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{a} f_X\Big(\frac{y}{a}\Big). \tag{9.1.3}$$

Example 9.1.1 Consider a standard 1D Gaussian random quantity $X \sim N(0, 1)$ with the p.d.f.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{9.1.4}$$

Then the random quantity, $Y = aX,\ a > 0$, has the p.d.f.,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\, a} e^{-\frac{y^2}{2a^2}}. \tag{9.1.5}$$

Obviously, the expectation

$$\mathbf{E}Y = \mathbf{E}(aX) = a\mathbf{E}X = 0,$$

and the variance of Y is

$$\sigma_Y^2 = \mathbf{E}(aX)^2 = a^2 \mathbf{E}X^2 = a^2. \tag{9.1.6}$$

If we conduct the same argument for $a < 0$, the p.d.f. of $Y = aX$ will be

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,(-a)} e^{-\frac{y^2}{2a^2}}. \tag{9.1.7}$$

Thus formulas (9.1.6) and (9.1.7) can be unified in a single statement: If $X \sim N(0, 1)$, then, for any $a \neq 0$, random quantity $Y = aX$ has p.d.f.

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,|a|} e^{-\frac{y^2}{2a^2}}. \tag{9.1.8}$$

Using the above elementary reasoning as a model we will now derive the formula for a d-dimensional p.d.f.

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{Y}}(y_1, \ldots, y_d)$$

of a random (column) vector

$$\vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix}$$

obtained by a nondegenerate linear transformation

$$\vec{Y} = \mathbf{A}\vec{X} \tag{9.1.9}$$

consisting of multiplication of the random vector

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix},$$

with a known p.d.f.

$$f_{\vec{X}}(\vec{x}) = f_{\vec{X}}(x_1, \ldots, x_d),$$

by a fixed nondegenerate nonrandom matrix

$$\mathbf{A} = \begin{pmatrix} a_{11}, \cdots, a_{1d} \\ \cdots \\ a_{d1}, \cdots, a_{dd} \end{pmatrix}.$$

In other words, we assume that $\det(A) \neq 0$, or, equivalently, that the rows of the matrix A form a linearly independent system of vectors.

In terms of its coordinates the result of the linear transformation (9.1.9) can be written in the explicit form

$$\vec{Y} = \begin{pmatrix} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1d}X_d \\ a_{21}X_1 + a_{22}X_2 + \cdots + a_{2d}X_d \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{d1}X_1 + a_{d2}X_2 + \cdots + a_{dd}X_d \end{pmatrix}.$$

To calculate the probability distribution of $\vec{Y}$ following the above 1D approach we must make use of the essential assumption of invertibility of the matrix $\mathbf{A}$, an

analogue of the assumption $a \neq 0$ in the 1D case. Then, for a domain D in the d-dimensional space $\mathbf{R}^d$,

$$\mathbf{P}(\vec{Y} \in D) = \mathbf{P}(\mathbf{A}\vec{X} \in D) = \mathbf{P}(\vec{X} \in \mathbf{A}^{-1}D). \tag{9.1.10}$$

This identity can be rewritten in terms of p.d.f.s of $\vec{Y}$ and $\vec{X}$ as follows:

$$\int_D f_{\vec{Y}}(\vec{y})\, dy_1 \cdot \ldots \cdot dy_d = \int_{\mathbf{A}^{-1}D} f_{\vec{X}}(\vec{x})\, dx_1 \cdot \ldots \cdot dx_d.$$

Making a substitution $\vec{x} = \mathbf{A}^{-1}\vec{z}$ in the second integral, in view of the d-dimensional change of variables formula, we get that

$$\int_D f_{\vec{Y}}(\vec{y})\, dy_1 \cdot \ldots \cdot dy_d = \int_D f_{\vec{X}}(\mathbf{A}^{-1}\vec{z}) \cdot |\det(\mathbf{A}^{-1})|\, dz_1 \cdot \ldots \cdot dz_d,$$

where $\det(\mathbf{A}^{-1})$ is just the Jacobian of the substitution $\vec{x} = \mathbf{A}^{-1}\vec{z}$. Remembering that the determinant of the inverse matrix $\mathbf{A}^{-1}$ is the reciprocal of the determinant of the matrix $\mathbf{A}$, we get the identity

$$\int_D f_{\vec{Y}}(\vec{y})\, dy_1 \cdot \ldots dy_d = \int_D \frac{f_{\vec{X}}(\mathbf{A}^{-1}\vec{z})}{|\det(\mathbf{A})|}\, dz_1 \cdot \ldots \cdot dz_d.$$

Since this identity holds true for any domain D, the integrands on both sides must be equal which gives the final formula for the p.d.f. of $\vec{Y}$:

$$f_{\vec{Y}}(\vec{y}) = \frac{f_{\vec{X}}(\mathbf{A}^{-1}\vec{y})}{|\det(\mathbf{A})|}, \quad \text{if} \quad \det(\mathbf{A}) \neq 0. \tag{9.1.11}$$

The 1D formula (9.1.3) is, obviously, the special case of the above general result.

9.2 Gaussian Random Vectors

As in the one-dimensional case, all nondegenerate zero-mean d-dimensional Gaussian random vectors can be obtained as nondegenerate linear transformations of a standard d-D Gaussian random vector

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

in which the coordinates $X_1, \ldots, X_d$ are independent $N(0, 1)$ random quantities. Because of their independence, the d-dimensional p.d.f. of $\vec{X}$ is the product of 1D $N(0, 1)$ p.d.f.s and is thus of the product form

$$f_{\vec{X}}(\vec{x}) = \frac{e^{\frac{-x_1^2}{2}}}{\sqrt{2\pi}} \cdot \ldots \cdot \frac{e^{\frac{-x_d^2}{2}}}{\sqrt{2\pi}} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}(x_1^2 + \cdots + x_d^2)}$$

$$= \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\|\vec{x}\|^2} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\vec{x}^T\vec{x}}, \tag{9.2.1}$$

where $\|\vec{x}\|$ stands for the norm (magnitude) of the vector $\vec{x}$, and the superscript T denotes the transpose of a matrix. Indeed,

$$\vec{x}^T\vec{x} = (x_1, \ldots, x_d) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} = x_1^2 + \cdots + x_d^2 = \|\vec{x}\|^2.$$

It is the latter form in (9.2.1) that will be useful now in applying formula (9.1.11). Indeed, substituting the last expression for $f_{\vec{X}}(\vec{x})$ in (9.2.1) into (9.1.11), one immediately gets,[2]

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{(2\pi)^{d/2}|\det(A)|} e^{-\frac{1}{2}\|\mathbf{A}^{-1}\vec{y}\|^2}$$

$$= \frac{1}{(2\pi)^{d/2}|\det(A)|} e^{-\frac{1}{2}(\mathbf{A}^{-1}\vec{y})^T \cdot (\mathbf{A}^{-1}\vec{y})}$$

$$= \frac{1}{(2\pi)^{d/2}|\det(A)|} e^{-\frac{1}{2}\vec{y}^T(\mathbf{A}\mathbf{A}^T)^{-1}\vec{y}}. \tag{9.2.2}$$

Thus formula (9.2.2) gives the general form of the d-dimensional zero-mean Gaussian p.d.f., and just as we identified the parameter a^2 in the 1D case (9.1.5) and (9.1.6) as the variance of the random quantity Y, we can identify entries of the matrix

$$\mathbf{\Gamma} = \mathbf{A}\mathbf{A}^T \tag{9.2.3}$$

appearing in the exponent in (9.2.2) as statistically significant parameters of the random vector $\vec{Y}$.

[2]Remember that, for any matrices $\mathbf{M}$, and $\mathbf{N}$, we have $(\mathbf{MN})^T = \mathbf{N}^T\mathbf{M}^T$, $(\mathbf{MN})^{-1} = \mathbf{N}^{-1}\mathbf{M}^{-1}$, and $(\mathbf{M}^T)^{-1} = (\mathbf{M}^{-1})^T$.

To see what they are let us first calculate the entries γ_{ij}, $i, j = 1, 2, \ldots, d$, of matrix $\mathbf{\Gamma}$:

$$\gamma_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots + a_{id}a_{jd}. \tag{9.2.4}$$

On the other hand, covariances (we are working with zero-mean vectors!) of different components of random vector $\vec{Y}$

$$\begin{aligned}\mathbf{E}(Y_iY_j) &= \mathbf{E}\Big((a_{i1}X_1 + \ldots + a_{id}X_d)\cdot(a_{j1}X_1 + \ldots + a_{jd}X_d)\Big)\\ &= a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots + a_{id}a_{jd},\end{aligned} \tag{9.2.5}$$

because $\mathbf{E}X_iX_j = 1$, if $i = j$, and $= 0$, if $i \neq j$.

So, it turns out that

$$\mathbf{\Gamma} = (\gamma_{ij}) = (\mathbf{E}Y_iY_j), \tag{9.2.6}$$

and matrix $\Gamma = (\gamma_{ij})$ is simply the covariance matrix of the general zero-mean Gaussian random vector $\vec{Y}$. Thus, since

$$\det(\mathbf{\Gamma}) = \det(\mathbf{A}\mathbf{A}^T) = \det(\mathbf{A})\cdot\det(\mathbf{A}^T) = (\det(\mathbf{A}))^2,$$

we finally get that the p.d.f. of $\vec{Y}$ can be written in the form

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{(2\pi)^{d/2}|\det(\mathbf{\Gamma})|^{1/2}}e^{-\frac{1}{2}\vec{y}^T\mathbf{\Gamma}^{-1}\vec{y}}, \tag{9.2.7}$$

where $\mathbf{\Gamma}$ is the covariance matrix of $\vec{Y}$ satisfying the nondegeneracy condition $\det(\mathbf{\Gamma}) \neq 0$.

Remark 9.2.1 (Gaussian Random Vectors with Nonzero Mean) Of course, to get the p.d.f. of a general Gaussian random vector with nonzero expectation

$$\mathbf{E}\vec{Y} = \vec{\mu} = (\mu_1, \ldots, \mu_d)^T,$$

it suffices to shift the p.d.f. (9.2.7) by $\vec{\mu}$ to obtain that

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{(2\pi)^{d/2}|\det(\mathbf{\Sigma})|^{1/2}}e^{-\frac{1}{2}(\vec{y}-\vec{\mu})^T\Sigma^{-1}(\vec{y}-\vec{\mu})}, \tag{9.2.7}$$

where

$$\mathbf{\Sigma} = (\sigma_{ij}) = (\mathbf{E}(Y_i - \mu_i)(Y_j - \mu_j)) \tag{9.2.8}$$

is the *covariance matrix* of $\vec{Y}$. A Gaussian random vector with joint p.d.f. given by formulas (8.2.7) and (8.2.8) is often called a normal $N(\vec{\mu}, \Sigma)$ random vector.

Example 9.2.1 (2D Zero-Mean Gaussian Random Vectors (See, Also, Example 4.3.2)) Let us carry out the above calculation explicitly in the special case of dimension $d = 2$. Then the covariance matrix

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{E}Y_1Y_1 & \mathbf{E}Y_1Y_2 \\ \mathbf{E}Y_2Y_1 & \mathbf{E}Y_2Y_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

where the variances of coordinate vectors

$$\sigma_1^2 = \mathbf{E}Y_1^2, \qquad \sigma_2^2 = \mathbf{E}Y_2^2,$$

and the correlation coefficient of the two components

$$\rho = \frac{\mathbf{E}Y_1Y_2}{\sigma_1\sigma_2}.$$

The determinant of the covariance matrix

$$\det(\mathbf{\Gamma}) = \sigma_1^2\sigma_2^2(1-\rho^2),$$

and its inverse

$$\mathbf{\Gamma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho \\ -\sigma_1\sigma_2\rho & \sigma_1^2 \end{pmatrix}.$$

Hence, the p.d.f. of a general zero-mean 2D Gaussian random vector is of the form

$$f_{\vec{\mathbf{Y}}}(y_1, y_2) = \frac{1}{(2\pi)^{2/2}\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$

$$\exp\left[-\frac{1}{2}(y_1, y_2) \frac{\begin{pmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho \\ -\sigma_1\sigma_2\rho & \sigma_1^2 \end{pmatrix}}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right],$$

which, after performing prescribed matrix algebra, leads to the final expression

$$f_{\vec{\mathbf{Y}}}(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{1}{2(1-\rho^2)} \left(\frac{y_1^2}{\sigma_1^2} - 2\rho\frac{y_1y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2} \right) \right]. \tag{9.2.9}$$

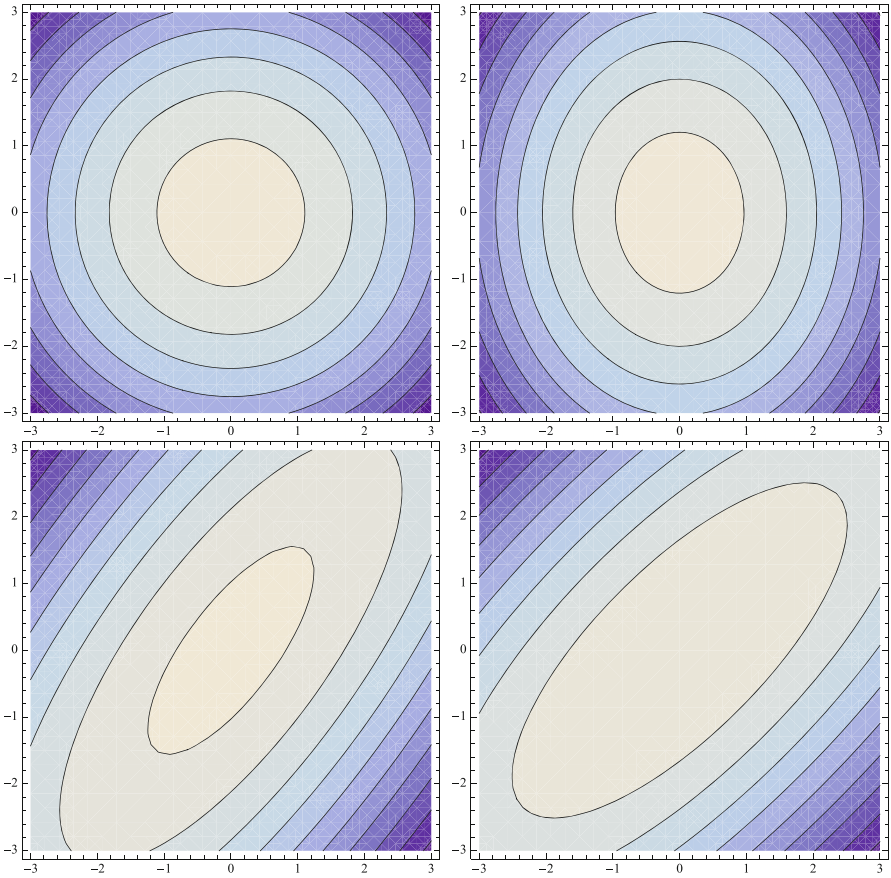


Fig. 9.1 Level curves for the 2D Gaussian probability density functions $f_{\vec{Y}}(y_1, y_2)$ (9.2.9), for the following selection of parameters $(\rho, \sigma_1, \sigma_2)$ (clockwise, from the top left corner): (0,9,9), (0,8,10), (3/4, 9,9), and (3/4,8,10). There are nine level curves in each plot, equally spaced between level zero and the maximum of the p.d.f

The plots of the above densities are bell-shaped surfaces and we have seen one example of such a surface in Chap. 4 (Fig. 4.14). The level curves of these densities, described by the equations

$$\frac{y_1^2}{\sigma_1^2} - 2\rho\frac{y_1 y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2} = \text{const},$$

are ellipses in the (y_1, y_2)-plane, with semiaxes and orientations depending on the parameters ρ, σ_1, and σ_2, representing, respectively, the correlation coefficient between the two components of the Gaussian random vector $\vec{Y}$, the variance of the first, and of the second component. Figure 9.1 shows the level curves of 2D Gaussian densities for four selections of the above three parameters.

9.3 Gaussian Stationary Signals

By definition, a nondegenerate zero-mean random signal $X(t)$ is Gaussian if, for any positive integer N, and any selection of sampling times $t_1 < t_2 < \cdots < t_N$, the random vector

$$\vec{X}_{(t_1,\ldots,t_N)} = \begin{pmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_N) \end{pmatrix} \tag{9.3.1}$$

is a Gaussian zero-mean random vector with nondegenerate covariance matrix. Thus, in view of results of Sect. 9.2, its N-dimensional joint p.d.f. $f_{(t_1,\ldots,t_N)}(x_1,\ldots,x_N)$ is given by the formula[3]

$$f_{(t_1,\ldots,t_N)}(x_1,\ldots,x_N) = \frac{1}{(2\pi)^{N/2}|\det(\mathbf{\Gamma})|^{1/2}} \cdot e^{-\frac{1}{2}\vec{x}^T\mathbf{\Gamma}^{-1}\vec{x}}, \quad \det(\mathbf{\Gamma}) \neq 0, \tag{9.3.2}$$

where $\mathbf{\Gamma}$ is the $N \times N$ covariance matrix

$$\mathbf{\Gamma} = \mathbf{\Gamma}_{(t_1,\ldots,t_N)} = (\gamma_X(t_i,t_j)) = (\mathbf{E}X(t_i)X(t_j)). \tag{9.3.3}$$

Thus, in view of (9.3.1) and (9.3.2), *the only information needed to completely determine all finite-dimensional joint probability distributions of a zero-mean Gaussian random signal $X(t)$ is the knowledge of its autocovariance function*

$$\gamma_X(s,t) = \mathbf{E}X(s)X(t).$$

For stationary Gaussian signals the situation is simpler yet as the autocovariance function $\gamma_X(s,t)$ is just a function of a single variable:

$$\gamma_X(s,t) = \gamma_X(t-s).$$

Thus the covariance matrix $\mathbf{\Gamma}$ for a stationary random signal $X(t)$ sampled at $t_1, t_2, \ldots, t_N$, is of the form

$$\mathbf{\Gamma}_{(t_1,\ldots,t_N)} = \begin{pmatrix} \gamma_X(0) & \gamma_X(t_2-t_1) & \gamma_X(t_3-t_1) & \ldots & \gamma_X(t_N-t_1) \\ \gamma_X(t_1-t_2) & \gamma_X(0) & \gamma_X(t_3-t_2) & \cdots & \gamma_X(t_N-t_2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \gamma_X(t_1-t_N) & \gamma_X(t_2-t_N) & \gamma_X(t_3-t_N) & \cdots & \gamma_X(0) \end{pmatrix}.$$

[3]Note that, for some simple (complex-valued) Gaussian stationary signals, like, e.g., $X(t) = X \cdot e^{jt}$, where $X \sim N(0,1)$, one can choose the t_is so that the determinant of the covariance matrix is zero; take, for example $N = 2$, and $t_1 = \pi$, $t_2 = 2\pi$. Then the joint p.d.f. of the Gaussian random vector $(X(t_1),\ldots,X(t_N))^T$ is not of the form (9.3.2). Such signals are called degenerate.

For the real-valued signals under consideration it is always symmetric because, in that case, ACvF is an even function so that $\gamma_X(t_i - t_j) = \gamma_X(t_j - t_i)$. Also, it is obviously invariant under translations, that is, for any t,

$$\mathbf{\Gamma}_{(t_1,\ldots,t_N)} = \mathbf{\Gamma}_{(t_1+t,\ldots,t_N+t)}, \tag{9.3.4}$$

which, in view of (9.3.2) and (9.3.3), implies that all finite-dimensional p.d.f.s of $X(t)$ are also invariant under translations, that is, for any positive integer N, any sampling times $t_1, \ldots, t_N$, and any time shift t,

$$f_{(t_1,\ldots,t_N)}(x_1, \ldots, x_N) = f_{(t_1+t,\ldots,t_N+t)}(x_1, \ldots, x_N). \tag{9.3.5}$$

In other words:

A Gaussian weakly stationary signal is strictly stationary.

In the particular case when the sampling times are uniformly spaced with the intersampling time interval Δt, the covariance matrix $\mathbf{\Gamma}$ of the signal $X(t)$ sampled at times

$$t,\ t + \Delta t,\ t + 2\Delta t,\ \ldots,\ t + (N-1)\Delta t$$

is

$$\begin{pmatrix} \gamma_X(0) & \gamma_X(\Delta t) & \gamma_X(2\Delta t) & \cdots \gamma_X((N-1)\Delta t) \\ \gamma_X(\Delta t) & \gamma_X(0) & \gamma_X(\Delta t) & \cdots \gamma_X((N-2)\Delta t) \\ \cdots & \cdots & \cdots & \cdots \quad \cdots \\ \gamma_X((N-1)\Delta t) & \gamma_X((N-2)\Delta t) & \gamma_X((N-3)\Delta t) & \cdots \quad \gamma_X(0) \end{pmatrix}$$

Example 9.3.1 (Ornstein-Uhlenbeck Random Signal (Process)) Consider a Gaussian signal $X(t)$ with autocovariance function

$$\gamma_X(t) = e^{-0.3|t|}.$$

We are interested in finding the joint p.d.f. of the signal at times $t_1 = 1$, $t_2 = 2$, and the probability that the signal has values between -0.6 and 1.4 at t_1, and between 0.7 and 2.6 at t_2.

The first step is then to find the covariance matrix

$$\mathbf{\Gamma}_{(1,2)} = \begin{pmatrix} \gamma_X(0) & \gamma_X(1) \\ \gamma_X(1) & \gamma_X(0) \end{pmatrix} = \begin{pmatrix} e^0 & e^{-0.3} \\ e^{-0.3} & e^0 \end{pmatrix} = \begin{pmatrix} 1 & 0.74 \\ 0.74 & 1 \end{pmatrix}.$$

The covariance coefficient of $X(1)$ and $X(2)$ is then

$$\rho = \frac{\gamma_X(2-1)}{\gamma_X(0)} = 0.74$$

and, in view of Example 8.2.1 (8.2.9), the joint p.d.f. of $X(1)$ and $X(2)$ is of the form

$$f_{(1,2)}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-0.74^2}} \cdot \exp\left[\frac{-1}{2(1-0.74^2)}\left(x_1^2 - 2 \cdot 0.74 x_1 x_2 + x_2^2\right)\right]$$

$$= 0.24 \cdot \exp\left[-1.11\left(x_1^2 - 1.48 x_1 x_2 + x_2^2\right)\right].$$

Finally, the desired probability

$$\mathbf{P}\Big(-0.6 \le X(1) \le 1.4, \quad \text{and} \quad 0.7 \le X(2) \le 2.6\Big)$$

$$= \int_{-0.6}^{1.4} \int_{0.7}^{2.6} 0.24 \cdot e^{-1.11(x_1^2 - 1.48 x_1 x_2 + x_2^2)}\, dx_1\, dx_2 = 0.17,$$

where the last integral has been evaluated numerically in *Mathematica* with a two-digit precision.

9.4 Sample Path Properties of General and Gaussian Stationary Signals

Mean-Square Continuity and Differentiability It is clear that the local properties of the autocovariance function $\gamma_X(\tau)$ of a stationary signal $X(t)$ affect properties of the sample paths of the signal itself in the mean-square sense, that is in terms of the behavior of the expectation of the square of signal's increments, i.e., the variances of the increments.[4] Indeed, with no distributional assumptions on $X(t)$, we have

$$\sigma^2(\tau) = \mathbf{E}(X(t+\tau) - X(t))^2 = 2(\gamma_X(0) - \gamma_X(\tau));$$

the variance of the increment is independent of t. Hence, we have the following result:

[4]Recall that the sequence (X_n) of random quantities is said to converge to X, in the mean-square, if $\mathbf{E}|X_n - X|^2 \to 0$, as $n \to \infty$.

A stationary signal $X(t)$ is continuous in the mean-square sense, that is, for any $t > 0$,

$$\lim_{\tau \to 0} \mathbf{E}(X(t+\tau) - X(t))^2 = 0,$$

if, and only if, the autocovariance function $\gamma_X(\tau)$ is continuous at $\tau = 0$, that is,

$$\lim_{\tau \to 0} \gamma_X(\tau) = \gamma_X(0).$$

In particular, signals with autocovariance functions $\gamma_X(\tau) = e^{-|\tau|}$, or $\gamma_X(\tau) = 1/(1+\tau^2)$, are mean-square continuous.

A similar, mean-square analysis of the limit at $\tau = 0$ of the differential ratio,

$$\mathbf{E}\left(\frac{(X(t+\tau) - X(t))}{\tau}\right)^2 = 2\frac{\gamma_X(0) - \gamma_X(\tau)}{\tau^2}$$

shows that a stationary signal with autocovariance function $\gamma_X(\tau) = e^{-|\tau|}$ cannot be possibly mean-square differentiable because in this case

$$\lim_{\tau \to 0} \frac{\gamma_X(0) - \gamma_X(\tau)}{\tau^2} = \lim_{\tau \to 0} \frac{1 - e^{-|\tau|}}{\tau^2} = \infty,$$

whereas the differentiability cannot be excluded for the signal with autocovariance $\gamma_X(\tau) = 1/(1+\tau^2)$ because, in this case,

$$\lim_{\tau \to 0} \frac{\gamma_X(0) - \gamma_X(\tau)}{\tau^2} = \lim_{\tau \to 0} \frac{1 - 1/(1+\tau^2)}{\tau^2} = 1.$$

Of course, the above brief discussion just verifies the boundedness of the variance of the signal's differential ratio as $\tau \to 0$, not whether the latter has a limit. So, let us take a closer look at the issue of the mean-square differentiability of a stationary signal, that is the existence of the random quantity $X'(t)$, for a fixed t. First, observe that this existence is equivalent to the statement that[5]

$$\lim_{\tau_1 \to 0} \lim_{\tau_2 \to 0} \mathbf{E}\left(\frac{X(t+\tau_1) - X(t)}{\tau_1} - \frac{X(t+\tau_2) - X(t)}{\tau_2}\right)^2 = 0.$$

[5]This argument relies on the so-called Cauchy criterion of convergence for random quantities with finite variance: *A sequence X_n converges in the mean-square as $n \to \infty$, that is, there exists a random quantity X such that $\lim_{n\to\infty} \mathbf{E}(X_n - X)^2 = 0$, if and only if $\lim_{n\to\infty} \lim_{m\to\infty} \mathbf{E}(X_n - X_m)^2 = 0$.* This criterion permits the verification of the convergence without knowing what the limit is; see, e.g., Theorem 11.4.2 in W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York 1976.

But the expression under the limit signs is equal to

$$\mathbf{E}\left(\frac{X(t+\tau_1)-X(t)}{\tau_1}\right)^2+\mathbf{E}\left(\frac{X(t+\tau_2)-X(t)}{\tau_2}\right)^2$$

$$-2\mathbf{E}\left(\frac{X(t+\tau_1)-X(t)}{\tau_1}\cdot\frac{X(t+\tau_2)-X(t)}{\tau_2}\right).$$

So, the existence of the derivative $X'(t)$ in the mean-square is equivalent to the fact that the first two terms converge to $\gamma_{X'}(0)$ and the third to $-2\gamma_{X'}(0)$. But the convergence of the last term means the existence of the limit

$$\lim_{\tau_1\to 0}\lim_{\tau_2\to 0}\frac{1}{\tau_1\tau_2}\mathbf{E}\Big((X(t+\tau_1)-X(t))\cdot(X(t+\tau_2)-X(t))\Big)$$

$$=\lim_{\tau_1\to 0}\lim_{\tau_2\to 0}\frac{1}{\tau_1\tau_2}\Big(\gamma_X(\tau_2-\tau_1)-\gamma_X(\tau_1)-\gamma_X(\tau_2)+\gamma_X(0)\Big)$$

$$=\lim_{\tau_1\to 0}\lim_{\tau_2\to 0}\frac{1}{\tau_1\tau_2}\Delta_{-\tau_1}\Delta_{\tau_2}\gamma_X(0),$$

where $\Delta_\tau f(t):=f(t+\tau)-f(t)$ is the usual difference operator. Indeed,

$$\Delta_{-\tau_1}\Delta_{\tau_2}\gamma_X(0)=\Delta_{-\tau_1}(\gamma_X(\tau_2)-\gamma_X(0))$$

$$=(\gamma_X(\tau_2-\tau_1)-\gamma_X(-\tau_1))-(\gamma_X(\tau_2)-\gamma_X(0)).$$

Since the existence of the last limit appearing above means twice differentiability of the autocovariance function of X at $\tau=0$ we arrive at the following criterion:

A stationary signal $X(t)$ is mean-square differentiable if and only if its autocovariance function $\gamma_X(\tau)$ is twice differentiable at $\tau=0$. Moreover, in this case, the crosscovariance of the signal $X(t)$ and its derivative $X'(t)$

$$\mathbf{E}X(t)X'(s)=\lim_{\tau\to 0}\frac{\gamma_X(t+\tau-s)-\gamma_X(t-s)}{\tau}=\frac{\partial}{\partial t}\gamma_X(t-s), \tag{9.4.1}$$

and the autocovariance of the derivative signal

$$\mathbf{E}X'(t)X'(s)=\lim_{\tau\to 0}\frac{1}{\tau}\left(\frac{\partial}{\partial t}\gamma_X(t+\tau-s)-\frac{\partial}{\partial t}\gamma_X(t-s)\right)=\frac{\partial^2}{\partial t\,\partial s}\gamma_X(t-s). \tag{9.4.2}$$

In a similar fashion one can calculate the crosscovariance of higher derivatives of the signal $X(t)$ to obtain that[6]

$$\mathbf{E}X^{(n)}(t)X^{(m)}(s) == \frac{\partial^{n+m}}{\partial t^n \partial s^m}\gamma_X(t-s). \tag{9.4.3}$$

Sample Path Continuity A study of properties of the individual sample paths (trajectories, realizations) of stationary random signals is a more delicate matter, with the most precise results obtainable only in the case of Gaussian signals. Indeed, we have observed in the previous sections that, for a Gaussian signal, the autocovariance function determines all the finite-dimensional probability distributions of the signal, meaning that for any finite sequence of windows, $[a_1, b_1], [a_2, b_2], \ldots, [a_N, b_N]$, and any collections of time instants $t_1, t_2, \ldots, t_N$, we can find the probability that the signal fits into those windows at prescribed times, that is,

$$\mathbf{P}(a_1 < t_1 < b_1,\ a_2 < t_2 < b_2, \ldots,\ a_N < t_N < b_N).$$

So it seems that by taking N to ∞, and making the time instants closer to each other, and windows narrower, one could find the probability that the signal's sample path has any specific shape or property. This idea is roughly speaking correct but only in a subtle sense that will be explained below.

The discussion of the sample path properties of stationary signals will be based here on the following theorem of the theory of general random signals (stochastic processes) due to N.N. Kolmogorov:

Theorem 9.4.1 *Let $g(h)$ be an even function, nondecreasing for $h > 0$, and such that $g(h) \to 0$ as $h \to 0$. Furthermore, suppose that $X(t)$ is a random signal such that*

$$\mathbf{P}\Big(|X(t+h) - X(t)| > g(h)\Big) \le q(h), \tag{9.4.4}$$

for a function $q(h)$ satisfying the following three conditions:

$$q(h) \to 0, \qquad \text{as} \qquad h \to 0; \tag{9.4.5}$$

$$\sum_{n=1}^{\infty} 2^n q(2^{-n}) < \infty; \tag{9.4.6}$$

$$\sum_{n=1}^{\infty} g(2^{-n}) < \infty. \tag{9.4.7}$$

Then, with probability 1, the sample paths of the signal $X(t)$ are continuous.

[6]For details, see M. Loeve, *Probability Theory*, Van Nostrand, Princeton 1963, Section 34.3.

Although the proof of the above theorem is beyond the scope of this book[7] the intuitive meaning of the assumptions (9.4.4)–(9.4.7) is clear: for the signal to have continuous sample paths the increments of the signal over small time intervals can be permitted to be large only with a very small probability.

Applied to the second order (not necessarily stationary) signals Theorem 9.4.1 immediately gives the following.

Corollary 9.4.1 *If there exists a τ_0 such that, for all τ, $0 \le \tau < \tau_0$, and all t in a finite time interval,*

$$\mathbf{E}\Big(X(t+\tau) - X(t)\Big)^2 \le C|\tau|^{1+\epsilon} \tag{9.4.8}$$

for some constants $C, \epsilon > 0$, then the sample paths of the signal $X(t)$ are continuous with probability 1.

To see how Corollary 9.4.1 follows from Theorem 9.4.1 observe first that, for any random quantity Z, and any constant $a > 0$.[8]

$$\mathbf{P}(Z > a) \le \int_a^\infty f_Z(z)\,dz \le \int_a^\infty \frac{z^2}{a^2} f_Z(z)\,dz \le \frac{\mathbf{E}Z^2}{a^2}.$$

Condition (9.4.8) implies then that

$$\mathbf{P}\Big(|X(t+\tau) - X(t)| > g(\tau)\Big) \le \frac{C|\tau|^{1+\epsilon}}{g^2(\tau)},$$

so that selecting $g(\tau) = |\tau|^{\epsilon/4}$, and

$$q(\tau) = \frac{C|\tau|^{1+\epsilon}}{g^2(\tau)} = C|\tau|^{1+\epsilon/2},$$

we easily see that $g(\tau)$ and $q(\tau)$ are continuous functions vanishing at $\tau = 0$, and that the conditions (9.4.6) and (9.4.7) of the theorem are also satisfied. Indeed,

$$\sum_{n=1}^{\infty} 2^n q(2^{-n}) = C\sum_{n=1}^{\infty} 2^n (2^{-n})^{1+\epsilon/2} = C\sum_{n=1}^{\infty} 2^{-n\epsilon/2} < \infty,$$

[7] For a more complete discussion of this theorem and its consequences for sample path continuity and differentiability of random signals, see, for example, M. Loève, *Probability Theory*, Van Nostrand, Princeton 1963, Section 35.3.

[8] This inequality is known as the Chebyshev Inequality and its proof here has been carried out only in the case of absolutely continuous probability distributions. The proof in the discrete case is left to the reader as an exercise, see Sect. 9.5.

and

$$\sum_{n=1}^{\infty} g(2^{-n}) = \sum_{n=1}^{\infty} 2^{-n\epsilon/4} < \infty.$$

In the special case of a stationary signal we have $\mathbf{E}(X(t+\tau) - X(t))^2 = 2(\gamma_X(0) - \gamma_X(\tau))$, so the sample path continuity is guaranteed by the following condition on the autocovariance function:

$$|\gamma_X(0) - \gamma_X(\tau)| \leq C|\tau|^{1+\epsilon}, \tag{9.4.9}$$

for some constant $\epsilon > O$, and small enough τ.

In particular, for the autocovariance function $\gamma_X(\tau) = 1/(1+\tau^2)$,

$$|\gamma_X(0) - \gamma_X(\tau)| = 1 - \frac{1}{1+\tau^2} = \frac{\tau^2}{1+\tau^2} \leq \tau^2,$$

and the condition (9.4.8) is satisfied thus giving the sample path continuity.

However, for a signal with autocovariance function $\gamma_X(\tau) = e^{-|\tau|}$, the difference $\gamma_X(0) - \gamma_X(\tau)$ behaves asymptotically like τ, for $\tau \to 0$. Therefore, there is no positive ϵ for which condition (9.4.9) is satisfied and we cannot claim the continuity of the sample path in this case—not a surprising result if one remembers that the exponential autocovariance was first encountered in the context of the obviously sample path discontinuous switching signal. Nevertheless, as we observed at the beginning of this section a signal with an exponential autocovariance is mean-square continuous.

For a Gaussian stationary signal $X(t)$, Theorem 9.4.1 can be applied in a more precise fashion since the probabilities $\mathbf{P}(X(t+\tau) - X(t) > a)$ are known exactly. Indeed, since for any positive z,

$$\int_z^{\infty} e^{-x^2/2}\, dx \leq \int_z^{\infty} \frac{x}{z}\, e^{-x^2/2}\, dx = \frac{1}{z} e^{-z^2/2},$$

because $x/z \geq 1$ in the interval of integration, we have, for any nonnegative function $g(\tau)$, and positive constant C,

$$\mathbf{P}\Big(|X(t+\tau) - X(t)| > Cg(\tau) \Big) \leq \sqrt{\frac{2}{\pi}} \frac{\sigma(\tau)}{Cg(\tau)} \exp\left(-\frac{1}{2} \frac{C^2 g^2(\tau)}{\sigma^2(\tau)} \right), \tag{9.4.10}$$

where $\sigma^2(\tau) = \mathbf{E}(X(t+\tau) - X(t))^2 = 2(\gamma_X(0) - \gamma_X(\tau))$. This estimate yields the following result:

Corollary 9.4.2 *If there exists τ_0 such that, for all τ, $0 \leq \tau \leq \tau_0$, the autocovariance function $\gamma_X(\tau)$ of a stationary Gaussian signal $X(t)$ satisfies condition*

$$\gamma_X(0) - \gamma_X(\tau) \leq \frac{K}{|\ln|\tau||^{\delta}}, \tag{9.4.11}$$

for some constants $K > 0$ and $\delta > 3$, then the signal $X(t)$ has continuous sample paths with probability 1.

The proof of the Corollary is completed by selecting

$$g(\tau) = |\ln|\tau||^{-\nu},$$

with any number ν satisfying condition $1 < \nu < (\delta - 1)/2$, choosing

$$q(C, \tau) = \frac{K'}{C|\ln|\tau||^{\delta/2-\nu}} \exp\left(-\frac{C^2}{2K}|\ln|\tau||^{\delta-2\nu}\right),$$

and verifying the convergence of the two series in conditions (9.4.6) and (9.4.7); see an exercise in Sect. 9.5.

Returning to the case of a stationary random signal with an exponential autocovariance function we see that if the signal is Gaussian, then Corollary 9.4.2 guarantees the continuity of its sample paths with probability 1. Indeed, condition (9.4.11) is obviously satisfied since (e.g., picking $\delta = 4$) we have

$$\lim_{\tau\to 0}(\gamma_X(0) - \gamma_X(\tau)) \cdot |\ln|\tau||^4 = \lim_{\tau\to 0}(1 - e^{-|\tau|}) \cdot |\ln|\tau||^4 = 0$$

in view of de l'Hospital's rule.

9.5 Problems and Exercises

1* A zero-mean Gaussian random signal has the autocovariance function of the form

$$\gamma_X(\tau) = e^{-0.1|\tau|}\cos 2\pi\tau.$$

Plot it. Find the power spectrum $S_X(f)$. Write the covariance matrix for the signal sampled at four time instants separated by 0.5 s. Find its inverse (numerically; use any of the familiar computing platforms, such as *Mathematica* and *Matlab*).

2 Find the joint p.d.f. of the signal from Exercise 9.5.1 at $t_1 = 1$ and $t_2 = 2$. Write the integral formula for

$$P(0 \leq X(1) \leq 1, 0 \leq X(2) \leq 2).$$

Evaluate the above probability numerically.

3* Find the joint p.d.f. of the signal from Problem 1 at $t_1 = 1, t_2 = 1.5, t_3 = 2$, and $t_4 = 2.5$. Write the integral formula for

$$P(-2 \le X(1) \le 2, -1 \le X(1.5) \le 4, -1 \le X(2) \le 1, 0 \le X(2.5) \le 3).$$

Evaluate the above probability numerically.

4* Show that if a 2D Gaussian random vector $\vec{Y} = (Y_1, Y_2)$ has uncorrelated components Y_1, Y_2, then those components are statistically independent random quantities.

5 Produce 3D surface plots for p.d.f.s of three 2D Gaussian random vectors: $(X(1.0), X(1.1))^T$, $(X(1.0), X(2.0))^T$, and $(X(1.0), X(5.0))^T$, where $X(t)$ is the stationary signal described in Example 9.3.1. Comment on the similarities and differences in the three plots.

6 Prove that if there exists a τ_0 such that, for all $\tau < \tau_0$, and all t in a finite time interval,

$$\mathbf{E}\Big(X(t+\tau) - X(t)\Big)^2 \le C \frac{|\tau|}{|\ln|\tau||^{1+\delta}},$$

for some $C > 0$ and $\delta > 2$, then the sample paths of the signal $X(t)$ are continuous with probability 1. *Hint:* This result is a little more delicate than Corollary 9.4.1 but the idea of the proof is similar: take $g(\tau) = |\ln|\tau||^{-\beta}$, for a β between 1 and $\delta/2$, wherefrom $q(\tau) = |\tau|/|\ln|\tau||^{1+\delta-2\beta}$, and check conditions (9.4.4)–(9.4.7) in Theorem 9.4.1.

7 Verify the Chebyshev inequality, $\mathbf{P}(|Z| > a) \le \mathbf{E}Z^2/a^2$, $a > 0$, for a discrete random quantity Z.

8 Produce 3D plots of several 2D Gaussian densities with selected means and covariance matrices. Then plot level curves for them.

9 Random signal $X(t)$ has an autocovariance function of the form $\gamma_X(\tau) = \exp(-|\tau|^\alpha)$ with $0 < \alpha \le 2$. For which values of parameter α can you claim the continuity of sample paths of $X(t)$ with probability 1? For $\alpha > 2$, the above formula does not give a covariance function of any stationary signal. Why? *Hint:* Check positive-definiteness condition from Remark 6.2.1.

10 Verify formula (9.4.3) for the crosscovariance of higher derivatives of a stationary signal.

11 Verify the convergence of the series (9.4.6) and (9.4.7) in the proof of Corollary 9.4.2.

Chapter 10
Spectral Representation of Discrete-Time Stationary Signals and Their Computer Simulations

Abstract This chapter provides a spectral representation of stationary, discrete-time random signals and determines that its autocovariance function is a positive-definite sequence cumulative power spectrum is also introduced. Stochastic integrals with respect to signals with uncorrelated increments are developed permitting introduction of computer algorithms to simulate stationary signals with a given spectral density.

10.1 Spectral Representation

Given an arbitrary power spectrum $S_X(f)$ or, equivalently, its Inverse Fourier Transform, the autocovariance function, $\gamma_X(\tau)$, our ability to simulate the corresponding stationary random signals $X(t)$, using only the pseudo-random number generator which produces, say, discrete-time white noise, depends on the observation that, in some sense, all stationary random signals can be approximated by superpositions of random harmonic oscillations such as those discussed in Examples 5.1.2 and 5.1.9. Recall that if $A_1, \ldots, A_N$, are independent, zero mean random variables with finite variance, and $0 < f_1, \ldots, f_N \leq 1$, is a sequence of distinct frequencies, then a random superposition of N simple complex-valued harmonic oscillations in discrete time, $n = \ldots, -1, 0, 1, \ldots,$

$$X_N(n) = \sum_{k=1}^{N} A_k \cdot e^{j2\pi f_k n}, \qquad 0 < f_k \leq 1, \tag{10.1.1}$$

is a stationary signal with the autocovariance function of the form

$$\gamma_{X_N}(n) = \sum_{k=1}^{N} \mathbf{E}|A_k|^2 \cdot e^{j2\pi f_k n}. \tag{10.1.2}$$

This suggests the following, intuitive approach to our simulation problem: Given a power spectrum, $S_X(f)$, concentrated, say, on the frequency interval

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_10

[0, 1], mimicking the continuous-time analysis of Sect. 6.2, we can expect the corresponding ACvF to be the "discrete-time inverse Fourier transform," i.e., the Fourier coefficients of $S_X(f)$,

$$\gamma_X(n) = \int_0^1 S(f) e^{j2\pi f n}\, df.$$

The latter integral can now be approximated by its discretized version, so that

$$\gamma_X(n) \approx \sum_{k=1}^{N} S(f_k)\ \Delta f_k\, e^{j2\pi f_k n}, \tag{10.1.3}$$

where

$$0 = f_0 < f_1 < \ldots < f_N = 1$$

is a partition of the [0, 1] interval and $\Delta f_k = f_k - f_{k-1}$. Comparing (10.1.2) and (10.1.3), it seems that to produce an approximated version of $X(n)$, it now suffices to generate a standard white noise $W(k), k = 1, \ldots N$, take as the random amplitudes in (10.1.1) the sequence

$$A_k = \sqrt{S(f_k)\Delta f_k}\, W(k), \qquad k = 1, \ldots, N, \tag{10.1.4}$$

so that $\mathbf{E}|A_k|^2 = S(f_k)\Delta f_k$, and produce the sequence,

$$X_N(n) = \sum_{k=1}^{N} \sqrt{S(f_k)\Delta f_k}\, W(k) \cdot e^{j2\pi f_k n} \approx X(n). \tag{10.1.5}$$

Alternatively, we can consider the Fourier series expansion of the power spectrum (see Chap. 2, but here the variable is the frequency f),

$$S_X(f) = \sum_{n=-\infty}^{\infty} c_n \cdot e^{j2\pi f n}, \tag{10.1.6}$$

with the Fourier coefficients

$$c_n = \int_0^1 S_X(f) \cdot e^{-j2\pi f n}\, df. \tag{10.1.7}$$

Now, the above integral can be replaced, approximately, by the discretized sum,

$$c_n \approx \sum_{k=1}^{K} a_k \cdot e^{-j2\pi f_k n}, \tag{10.1.8}$$

with the Fourier coefficients

$$a_k = \int_{f_{k-1}}^{f_k} S_X(f)\, df, \qquad k = 1, 2, \ldots, K \tag{10.1.9}$$

corresponding to the power of the signal $X(n)$ concentrated in each of the frequency bands $[f_{k-1}, f_k]$, $k = 1, \ldots, K$. Finally, we recognize in (10.1.8) the discrete-time version of the ACvF of the form (10.1.2) of the signal of the form (10.1.1), which gives us yet another approximate expression for the sought signal $X(n)$:

$$X_K(n) \approx \sum_{k=1}^{K} A_k \cdot e^{-j2\pi f_k n}, \tag{10.1.10}$$

where A_k are selected to be arbitrary zero-mean, independent random variables, with $\mathbf{E}|A_k|^2 = a_k$, $k = 1, \ldots, K$, so that

$$\gamma_{X_K}(n) = \sum_{k=1}^{K} a_k \cdot e^{-j2\pi f_k n} \approx c_n. \tag{10.1.11}$$

If $W(k)$ is the standard white noise (of an arbitrary distribution), then choosing

$$A_k = \sqrt{a_k} \cdot W(k), \qquad k = 1, \ldots, K \tag{10.1.12}$$

will also do the job.

Obviously, the key to applying the above schemes is in the details: In what sense the approximations is meant? What are the precise algorithms? What is the rigorous justification for them? Also, clearly, for smooth spectra, $S(f)$, and large K, and N, the difference between the expressions, (10.1.5) and (10.1.10), is negligible.

In this chapter we work with discrete-time signals and the rigorous answer to the above questions is contained in the so-called *Spectral Representation Theorem* for stationary random signals which is derived in this chapter. On the way to its formulation we introduce the necessary concepts including the crucial construction of *stochastic integrals* with respect to a white noise signal, often called the *white-noise integrals*. We conclude with a computer algorithm based on the *Spectral Representation Theorem*.

10.2 Autocovariance as a Positive-Definite Sequence

In this chapter we will study random stationary signals in discrete time, that is sequences of *complex-valued* random quantities

$$\ldots, X(-2), X(-1), X(0), X(1), X(2), \ldots$$

with time n extending all the way from minus to plus infinity. The stationarity is meant in the second-order, weak sense, that is we will assume that the means $\mathbf{E}X(n) = 0$ and the autocovariance function, now, really a sequence,

$$\mathbf{E}[X^*(m)X(n)] = \gamma(n-m), \qquad m, n = \ldots, -2, -1, 0, 1, 2, \ldots,$$

depends only on the time-lag $\tau = n - m$. The following properties of the *autocovariance sequence* are immediately verified:

(i) For any n,

$$\mathbf{E}|X(n)|^2 = \mathbf{E}[X^*(n)X(n)] = \mathbf{E}|X(0)|^2 = \gamma_X(0) \geq 0, \tag{10.2.1}$$

$$\gamma_X(-n) = \gamma_X^*(n), \tag{10.2.2}$$

$$|\gamma_X(n)| \leq \gamma_X(0). \tag{10.2.3}$$

The last inequality is a direct consequence of the Cauchy-Schwartz inequality.

Also, importantly, the autocovariance sequence is *positive definite*, that is, for any positive integer N, arbitrary integers, $n_1, n_2, \ldots, n_N$, and arbitrary complex numbers $\lambda_1, \lambda_2, \ldots, \lambda_N$,

$$\sum_{i,k=1}^{N} \gamma_X(n_i - n_k)\lambda_i\lambda_k^* \geq 0. \tag{10.2.4}$$

Indeed,

$$\sum_{i,k=1}^{N} \gamma_X(n_i - n_k)\lambda_i\lambda_k^* = \sum_{i,k=1}^{N} \mathbf{E}[X(n_i)X^*(n_k)]\lambda_i\lambda_k^*$$

$$= \mathbf{E}\sum_{i,k=1}^{N} [\lambda_i X(n_i)] \cdot [\lambda_k X(n_k)]^* = \mathbf{E}\sum_{i=1}^{N} \lambda_i X(n_i) \cdot \sum_{k=1}^{N} [\lambda_k X(n_k)]^*$$

$$= \mathbf{E}\left|\sum_{i=1}^{N} \lambda_i X(n_i)\right|^2 \geq 0.$$

Recall, see Remark 6.2.1, that the ACvF in continuous time was also proven to be positive-definite.

10.3 Cumulative Power Spectrum of Discrete-Time Stationary Signal

The development of this section will be analogous to the development of the concept of power spectrum of continuous-time signals in Sect. 5.2. However, we will proceed in a slightly different fashion, and with more mathematical precision. The basic structural result regarding the autocovariance function of a discrete-time stationary signal can be formulated as follows:

Herglotz Theorem *The following statements about sequence* $\gamma(n), n = \ldots, -2, -1, 0, 1, 2, \ldots,$ *of complex numbers are equivalent:*

(i) Sequence $\gamma(n)$ *is an autocovariance sequence of a stationary discrete-time signal, that is, there exists a stationary signal* $X(n)$ *such that* $\gamma(n) = \gamma_X(n)$*;*
(ii) Sequence $\gamma(n)$ *is positive-definite, that is, it satisfies condition (10.2.4);*
(iii) There exists a nondecreasing bounded function $\mathcal{S}_X(f)$*, defined on the interval* $[0, 1]$*, such that*

$$\gamma(n) = \int_0^1 e^{j2\pi nf}\, d\mathcal{S}(f), \qquad n = \ldots, -2, -1, 0, 1, 2, \ldots. \tag{10.3.1}$$

Function $\mathcal{S}_X(f)$ *is called cumulative power spectrum of signal* X*.*

Remark 10.3.1 (Power Spectrum Density) The integral of the form $\int a(f)\, d\mathcal{S}(f)$, called the Stieltjes integral, is to be understood as the limit of sums $\sum a(f_i) \cdot \Delta\mathcal{S}(f_i)$, when $\max_i |\Delta\mathcal{S}(f_i)| = \mathcal{S}(f_i) - \mathcal{S}(f_{i-1}) \to 0$. As before, $0 = f_0 < f_1 < \ldots < f_N = 1$, stands for a partition of the interval $[0, 1]$.

If the cumulative power spectrum has a spectral density $S(f), 0 \le f \le 1$, that is,

$$\mathcal{S}(f) = \int_0^f S(g)\, dg, \qquad \frac{d\mathcal{S}(f)}{df} = S(f) \ge 0,$$

then formula (10.2.1) takes the form of the usual Riemann integral

$$\gamma(n) = \int_0^1 e^{j2\pi nf}\, S(f)\, df, \qquad n = \ldots, -2, -1, 0, 1, 2, \ldots, \tag{10.3.2}$$

and the sequence $\gamma(-n)$ can be simply viewed as the sequence of Fourier coefficients of power spectrum density $S(f)$.

In the special case when the cumulative power spectrum is constant, except for jumps, that is

$$\mathcal{S}(f) = \sum_k s_k\, u(f - f_k), \qquad 0 = f_0 < f_1 < \ldots < f_N = 1,$$

where $u(t)$ is the unit step function, then

$$\int a(f)\, d\mathcal{S}(f) = \sum_l a(f_k) s_k,$$

so that

$$\gamma(n) = \sum_k s_k e^{j2\pi n f_k}, \qquad n = \ldots, -2, -1, 0, 1, 2, \ldots. \tag{10.3.3}$$

and the power spectrum density can be understood as a sum of the Dirac-deltas:

$$S(f) = \sum_k s_k\, \delta(f - f_k).$$

However, it is worth remembering that there are so-called singular cumulative power spectra that are not of either of the two types described above (nor their mixtures).[1]

Proof of Herglotz Theorem The implication *(i)* $\Longrightarrow$ *(ii)* has been proved following the definition (10.2.4).

We shall now prove that *(ii)* $\Longrightarrow$ *(iii)*. So, assume that $\gamma(n)$ is positive definite. In view of (10.2.4), selecting $n_i = i$, $\lambda_i = e^{-j2\pi i f}$, $i = 1, 2, \ldots, N$, we have

$$0 \le \sum_{i,k=1}^{N} \gamma(i-k) e^{-j2\pi i f} e^{j2\pi k f} = \sum_{i,k=1}^{N} \gamma(i-k) e^{-j2\pi (i-k) f}$$

$$= \sum_{m=-N+1}^{N-1} (N - |m|)\gamma(m) e^{-j2\pi m f},$$

after substitution $m = i - k$. Define,

$$S_N(f) := \frac{1}{N} \sum_{m=-N+1}^{N-1} (N - |m|)\gamma(m) e^{-j2\pi m f}.$$

Then,

$$S_N(f) \ge 0, \qquad \text{and} \qquad \int_0^1 S_N(f)\, df = \gamma(0). \tag{10.3.4}$$

[1] See Sect. 4.1 or, e.g., M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena. Uncertainty, Complexity and Chaotic Behavior in Engineering and Science,* Birkhäuser-Boston 1998.

By a fundamental real analysis result called Banach-Alaoglu Theorem,[2] conditions (10.3.4) guarantee the existence of a function $\mathcal{S}(f)$ and a sequence $N_i \nearrow \infty$, $i \to \infty$, such that, for each bounded and smooth function $a(f)$,

$$\int_0^1 a(f) S_{N_i}(f)\, df \longrightarrow \int_0^1 a(f)\, d\mathcal{S}(f).$$

Therefore, selecting $a(f) = e^{j2\pi mf}$, we have

$$\int_0^1 e^{j2\pi mf}\, d\mathcal{S}(f) = \lim_{i\to\infty} \int_0^1 e^{j2\pi mf} S_{N_i}(f)\, df = \gamma(m)$$

because, for each m such that $|m| \le N_i$,

$$\int_0^1 e^{j2\pi mf} S_{N_i}(f)\, df = \gamma(m)\left(1 - \frac{|m|}{N_i}\right).$$

Thus the existence of the cumulative spectral measure for each discrete-time stationary signal has been established.

The implication *(iii)* $\Longrightarrow$ *(ii)* can be verified directly. Indeed, given assumption (iii),

$$\sum_{i,k=1}^{N} \gamma(n_i - n_k)\lambda_i \lambda_k^* = \sum_{i,k=1}^{N} \int_0^1 e^{j2\pi(n_i - n_k)f} d\mathcal{S}(f) \cdot \lambda_i \lambda_k^*$$

$$= \int_0^1 \sum_{i,k=1}^{N} [\lambda_i e^{j2\pi n_i f}] \cdot [\lambda_k e^{j2\pi n_k f}]^* d\mathcal{S}(f) = \int_0^1 \left| \sum_{i=1}^{N} \lambda_i e^{j2\pi n_i f} \right|^2 d\mathcal{S}(f) \ge 0,$$

because $\mathcal{S}(f)$ is nondecreasing, so that its increments, "$d\mathcal{S}(f)$," are nonnegative.

The implication *(ii)* $\Longrightarrow$ *(i)* follows from the following fact established in Sect. 10.2. For any given positive-definite matrix $\mathbf{\Gamma} = (\gamma_{ik})$, $i, k = 1, 2, \ldots, N$, there exists a Gaussian random vector $\mathbf{X} = (X_1, X_2, \ldots, X_N)$, with covariance matrix $\mathbf{\Gamma}$. Now, for any N, it suffices to take $\mathbf{\Gamma} = (\gamma(i-k))\,, i, k = 1, 2, \ldots, N$, and define

$$X(1) = X_1,\ X(2) = X_2,\ \ldots,\ X(N) = X_N.$$

[2] See, e.g., G.B. Folland, *Real Analysis*, J. Wiley, New York 1984.

This proves the existence of a finite discrete-time stationary random signal with an autocovariance sequence given by a prescribed positive-definite sequence.[3]

10.4 Stochastic Integration with Respect to Signals with Uncorrelated Increments

Recall that our goal in this chapter is to develop a simulation algorithm for discrete-time stationary signals with a given power spectrum, and one of the methods used for that purpose involves representation of the random signal as a stochastic integral with respect to another random signal which has uncorrelated increments which is easy to simulate via a pseudo-random number generator. The purpose of this section is to introduce such integrals.

The finite variance, zero-mean, real-valued signal $\mathcal{W}(w)$ of continuous, or discrete parameter w is said to have *uncorrelated increments* if, for any $w_1 \le w_2 \le w_3$,

$$\mathbf{E}\Big[(\mathcal{W}(w_3) - \mathcal{W}(w_2)) \cdot (\mathcal{W}(w_2) - \mathcal{W}(w_1))\Big] = 0. \tag{10.4.1}$$

In other words, such signals have uncorrelated increments over disjoint intervals of parameter w. Observe that condition (10.4.1) can be rewritten in terms of the autocovariance function $\gamma_{\mathcal{W}}(v, w) = \mathbf{E}\mathcal{W}(v)\mathcal{W}(w)$ (which here is truly a function of two variables v, w, and not just the parameter lag $w-v$ as is the case for stationary signals) as follows:

$$\mathbf{E}\Big[(\mathcal{W}(w_3) - \mathcal{W}(w_2)) \cdot (\mathcal{W}(w_2) - \mathcal{W}(w_1))\Big] = \tag{10.4.2}$$

$$\mathbf{E}\mathcal{W}(w_3)\mathcal{W}(w_2) - \mathbf{E}\mathcal{W}(w_2)\mathcal{W}(w_2) - \mathbf{E}\mathcal{W}(w_3)\mathcal{W}(w_1) + \mathbf{E}\mathcal{W}(w_2)\mathcal{W}(w_1)$$

$$= \gamma_{\mathcal{W}}(w_3, w_2) - \gamma_{\mathcal{W}}(w_2, w_2) - \gamma_{\mathcal{W}}(w_3, w_1) + \gamma_{\mathcal{W}}(w_2, w_1) = 0.$$

Example 10.4.1 (Random Walk: The Cumulative White Noise in Discrete Time) In discrete time, the white noise, $W(n)$, was defined simply as a sequence of zero-mean, independent (and thus uncorrelated), identically distributed random quantities with finite variance, so that its autocovariance sequence,

$$\gamma_W(n, m) = \gamma_W(m - n) = \mathbf{E}W(n)W(m) = \begin{cases} 0, & \text{if } n - m \neq 0; \\ \sigma^2, & \text{if } n - m = 0. \end{cases}$$

[3] A step proving the existence of an *infinite* such sequence requires an application of the so-called Kolmogorov Extension Theorem, see, e.g., P. Billingsley, *Probability and Measure*, Wiley, New York, 1986.

We will define the *random walk*, or *cumulative white noise*, generated by the white noise $W(n)$ as the random signal,

$$\mathcal{W}(n) = W(1) + W(2) + \ldots + W(n), \qquad n = 1, 2, \ldots,$$

with the convention $\mathcal{W}(0) = 0$.

The following mental picture is worth keeping in mind: In the case of the symmetric Bernoulli white noise, $W(n)$, with $\mathbf{P}(W(n) = \pm 1) = 1/2$, the generated random walk $\mathcal{W}(n)$ moves "forward" by 1, whenever $W(n) = +1$, and "backward" by 1, whenever $W(n) = -1$; each possibility occurring with probability 1/2.

The cumulative white noise has uncorrelated increments. Indeed, if $n_1 \le n_2 \le n_3$, then

$$\mathbf{E}\Big[(\mathcal{W}(n_3) - \mathcal{W}(n_2)) \cdot (\mathcal{W}(n_2) - \mathcal{W}(n_1))\Big]$$

$$= \mathbf{E}\Big[(\sum_{n=1}^{n_3} W(n) - \sum_{n=1}^{n_2} W(n)) \cdot (\sum_{n=1}^{n_2} W(n) - \sum_{n=1}^{n_1} W(n))\Big]$$

$$= \mathbf{E}\Big[\Big(W(n_2+1) + \ldots + W(n_3)\Big) \cdot \Big(W(n_1+1) + \ldots + W(n_2)\Big)\Big]$$

$$= \mathbf{E}\Big(W(n_1+1) + \ldots + W(n_2)\Big) \cdot \mathbf{E}\Big(W(n_2+1) + \ldots + W(n_3)\Big) = 0,$$

because $W(n_1+1) + \ldots + W(n_2)$ and $W(n_2+1) + \ldots + W(n_3)$ are independent and zero-mean.

For any signal $\mathcal{W}(w)$ with uncorrelated increments we will introduce a *cumulative control function*

$$\mathcal{C}(w) := \mathbf{E}[\mathcal{W}(w) - \mathcal{W}(0)]^2 = \mathbf{E}[\mathcal{W}(w)]^2 \ge 0, \tag{10.4.3}$$

which simply measures the variance of the increment of the signal from 0 to w. Since the variance of the sum of uncorrelated random quantities is the sum of their variances, the cumulative control function is always nondecreasing because, for $0 \le v \le w$,

$$\mathcal{C}(w) = \mathbf{E}[(\mathcal{W}(w) - \mathcal{W}(0)]^2 = \mathbf{E}[(\mathcal{W}(w) - \mathcal{W}(v)) + (\mathcal{W}(v) - W(0))]^2$$

$$= \mathbf{E}[(\mathcal{W}(w) - \mathcal{W}(v))]^2 + \mathbf{E}[(\mathcal{W}(v) - W(0))]^2 \ge \mathbf{E}[(\mathcal{W}(v) - W(0))]^2 = \mathcal{C}(v). \tag{10.4.4}$$

Observe that, under condition $\mathcal{W}(0) = 0$, the cumulative control function determines the correlation structure of $\mathcal{W}(w)$ and vice versa. If, say, $0 \le v \le w$, then

$$\gamma_{\mathcal{W}}(v, w) = \mathbf{E}\mathcal{W}(v)\mathcal{W}(w)$$

$$= \mathbf{E}[\mathcal{W}(v) - \mathcal{W}(0)] \cdot [(\mathcal{W}(w) - \mathcal{W}(v)) + (\mathcal{W}(v) - \mathcal{W}(0)] =$$

$$= \mathbf{E}[\mathcal{W}(v) - \mathcal{W}(0)] \cdot [(\mathcal{W}(v) - \mathcal{W}(0)] = \mathcal{C}(v),$$

because the increments over intervals $[0, v]$ and $[v, w]$ are uncorrelated. Since an analogous reasoning holds true in the case $0 \leq w \leq v$, we get the general formula

$$\gamma_{\mathcal{W}}(v, w) = \mathcal{C}(\min(v, w)). \tag{10.4.5}$$

An important class of signals with independent (and thus uncorrelated) increments are those that also have *stationary increments*, that is for which the c.d.f. of the increment $\mathcal{W}(w) - \mathcal{W}(v)$ is the same as the c.d.f. of the increment $\mathcal{W}(w + z) - \mathcal{W}(v + z)$, for any z. Random walk from Example 10.4.1 is such a signal. For signals with independent and stationary increments the cumulative control function satisfies condition

$$\mathcal{C}(w + v) = \mathcal{C}(w) + \mathcal{C}(v) \tag{10.4.6}$$

because

$$\mathbf{E}[\mathcal{W}(w + v) - \mathcal{W}(0)]^2 = \mathbf{E}[\mathcal{W}(w + v) - \mathcal{W}(v)]^2 + \mathbf{E}[\mathcal{W}(v) - \mathcal{W}(0)]^2$$

$$= \mathbf{E}[\mathcal{W}(w) - \mathcal{W}(0)]^2 + \mathbf{E}[\mathcal{W}(v) - \mathcal{W}(0)]^2.$$

Condition (10.4.6) forces the cumulative function to be linear, that is, of the form

$$\mathcal{C}_{\mathcal{W}}(w) = const \cdot w, \tag{10.4.7}$$

and, in view of (10.4.5), the autocovariance structure of a signal with stationary and uncorrelated increments is of the form

$$\gamma_W(v, w) = const \cdot \min(v, w). \tag{10.4.8}$$

Example 10.4.2 (The Wiener or Brownian Motion Process) A continuous-time Gaussian signal with stationary and independent increments with

$$\mathcal{C}_{\mathcal{W}}(w) = w, \qquad \gamma_W(v, w) = \min(v, w)$$

is called the Wiener stochastic process (or the Brownian motion process). Its sample trajectories are shown in Fig. 1.4. Notice that in this case, in view of Sect. 9.3, the autocovariance function gives a complete description of all finite-dimensional distributions of $\mathcal{W}(w)$. Indeed, given parameter values,

$$w_1 \le w_2 \le \ldots \le w_N,$$

the random vector

$$(\mathcal{W}(w_1), \mathcal{W}(w_2), \ldots, \mathcal{W}(w_N))$$

is a Gaussian random vector with the covariance matrix $\mathbf{\Gamma} = (\min(w_i, w_k))$, so that its joint c.d.f. can be explicitly calculated:

$$\mathbf{P}\Big(\mathcal{W}(w_1) \le a_1, \mathcal{W}(w_2) \le a_2, \ldots, \mathcal{W}(w_N) \le a_N\Big) \tag{10.4.9}$$

$$= \int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_N} \frac{e^{-\frac{\zeta_1^2}{2w_1}}}{\sqrt{2\pi\, w_1}} \cdot \frac{e^{-\frac{(\zeta_2-\zeta_1)^2}{2(w_2-w_1)}}}{\sqrt{2\pi(w_2-w_1)}} \cdot \ldots \cdot \frac{e^{-\frac{(\zeta_N-\zeta_{N-1})^2}{2(w_N-w_{N-1})}}}{\sqrt{2\pi(w_N-w_{N-1})}}$$

$$\times d\zeta_N \cdot \ldots \cdot d\zeta_2 \cdot d\zeta_1.$$

At this point, we are able to introduce the stochastic integral

$$\int_0^1 x(w)\, d\mathcal{W}(w),$$

with respect to a signal $\mathcal{W}(w)$ with uncorrelated increments, for a deterministic, possibly complex-valued function $x(w)$. If $x(w)$ is a step function of the form

$$x(w) = \sum_{i=1}^{N} x_i \mathbf{1}_{(w_{i-1}, w_i]}(w), \tag{10.4.10}$$

with $0 = w_0 < w_1 < \ldots < w_{N-1} < w_N = 1$, and $\mathbf{1}_A(w)$ denoting the indicator function of set A,[4] then, obviously

$$\int_0^1 x(w)\, d\mathcal{W}(w) := \sum_{i=1}^{N} x_i \cdot (\mathcal{W}(w_i) - \mathcal{W}(w_{i-1})). \tag{10.4.11}$$

Note that the variance of the stochastic integral in (10.4.11) is

$$\mathbf{E}\left|\int x(w)\, d\mathcal{W}(w)\right|^2 = \mathbf{E}\left|\sum_{i=1}^{N} x_i \cdot (\mathcal{W}(w_i) - \mathcal{W}(w_{i-1}))\right|^2$$

[4] Recall that the indicator function $\mathbf{1}_A(w)$ is defined as being equal to 1 for w belonging to set A, and being 0 for w outside A.

$$= \sum_{i=1}^{N} |x_i|^2 \mathbf{E}(\mathcal{W}(w_i) - \mathcal{W}(w_{i-1}))^2 = \sum_{i=1}^{N} |x_i|^2 (\mathcal{C}(w_i) - \mathcal{C}(w_{i-1}))$$

$$= \int_0^1 |x(w)|^2 \, d\mathcal{C}(w), \tag{10.4.12}$$

because, in view of (9.3.4), for any $0 < v < w$,

$$\mathbf{E}(\mathcal{W}(w) - \mathcal{W}(v))^2 = \mathcal{C}(w) + \mathcal{C}(v) - 2\mathcal{C}(v \wedge w) = \mathcal{C}(w) - \mathcal{C}(v). \tag{10.4.13}$$

Since any function $x(w)$ such that

$$\int_0^1 |x(w)|^2 \, d\mathcal{C}(w) < \infty \tag{10.4.14}$$

is a limit of a sequence $x_n(w)$ of step functions,[5] in the sense that

$$\int_0^1 |x_n(w) - x(w)|^2 \, d\mathcal{C}(w) \to 0, \quad \text{as} \quad n \to \infty,$$

the definition (10.4.11) of the stochastic integral for step functions can now be extended to any $x(w)$ satisfying condition (10.4.14), that is square integrable with respect to $d\mathcal{C}(w)$, by setting

$$\int_0^1 x(w) \, d\mathcal{W}(w) := \lim_{n \to \infty} \int_0^1 x_n(w) \, d\mathcal{W}(w), \tag{10.4.15}$$

where the limit is understood as the limit in the mean-square of random quantities (that is, variance, given that all the random quantities have zero means). In view of this procedure, the general stochastic integral for a function $x(w)$ satisfying condition (10.4.14) enjoys the "isometric" property

$$\mathbf{E} \left| \int_0^1 x(w) \, d\mathcal{W}(w) \right|^2 = \int_0^1 |x(w)|^2 \, d\mathcal{C}(w). \tag{10.4.16}$$

Example 10.4.3 (Gaussian Stochastic Integrals) Note that if the cumulative control function $\mathcal{C}(w)$ of a Gaussian process with independent increments $\mathcal{V}(w)$ has a density $c(w)$, that is,

[5] See, e.g., G. B. Folland, *Real Analysis*, W. Wiley, New York 1984.

$$\mathcal{C}(w) = \int_0^w c(v)\,dv, \qquad \frac{d\mathcal{C}(w)}{dw} = c(w) \geq 0, \quad 0 \leq w \leq 1,$$

then, in view of (10.4.16),

$$\mathbf{E}(\mathcal{V}(w))^2 = \mathbf{E}\left(\int_0^w d\mathcal{V}(v)\right)^2 = \int_0^w c(v)\,dv = \int_0^w \left(\sqrt{c(v)}\right)^2 dv,$$

which implies that, for any $x(w)$ satisfying (10.4.14), the statistical properties of the stochastic integrals,

$$\int_0^1 x(v)\,d\mathcal{V}(v), \qquad \text{and} \qquad \int_0^1 x(w)\sqrt{c(w)}\,d\mathcal{W}(w), \tag{10.4.17}$$

where $\mathcal{W}(w)$ is the Wiener process are the same. Later on this fact will serve as the basis of computer simulation of stationary random signals with a given spectrum.

Because, for any complex numbers ξ, η, we have the so-called "polarization formulas,"

$$\mathrm{Re}\,[\xi \cdot \eta^*] = \frac{1}{4}(|\xi + \eta|^2 - |\xi - \eta|^2),$$

$$\mathrm{Im}\,[\xi \cdot \eta^*] = \frac{1}{4}(|\xi + j\eta|^2 - |\xi - j\eta|^2),$$

which express the product in terms of the squared moduli, the "isometric" relation (10.3.16) extends from the mean-squares to scalar products. In other words, for any $x(w)$, $y(w)$, satisfying condition (10.4.14)

$$\mathbf{E}\left[\int_0^1 x(w)\,d\mathcal{W}(w) \cdot \left(\int_0^1 y(w)\,d\mathcal{W}(w)\right)^*\right] = \int_0^1 x(w) \cdot y^*(w)\,d\mathcal{C}(w). \tag{10.4.18}$$

10.5 Spectral Representation of Stationary Signals

The fundamental result about the structure of discrete-time stationary signals is that they are, essentially, sequences of random Fourier coefficients of stochastic processes with uncorrelated increments. More precisely, we have the following.

Spectral Representation Theorem *A discrete-time random signal* $X(n)$, $n = \ldots, -2, -1, 0, 1, 2, \ldots$, *is stationary if and only if it has the representation*

$$X(n) = \int_0^1 e^{j2\pi nf}\, d\mathcal{W}(f) \tag{10.5.1}$$

for a certain random process $\mathcal{W}(f)$, $0 \le f \le 1$, *which has uncorrelated increments. Moreover, the cumulative spectral function of* $X(n)$ *is identical to the cumulative control function of* $\mathcal{W}(f)$*, that is,*

$$\mathcal{S}_X(f) = \mathcal{C}_{\mathcal{W}}(f), \qquad 0 \le f \le 1. \tag{10.5.2}$$

Proof If random signal $X(n)$ is of the form (10.5.1), then it is stationary because it has zero mean and because, in view of the "isometry" (10.4.17),

$$\begin{aligned}\mathbf{E}[X(n)X^*(m)] &= \mathbf{E}\left[\int_0^1 e^{j2\pi nf}\, d\mathcal{W}(f) \cdot \left(\int_0^1 e^{j2\pi mf}\, d\mathcal{W}(f)\right)^*\right] \\ &= \int_0^1 e^{j2\pi(n-m)f}\, d\mathcal{C}_{\mathcal{W}}(f).\end{aligned}$$

The above calculation also identifies the *cumulative control function* of process $\mathcal{W}(f)$ as the cumulative spectral function of the random signal $X(n)$.

The proof of the reverse implication is more delicate as it requires identification, for each signal $X(n)$, of a process $\mathcal{W}(f)$ yielding representation (10.5.1). So, assume that $X(n)$ is a stationary signal with autocovariance sequence

$$\gamma_X(n) = \int_0^1 e^{j2\pi nf}\, d\mathcal{S}_X(f).$$

Denote by $L_0^2(\mathbf{P})$ the space of random quantities with zero mean and finite variance in the space $L^2(d\mathcal{S}_X(f))$ of complex functions on $[0, 1]$ which are square integrable with respect to cumulative spectral function $\mathcal{S}_X(f)$. Next, consider a linear mapping I from $L_0^2(\mathbf{P})$ into $L^2(d\mathcal{S}_X(f))$ defined by the identity

$$I[X(n)] := e^{j2\pi nf}, \qquad n = \ldots, -2, -1, 0, 1, 2, \ldots, \tag{10.5.3}$$

on complex exponentials and extended, in a natural way, to all their combinations. In other words, for any complex numbers $c_{-N}, \ldots, c_{-1}, c_0, c_1, \ldots, c_N$,

$$I\left[\sum_{n=-N}^{N} c_n X(n)\right] = \sum_{n=-N}^{N} c_n e^{j2\pi nf}. \tag{10.5.4}$$

Mapping I is an *isometry*[6] on such linear combinations because

$$\mathbf{E}\left|\sum_{n=-N}^{N} c_n X(n)\right|^2 = \sum_{n,m=-N}^{N} c_n c_m^* \mathbf{E}[X(n)X^*(m)]$$

$$= \sum_{n,m=-N}^{N} c_n c_m^* \int_0^1 e^{j2\pi(n-m)f}\, d\mathcal{S}_{\mathcal{W}}(f) = \int_0^1 \left|\sum_{n=-N}^{N} c_n e^{j2\pi nf}\right|^2 d\mathcal{S}_X(f),$$

and, as such, it extends to the linear isometry

$$I : \mathcal{L}[X(n), n = \ldots, -2, -1, 0, 1, 2, \ldots] \longmapsto L^2(d\mathcal{S}_X(f)),$$

where $\mathcal{L}[X(n), n = \ldots, -2, -1, 0, 1, 2, \ldots]$ is the subspace of $L^2(\mathbf{P})$ consisting of linear combinations of $X(n)$s and their mean-square limits. Since any isometry is necessarily a one-to-one mapping, I has a well-defined inverse

$$I^{-1} : L^2(d\mathcal{S}_X(f)) \longmapsto \mathcal{L}[X(n), n = \ldots, -2, -1, 0, 1, 2, \ldots],$$

which is also a linear isometry.

Now we will define a stochastic process $\mathcal{W}(f)$ by the formula

$$\mathcal{W}(f) := I^{-1}(\mathbf{1}_{[0,f]}),$$

where $\mathbf{1}_{[0,f]}(g), 0 \le g \le 1$, is the indicator function of the interval $[0, f]$. This process has zero mean and uncorrelated increments since, for $f_1 \le f_2 \le f_3$, in view of the isometric property of I^{-1},

$$\mathbf{E}[(\mathcal{W}(f_3) - \mathcal{W}(f_2)) \cdot (\mathcal{W}(f_2) - \mathcal{W}(f_1))]$$

$$= \mathbf{E}[(I^{-1}(\mathbf{1}_{[0,f_3]}) - I^{-1}(\mathbf{1}_{[0,f_2]})) \cdot (I^{-1}(\mathbf{1}_{[0,f_2]}) - I^{-1}(\mathbf{1}_{[0,f_1]}))]$$

$$= \mathbf{E}[I^{-1}(\mathbf{1}_{[0,f_3]}) - \mathbf{1}_{[0,f_2]})) \cdot (I^{-1}(\mathbf{1}_{[0,f_2]}) - \mathbf{1}_{[0,f_1]}))]$$

$$= \mathbf{E}[I^{-1}(\mathbf{1}_{(f_2,f_3]}) \cdot (I^{-1}(\mathbf{1}_{(f_1,f_2]})]$$

$$= \int_0^1 \mathbf{1}_{(f_2,f_3]}(f) \cdot \mathbf{1}_{(f_1,f_2]}(f)\, d\mathcal{S}_X(f) = 0.$$

The same calculation shows that

[6] In the sense that it preserves the norms: the standard deviation is the norm in space $L_0^2(\mathbf{P})$, and $\|a\| = (\int_0^1 |a(f)|^2\, d\mathcal{S}_X(f))^{1/2}$, for an $a(f)$ in $L^2(d\mathcal{S}_X(f))$.

$$\mathbf{E}\mathcal{W}^2(f) = \int_0^1 \mathbf{1}^2_{[0,f]}(g)\, d\mathcal{S}_X(g) = \mathcal{S}_X(f).$$

Now, proceeding again via step functions like in Sect. 10.3, using the linearity and isometry properties of I^{-1}, we have, for any function $a(f)$ in space $L^2(d\mathcal{S}_X(f))$,

$$I^{-1}(a) = \int_0^1 a(f)\, d\mathcal{W}(f).$$

In particular, selecting $a(f) = e^{j2\pi nf}$, we obtain that

$$X(n) = I^{-1}(e^{j2\pi nf}) = \int_0^1 e^{j2\pi nf}\, d\mathcal{W}(f),$$

which concludes the proof of the spectral representation theorem.

Example 10.5.1 (Spectral Representation of White Noise) Let $\mathcal{W}(f)$ be the Wiener process. Its cumulative control function

$$\mathcal{C}_{\mathcal{W}}(f) = f = \int_0^f df$$

has a control density function $c_{\mathcal{W}}(f) \equiv 1$. The stationary, discrete-time signal

$$X(n) = \int_0^1 e^{j2\pi nf}\, d\mathcal{W}(f)$$

has the spectral density function $S_X(f) = c_{\mathcal{W}}(f) \equiv 1$, and the autocovariance sequence

$$\gamma_X(n) = \mathbf{E}X(n)X^*(0) = \int_0^1 e^{j2\pi nf}\, df = \delta(n) = \begin{cases} 0, & \text{if } n \neq 0; \\ 1, & \text{if } n = 0. \end{cases}$$

Hence, $X(n)$ is the discrete-time white noise discussed in Chap. 6.

Example 10.5.2 (Spectral Representation of Filtered White Noise) Let $X(n)$ be the white noise discussed above. Consider the (acausal) filtered (i.e., moving average of) white noise

$$Y(n) = \sum_{k=-\infty}^{\infty} c_k X(n-k) = \int_0^1 \left(\sum_{k=-\infty}^{\infty} c_k e^{j2\pi(n-k)f} \right) d\mathcal{W}(f),$$

for $n = \ldots, -2, -1, 0, 1, 2, \ldots$. Its autocovariance sequence

$$\gamma_Y(n) = \mathbf{E}Y(n)Y^*(0) = \mathbf{E}\left(\sum_{k=-\infty}^{\infty} c_k X(n-k) \cdot \sum_{k=-\infty}^{\infty} c_k^* X^*(-k)\right)$$

$$= \mathbf{E}\sum_{k,l=-\infty}^{\infty} c_k c_l^* X(n-k)X^*(-l) = \sum_{k,l=-\infty}^{\infty} c_k c_l^* \delta(n-(k-l))$$

$$= \sum_{k,l=-\infty}^{\infty} c_k c_l^* \int_0^1 e^{j2\pi(n-(k-l))f}\, df = \int_0^1 |c(f)|^2 e^{j2\pi nf}\, df,$$

where

$$c(f) = \sum_{k=-\infty}^{\infty} c_k e^{-j2\pi kf}$$

is well defined as long as $\sum_{k=-\infty}^{\infty} |c_k|^2 < \infty$. Hence the power spectral density of the filtered white noise is

$$S_Y(f) = |c(f)|^2.$$

10.6 Computer Algorithms: Complex-Valued Case

Given a spectral density $S_X(f)$ of a discrete-time, stationary Gaussian signal $X(n)$ we can simulate a sample path of $X(n), n = 1, 2, \ldots, N$, by first calculating the autocovariance function $\gamma_X(n)$ using formula (10.3.2),

$$\gamma_X(n) = \int_0^1 e^{j2\pi nf} S_X(f)\, df, \tag{10.6.1}$$

and then by producing a sample of an N-dimensional Gaussian random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, with the covariance matrix $\mathbf{\Gamma} = (\gamma_X(n-m), n, m = 1, 2, \ldots, N)$, using the standard statistical software. This, however, would be computationally expensive, and even infeasible if n is large.

So, in this section we will describe a different, explicit algorithm for such a simulation based on the spectral representation of Sect. 9.4. The algorithm is mathematically justified by the discussions of the preceding sections, and it has the advantage of not being restricted to Gaussian signals.

The starting point is, of course, the Spectral Representation Theorem and, in particular, formula (10.4.1) which writes the signal $X(n)$ as a random Fourier coefficient,

$$X(n) = \int_0^1 e^{j2\pi nf}\, d\mathcal{W}(f), \qquad n = 1, 2, \ldots, N, \tag{10.6.2}$$

of a process $\mathcal{W}(f)$ with uncorrelated increments and cumulative control function $\mathcal{C}_{\mathcal{W}}(f)$ equal to the desired cumulative spectrum $\mathcal{S}_X(f)$.

We will assume that the spectrum of $X(n)$ is (absolutely) continuous, that is, it has a power spectrum density $S_X(f)$ such that

$$\mathcal{C}_{\mathcal{W}}(f) = \mathcal{S}_X(f) = \int_0^f S_X(g)\, dg. \tag{10.6.3}$$

For computational purposes the random integral (10.6.2) has to be discretized. More precisely, we have to choose an integer K, and partition

$$f_0 = 0,\ f_1 = \frac{1}{K},\ f_2 = \frac{2}{K}, \ldots, f_{K-1} = \frac{K-1}{K},\ f_K = 1,$$

of the interval [0, 1], and replace the right-hand side of (10.6.2) by the sums

$$X_K(n) = \sum_{k=1}^{K} e^{j2\pi nf_k}\Big(\mathcal{W}(f_k) - \mathcal{W}(f_{k-1})\Big)$$

$$= \sum_{k=1}^{K} e^{j2\pi n(k/K)}\Big(\mathcal{W}\Big(\frac{k}{K}\Big) - \mathcal{W}\Big(\frac{k-1}{K}\Big)\Big).$$

The increments

$$\mathcal{W}\Big(\frac{1}{K}\Big) - \mathcal{W}\Big(\frac{0}{K}\Big),\ \ \mathcal{W}\Big(\frac{2}{K}\Big) - \mathcal{W}\Big(\frac{1}{K}\Big),\ \ \ldots,\ \ \mathcal{W}\Big(\frac{K}{K}\Big) - \mathcal{W}\Big(\frac{k-1}{K}\Big),$$

are zero-mean, uncorrelated and have, respectively, variances

$$\sigma_1^2 = \int_0^{1/K} S_X(f)\, df,\ \ \sigma_2^2 = \int_{1/K}^{2/K} S_X(f)\, df,\ \ \ldots, \sigma_K^2 = \int_{(K-1)/K}^{1} S_X(f)\, df.$$

Hence, the total mean powers of $X(n)$ and $X_K(n)$ match exactly. Thus the simulation algorithm calls for the following steps:

Step 0: *Select a positive integer K determining the accuracy of our simulation.*
Step 1: *Generate, via a random number generator, a sequence*

$$\xi_1, \xi_2, \ldots, \xi_K,$$

of zero-mean, variance one, uncorrelated random values of an otherwise arbitrary distribution.

Step 2: *Calculate variances*

$$\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2$$

defined above via the desired power spectrum density.

Step 3: *Calculate the complex numbers,*

$$x_n = \sum_{k=1}^{K} e^{j2\pi n(k/K)} \sigma_k \xi_k, \qquad n = 1, 2, \ldots, N.$$

They represent an approximate sample of our desired random signal.

Step 4: *Plot the real and imaginary parts of the sequence* x_n, $n = 1, 2, \ldots, N$,

$$\text{Re } x_n = \sum_{k=1}^{K} \cos(j2\pi n(k/K))\sigma_k \xi_k, \quad \text{Im } x_n = \sum_{k=1}^{K} \sin(j2\pi n(k/K))\sigma_k \xi_k,$$

as functions of variable n.

Remark 10.6.1 It should be observed that if the power spectrum density is symmetric about the midpoint $f = 1/2$, that is, $S_X(1/2 + f) = S_X(1/2 - f)$, then the autocovariance function is real-valued because

$$\gamma_X(n) = \int_0^1 e^{j2\pi nf} S_X(f)\, df = \int_0^1 \cos(2\pi nf) S_X(f)\, df.$$

We shall illustrate the above algorithm on a concrete example implemented in the symbolic manipulation language *Mathematica*.

Example 10.6.1 (Mathematica Simulation of a Complex-Valued Stationary Signal) The goal is to simulate a discrete-time signal $X(n)$, $n = 1, 2, \ldots, 150$, with the spectral density function $S_X(f) = f(1 - f)$, $0 \le f \le 1$, pictured below.

Step 0. Selecting a positive integer K determining the accuracy of the simulation.

```
In[1]:=  K=100
Out[1]= 100
```

Step 1: Generating, via a pseudo-random number generator, a sequence

$$\xi_1, \xi_2, \ldots, \xi_K,$$

of zero-mean, variance one, uncorrelated random values of an otherwise arbitrary distribution. Here we start with a sample of 100 pseudo-random numbers with the Gaussian, $N[0, 1]$-distribution, see Fig. 10.1.

```
In[2]:= xi = Table[Random[NormalDistribution[0, 1]], {100}]
```

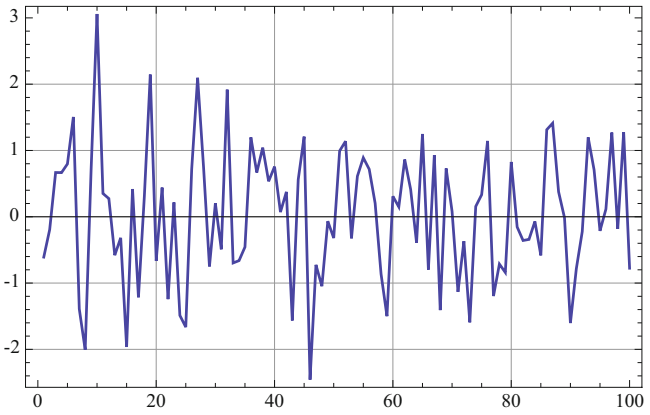


Fig. 10.1 A sample of 100 pseudo-random numbers with the Gaussian, $N(0, 1)$ distribution

```
Out[2]= {-0.608542, -0.193407, 0.667423, 0.665791, 0.796963,
1.50578, -1.38957, -2.00677, 0.710005, 3.05874, 0.351129,
0.274176, -0.57993, -0.317531, -1.9642, 0.418438, -1.21485,
0.311505, 2.14493, -0.665234, 0.440417, -1.24286, 0.217456,
-1.48803, -1.66472, 0.720181, 2.09662, 0.751509, -0.748984,
0.203246, -0.490937, 1.91771, -0.696637, -0.661528, -0.456505,
1.19835, 0.667494, 1.04284, 0.534665, 0.756436, 0.0707936,
0.375792, -1.56415, 0.559878, 1.20885, -2.45781, -0.724939,
-1.04777, -0.0669847, -0.321047, 0.993232, 1.1395, -0.325509,
0.611529, 0.890348, 0.716697, 0.203702, -0.863057, -1.49988,
0.308803, 0.148938, 0.863372, 0.413497, -0.392592, 1.24894,
-0.795932, 0.929254, -1.40817, 0.728825, 0.0811022, -1.13286,
-0.368274, -1.59267, 0.155889, 0.332486, 1.14419, -1.19604,
-0.713426, -0.839724, 0.827024, -0.154212, -0.357799,
-0.341499, -0.0706729, -0.58252, 1.31315, 1.41184, 0.376868,
-0.0139196, -1.60352, -0.783236, -0.223895, 1.19736, 0.707607,
-0.212544, 0.115375, 1.27051, -0.18183, 1.27593, -0.775792}

In[2]:=  ListPlot[xi, PlotJoined->True, Frame->True,
           GridLines->Automatic]
Out[2]  -Graphics-
```

Step 2: Calculation of standard deviations,

$$\sigma_1, \sigma_2, \ldots, \sigma_K,$$

defined via the above power spectrum density.

```
In[4]:= SX[f_]:=f*(1-f)
In[5]:= sigma =
          Table[ Sqrt[NIntegrate[SX[f], {f, (k-1)/100,
          (k )/100}]], {k,1,100}]
```

```
Out[5]= {0.00704746,0.0121518,0.0156098,0.0183757,0.0207284,
  0.0227962,0.0246509,0.0263376,0.0278867,0.0293201,0.030654,
  0.0319009,0.0330706,0.0341711,0.0352089,0.0361893,0.0371169,
  0.0379956,0.0388287,0.039619,0.0403691,0.0410812,0.0417572,
  0.0423989,0.0430078,0.0435852,0.0441324,0.0446505,0.0451405,
  0.0456034,0.0460398,0.0464507,0.0468366,0.0471982,0.047536,
  0.0478505,0.0481422,0.0484114,0.0486587,0.0488842,0.0490884,
  0.0492714,0.0494335,0.0495749,0.0496957,0.0497963,0.0498765,
  0.0499366,0.0499767,0.0499967,0.0499967,0.0499767,0.0499366,
  0.0498765,0.0497963,0.0496957,0.0495749,0.0494335,0.0492714,
  0.0490884,0.0488842,0.0486587,0.0484114,0.0481422,0.0478505,
  0.047536,0.0471982,0.0468366,0.0464507,0.0460398,0.0456034,
  0.0451405,0.0446505,0.0441324,0.0435852,0.0430078,0.0423989,
  0.0417572,0.0410812,0.0403691,0.039619,0.0388287,0.0379956,
  0.0371169,0.0361893,0.0352089,0.0341711,0.0330706,0.0319009,
  0.030654,0.0293201,0.0278867,0.0263376,0.0246509,0.0227962,
  0.0207284,0.0183757,0.0156098,0.0121518,0.00704746}
```

Step 3: Calculation of numbers

$$\text{Re } x_n = \sum_{k=1}^{K} \cos(2\pi n(k/K))\sigma_k \xi_k, \qquad n = 1, 2, \ldots, N,$$

and

$$\text{Im } x_n = \sum_{k=1}^{K} \sin(2\pi n(k/K))\sigma_k \xi_k, \qquad n = 1, 2, \ldots, N,$$

for $N = 150$. They represent an approximate samples of the real and imaginary parts of our desired random signal.

```
In[6] ReXi= Table[N[Sum[Cos[2*Pi*n*(k/100)] * sigma[[k]]
           * xi[[k]], {k,1,100}]],{n,1,150}]
```

```
Out[6] = {-0.023415, 0.204973, 0.262053, -0.306833, 0.0423987,
0.0801657,-0.114673, 0.180827, -0.182326, 0.0501663, 0.241876,
-0.422759,-0.267774, -0.2427, 0.018383, 0.664823, -0.415174,
0.173961,-0.0833322, 0.197514, -0.078882, 0.203239, 0.00381133,
-0.486851,0.193364, -0.182158, 0.0293311, -0.381732, 0.304001,
0.0549667,0.410134, -0.0548758, 0.104368, 0.00517703, -0.213219,
0.0621887,0.122844, 0.119623, -0.21869, -0.00453364, -0.416995,
0.0884643,0.459038, -0.279907, 0.0401727, -0.216858, 0.00620257,
-0.202628,0.0410997, 0.211609, 0.0410997, -0.202628, 0.00620257,
-0.216858,0.0401727, -0.279907, 0.459038, 0.0884643, -0.416995,
-0.00453364,-0.21869, 0.119623, 0.122844, 0.0621887, -0.213219,
0.00517703,0.104368, -0.0548758, 0.410134, 0.0549667, 0.304001,
-0.381732,0.0293311, -0.182158, 0.193364, -0.486851, 0.00381133,
0.203239,-0.078882, 0.197514, -0.0833322, 0.173961, -0.415174,
0.664823,0.018383, -0.2427, -0.267774, -0.422759, 0.241876,
0.0501663,-0.182326, 0.180827, -0.114673, 0.0801657, 0.0423987,
```

```
-0.306833,0.262053, 0.204973, -0.023415, 0.103954, -0.023415,
0.204973,0.262053, -0.306833, 0.0423987, 0.0801657, -0.114673,
0.180827,-0.182326, 0.0501663, 0.241876, -0.422759, -0.267774,
-0.2427,0.018383, 0.664823, -0.415174, 0.173961, -0.0833322,
0.197514,-0.078882, 0.203239, 0.00381133, -0.486851, 0.193364,
-0.182158,0.0293311, -0.381732, 0.304001, 0.0549667, 0.410134,
-0.0548758,0.104368, 0.00517703, -0.213219, 0.0621887, 0.122844,
0.119623,-0.21869, -0.00453364, -0.416995, 0.0884643, 0.459038,
-0.279907,0.0401727, -0.216858, 0.00620257, -0.202628,
0.0410997, 0.211609}


In[7]:= ImXi = Table[N[Sum[Sin[2*Pi*n*(k/100)]*sigma[[k]]
                        *xi[[k]], {k, 1, 100}]], {n, 1, 150}]

Out[7]= {0.15977, -0.103151, -0.157232, 0.333, -0.139511,
0.245695, -0.43247, 0.407358, -0.70167, -0.0945059, 0.27421,
0.58988, 0.0705348, -0.11186, -0.0567596, -0.0596612,
-0.574812, -0.467159, 0.0811688, 0.38486, -0.463603, 0.178059,
0.791538, -0.0854149, -0.0661586, -0.106904, 0.0448853,
0.110552, -0.261648, -0.19714, -0.26017, 0.357341, -0.276876,
0.314915, 0.108389, -0.143431, -0.232836, -0.121447, 0.474415,
-0.426709, 0.176697, -0.123609, -0.138301, 0.132275, 0.660073,
-0.661418, -0.361657, 0.239999, -0.134132, 0., 0.134132,
-0.239999, 0.361657, 0.661418, -0.660073, -0.132275, 0.138301,
0.123609, -0.176697, 0.426709, -0.474415, 0.121447, 0.232836,
0.143431, -0.108389, -0.314915, 0.276876, -0.357341, 0.26017,
0.19714, 0.261648, -0.110552, -0.0448853, 0.106904, 0.0661586,
0.0854149, -0.791538, -0.178059, 0.463603, -0.38486,
-0.0811688, 0.467159, 0.574812, 0.0596612, 0.0567596, 0.11186,
-0.0705348, -0.58988, -0.27421, 0.0945059, 0.70167, -0.407358,
0.43247, -0.245695, 0.139511, -0.333, 0.157232, 0.103151,
-0.15977, 0., 0.15977, -0.103151, -0.157232, 0.333, -0.139511,
0.245695, -0.43247, 0.407358, -0.70167, -0.0945059, 0.27421,
0.58988, 0.0705348, -0.11186, -0.0567596, -0.0596612, -0.574812,
-0.467159, 0.0811688, 0.38486, -0.463603, 0.178059, 0.791538,
-0.0854149, -0.0661586, -0.106904, 0.0448853, 0.110552,
-0.261648, -0.19714, -0.26017, 0.357341, -0.276876, 0.314915,
0.108389, -0.143431, -0.232836, -0.121447, 0.474415, -0.426709,
0.176697, -0.123609, -0.138301, 0.132275, 0.660073, -0.661418,
-0.361657, 0.239999, -0.134132, 0.}
```

Step 4: Plotting the complex-valued sequence x_n as a function of variable n. The consecutive values of real (left plot) and imaginary (right plot) parts of the numbers $x_1, \ldots, x_{150}$ were joined in Fig. 10.2 to better show their progression in time.

```
In[8]:=  ListPlot[ReXi, PlotJoined->True, Frame->True,
            GridLines->Automatic]
Out[8]  -Graphics-
In[9]:=  ListPlot[ImXi, PlotJoined->True, Frame->True,
            GridLines->Automatic]
Out[9]  -Graphics-
```

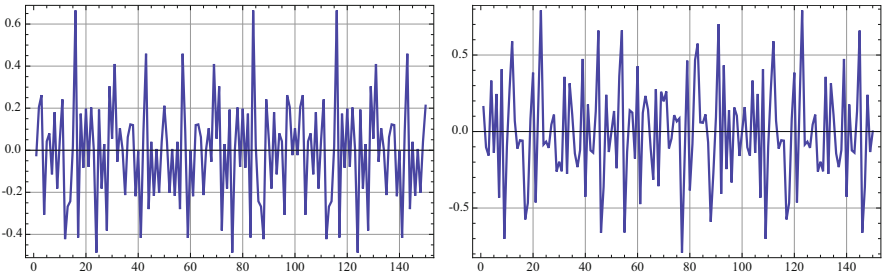


Fig. 10.2 Samples of the real (left) and imaginary (right) components of a stationary signal with spectral density $S_X(f) = f(1-f), 0 \le f \le 1$

Note that, for $K = 100$, the smallest frequency present in the representation is $f = 1/100$. Thus the produced signal sample is periodic with period $P = 100$ (Fig. 10.2).

Remark 10.6.2 The above simulation can be adapted to any discrete-time signal $X(t_n)$ with $t_n = n \cdot \Delta t$, extending the procedures described above in the case $\Delta t = 1$ (see Problem 5.5.3). In the theoretical limit, $\Delta t \to 0$, one obtains the spectral representation of continuous-time (see Problem 5.5.4).

Remark 10.6.3 The fact that the spectral density was concentrated on the interval [0, 1] was related to selection of the complex exponentials of the form $e^{j2\pi nf}$ in the Spectral Representation Theorem. A different selection of complex exponentials would lead to different intervals. For example, choosing the complex exponentials of the form $e^{jn\omega}$, that is, conducting spectral analysis in terms of the angular velocity rather than the frequency, would lead to spectral densities concentrated on the interval $[0, 2\pi]$, or any other interval of length 2π. Figure 10.3 shows several examples of such spectral densities concentrated on the symmetric frequency intervals $[-\pi, +\pi]$, and the real parts of the sample paths of the corresponding stationary signals.

Remark 10.6.4 Another way to produce a graphical representation of the complex-valued signal x_n considered in the above Example 10.6.1 would be to plot its moduli and arguments instead of its real and imaginary parts.

10.7 Computer Algorithms: Real-Valued Case

To produce a sample of a real-valued stationary discrete-time signal it is not enough to take a real part of the complex-valued signal because the real part of a complex-valued stationary signal need not be stationary at all. Indeed, as we have observed before (see Example 5.1.9), the simple complex random harmonic oscillation

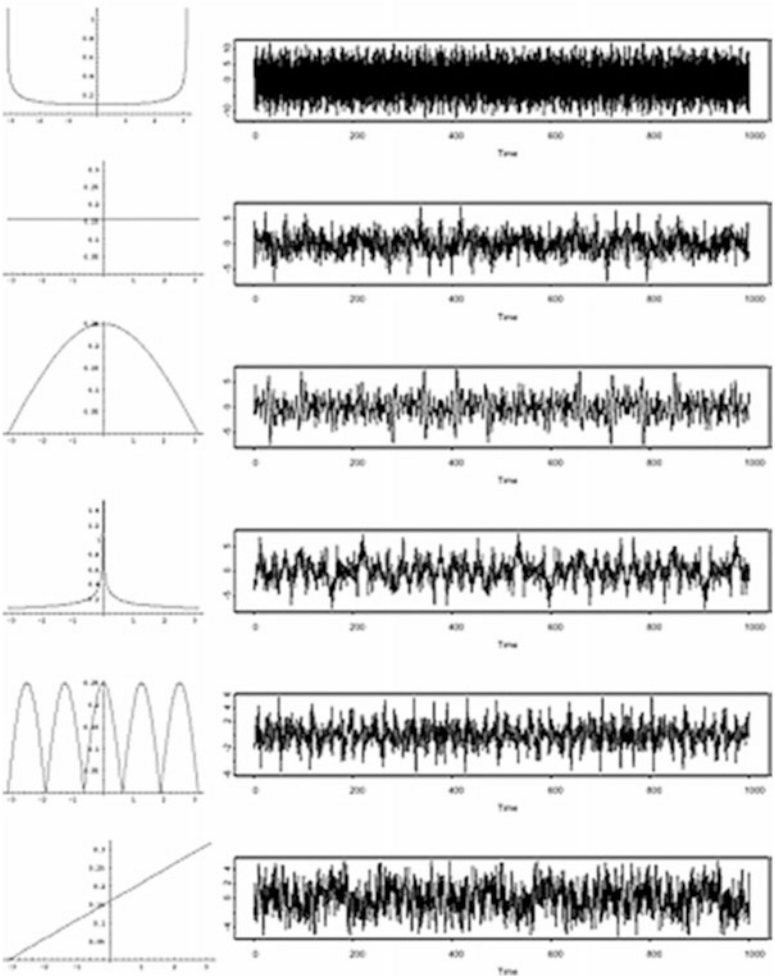


Fig. 10.3 Examples of real parts of simulated discrete-time stationary signals (right column) with prescribed spectral density functions (left column). Note that the spectral densities in these simulations are even and concentrated on the interval $[-\pi, +\pi]$

$$X(n) = A \cdot e^{j2\pi f_0 n},$$

with the zero-mean random amplitude A, is stationary, with ACvF

$$\mathbf{E}X^*(n)X(n+\tau) = \mathbf{E}\left(A^* e^{-j2\pi f_0 n} \cdot A e^{j2\pi f_0(n+\tau)}\right) = \mathbf{E}|A|^2 \cdot e^{j2\pi f_0 \tau} = \gamma_X(\tau),$$

but its real part,

$$\mathrm{Re}X(n) = A \cdot \cos(2\pi f_0 n),$$

is not because

$$\mathbf{E}[\mathrm{Re}X^*(n) \cdot \mathrm{Re}X(n+\tau)] = \mathbf{E}|A|^2 \cdot \cos(2\pi f_0 n)\cos(2\pi f_0(n+\tau))$$

$$= \frac{1}{2}\mathbf{E}|A|^2\Big(\cos(2\pi f_0(2n+\tau)) + \cos(2\pi f_0 \tau)\Big)$$

obviously depends not only on the time lag τ but also on the time n.

The solution here becomes clear if we abandon the complex domain altogether and return full circle to the very first examples of stationary signals discussed in Chap. 4 (see Examples 5.1.2 and 5.1.3), this time considering them in discrete time, $n = \ldots, -1, 0, 1, \ldots,$. Without going again through all the rigorous mathematical details developed in the complex case earlier in this chapter we just present the basic algorithm.

Consider the real-valued superposition of harmonic oscillations with distinct frequencies $f_1, \ldots, f_K$,

$$X_K(n) = \sum_{k=1}^{K} A_k \cos(2\pi f_k(n + \Theta_k)), \tag{10.7.1}$$

where A_k, $k = 1, \ldots, K$ are independent, zero-mean real-valued random amplitudes, and Θ_k, $k = 1, \ldots, K$ are independent random phases, independent of the amplitudes and uniformly distributed over the corresponding periods $P_k = 1/f_k$, $k = 1, \ldots, K$.

The signal $X_K(n)$ has, obviously, zero mean, and (following calculations analogous to those in Example 5.1.2) the autocovariance sequence

$$\mathbf{E}X_K(n)X_K(n+\tau) = \mathbf{E}\sum_{k=1}^{K} A_k \cos(2\pi f_k(n+\Theta_k)) \cdot \sum_{l=1}^{K} A_k \cos(2\pi f_l(n+\tau+\Theta_l))$$

$$= \sum_{k=1}^{K}\sum_{l=1}^{K} \mathbf{E}\Big[\Big(A_k \cos(2\pi f_k(n+\Theta_k))\Big)\Big(A_l \cos(2\pi f_l(n+\tau+\Theta_l))\Big)\Big]$$

$$= \sum_{k=1}^{K} \mathbf{E}A_k^2 \cdot \mathbf{E}\Big(\cos(2\pi f_k(n+\Theta_k)) \cdot \cos(2\pi f_k(n+\tau+\Theta_k))\Big).$$

Taking into account the trigonometric formula for the product of the cosines in Table 1.1, and the uniform distributions of Θ_k, over the periods P_k, we finally obtain the autocovariance sequence

$$\gamma_{X_K}(\tau) = \mathbf{E}X_K(n)X_K(n+\tau) = \frac{1}{2}\sum_{k=1}^{K} \mathbf{E}A_k^2 \cdot \cos(2\pi f_k \tau). \tag{10.7.2}$$

The above autocovariance sequence corresponds to the power spectrum (see Example 5.2.2)

$$S_{X_K}(f) = \frac{1}{4}\sum_{k=1}^{K} \mathbf{E}A_k^2 \cdot \Big(\delta(f-f_k) + \delta(f+f_k)\Big). \tag{10.7.3}$$

Now, let us consider an arbitrary even power spectrum $S_X(f)$, i.e., satisfying condition

$$S_X(-f) = -S_X(f),$$

and restricted, for the sake of convenience, to the symmetric interval $[-1/2, +1/2]$. The strategy is to approximate $S_X(f)$ by $S_{X_K}(f)$ described in (10.7.3), while preserving the total power, that is requiring that

$$PW_X = \int_{-1/2}^{+1/2} S_X(f)\,df = \int_{-1/2}^{+1/2} S_K(f)\,df = PW_{X_K} = \gamma_{X_K}(0). \tag{10.7.4}$$

The frequencies f_k will now be taken to correspond to the partition of the interval $[0, 1/2]$, that is

$$f_k = \frac{1}{2}\cdot\frac{k}{K}, \quad k = 1, \ldots K. \tag{10.7.5}$$

So, it suffices to select the random amplitudes A_k, so that

$$\frac{1}{4}\mathbf{E}A_k^2 = \int_{f_{k-1}}^{f_k} S_X(f)\,df \equiv \sigma_k^2, \quad k = 1, \ldots, K. \tag{10.7.6}$$

As a result,

$$PW_{X_K} = \frac{1}{2}\sum_{k=1}^{K} \mathbf{E}A_k^2 = 2\sum_{k=1}^{K}\int_{f_{k-1}}^{f_k} S_X(f)\,df = 2\int_0^{1/2} S_X(f)\,df = PW_X,$$

as required.

Now if we start with realizations of the standard white noise, $\xi_k = W(k), k = 1, \ldots, K$, (of any distribution), and the white noise $\eta_k = U(k), k = 1, \ldots, K$, uniformly distributed on $[-1/2, +1/2]$, then

$$\theta_k = \frac{2K}{k}\eta_k, \quad k = 1, \ldots, K,$$

form a realization of independent random quantities, uniformly distributed over the intervals

$$\left[-\frac{K}{k}, +\frac{K}{k}\right] = \left[-\frac{P_k}{2}, +\frac{P_k}{2}\right], \quad k = 1, \ldots, K,$$

respectively, and the superposition of real-valued harmonic oscillations,

$$X_K(n) = \sum_{k=1}^{K} \sigma_k \xi_k \cos\Big(2\pi f_k (n + 2K\eta_k/k)\Big) \tag{10.7.7}$$

will approximate, in the mean-square sense, signal $X(n)$ with the required power spectrum $S_X(f)$.

Example 10.7.1 (Mathematica Simulation of a Real-Valued Stationary Signal) We will implement the above algorithm for the power spectrum $S_X(f) = 100f^2$, $-1/2 \le f \le +1/2$, and a Gaussian signal. The general outlines are the same as in Example 10.6.1. First we obtain a sample ξ_k of length $K = 100$ of the standard Gaussian white noise:

```
In[1]:= xi = Table[Random[NormalDistribution[0, 1]], {100}]

Out[1]= {-0.856053, 1.08187, 2.46229, 0.714797, 0.714182,
-0.213566, -0.433184, -0.851746, -0.0462548, 1.50339,
-1.51236, -1.28448, 0.0673793, -0.108364, 0.270925,-0.330244,
1.35095, -0.44158, -0.357206, -0.647803, -1.09377, -1.34072,
0.849032, 0.0500218, -0.575234, -0.0171291, -1.79476, 1.31388,
-0.628999, -0.593384, -0.464793, 1.90548, 0.691585, -0.426236,
-0.420072, 0.133262, -0.0273259, -0.499321, -0.169682,
-0.91716, 1.63794, 0.746604, 0.0121301, 0.997426, 1.3202,
-0.510749, -0.198871, -0.439695, 0.908916, 1.75012, -0.244048,
0.0384926, 0.182402, 0.00244352, -2.0007, 0.259864, -0.755299,
-1.06697, 0.177168, 0.518347, 0.127846, -0.426915, 0.831972,
0.130949, -0.708484, 0.744263, 0.0306772, -2.40272, -0.388865,
1.04692, -2.36268, 1.26858, 0.020974, -1.19099, -0.0972772,
-1.11214, -0.253469, -1.07956, -1.73907, 1.55135, -0.273338,
0.814078, 0.280743, 0.199324, 1.59616, -0.569614, -1.32923,
```

```
-0.0159629, 1.58278, -0.966994, -1.19754, -1.77986, 1.41761,
-1.27518, 0.322685, -0.398681, 1.02684, -0.735058, -0.141971,
-0.41919}
```

The next step is to produce a sample of length $K = 100$ of the white noise η_k uniformly distributed on the interval $-1/2, +1/2$. This is accomplished by first producing a white noise uniformly distributed on $[0, 1]$ and then subtracting 1/2 from each of its terms:

```
In[2]:= eta = Table[Random[Real, {0, 1}] - 1/2, {100}]

Out[2]= {0.041948, 0.484289, -0.318925, -0.0276171, 0.0359713,
-0.088659, 0.252302, 0.353539, 0.255555, 0.089573, -0.0901944,
0.227213, 0.0284539, -0.273957, 0.441175, -0.189807,
-0.0364003, 0.273394, 0.445258, -0.40948, 0.152135, -0.333722,
-0.124852, -0.42935, -0.389813, -0.318011, -0.305928,
0.0982668, 0.0742158, 0.270648, -0.0582301, 0.244727, 0.318661,
-0.318925, -0.468036, -0.482485, -0.209793, 0.455031,
-0.409211, 0.207322, 0.326608, -0.318363, -0.354468, 0.116801,
-0.325528, -0.484641, 0.270384, 0.0461516, -0.435715, 0.33337,
0.0763118, 0.447885, -0.00993046, -0.437278, -0.365458,
-0.296843, 0.171408, 0.381647, -0.397422, -0.314357, -0.118799,
0.426616, -0.488212, -0.0216788, 0.0545938, 0.244979, 0.366257,
0.36152, -0.119879, 0.22962, -0.404127, -0.184632, -0.184164,
0.39625, 0.0195609, -0.132516, 0.325766, 0.333528, -0.114981,
-0.335674, -0.345642, 0.451881, -0.217559, 0.478683, 0.273157,
-0.474735, -0.229347, 0.000362175, -0.281437, -0.219714,
-0.095604, 0.138842, 0.338442, 0.0506663, -0.191477, -0.176526,
0.0226057, 0.154416, 0.288962, 0.45599}
```

The standard deviations σ_k are

```
In[3]:= sigma=Table[ Sqrt[NIntegrate[100 f^2, {f, (k - 1)/200,
                        (k )/200}]], {k, 1, 100}]
Out[3]={0.00204124, 0.00540062, 0.00889757, 0.0124164,
0.0159426, 0.0194722, 0.0230036, 0.0265361, 0.0300694,
0.0336031, 0.0371371, 0.0406714, 0.044206, 0.0477406,
0.0512754, 0.0548103, 0.0583452, 0.0618803, 0.0654153,
0.0689505, 0.0724856, 0.0760208, 0.0795561, 0.0830913,
0.0866266, 0.0901619, 0.0936972, 0.0972325, 0.100768,
0.104303, 0.107839, 0.111374, 0.114909, 0.118445, 0.12198,
0.125516, 0.129051, 0.132586, 0.136122, 0.139657, 0.143193,
0.146728, 0.150264, 0.153799, 0.157335, 0.16087, 0.164405,
0.167941, 0.171476, 0.175012, 0.178547, 0.182083, 0.185618,
0.189154, 0.192689, 0.196225, 0.19976, 0.203296, 0.206831,
0.210367, 0.213902, 0.217438, 0.220973, 0.224509, 0.228044,
0.23158, 0.235115, 0.238651, 0.242186, 0.245722, 0.249257,
0.252793, 0.256328, 0.259864, 0.263399, 0.266935, 0.27047,
0.274006, 0.277541, 0.281077, 0.284612, 0.288148, 0.291683,
0.295219, 0.298754, 0.30229, 0.305825, 0.309361, 0.312896,
0.316432, 0.319967, 0.323503, 0.327038, 0.330574, 0.33411,
0.337645, 0.341181, 0.344716, 0.348252, 0.351787}
```

Entering the above data in the formula (10.7.7) gives us a sample of 150 consecutive values of the desired signal.

```
In[4]:= xn = Table[
          N[Sum[Cos[2*Pi*(n + (200*eta[[k]]/k))*(k/200)]
             *sigma[[k]]*xi[[k]],{k, 1, 100}]], {n, 1, 150}]
Out[4] = {0.902888, -1.44987, 0.73034, -0.0467446, -0.981386,
1.85137, -1.14327, 0.505844, -0.702551, -0.480106, 2.02188,
-0.324226, -2.58959, 3.54, -2.15705, 0.535498, -0.696431,
1.32388, -0.624256, -1.37706, 0.71946, 0.258689, 2.11401,
-2.44596, 1.29879, -1.55369, 0.938206, 0.444133, -0.487131,
0.11673, 0.286159, -1.11297, 0.330835, -0.246013, -0.0297903,
2.28609, -0.11916, -2.36099, 2.56028, -2.65337, -0.0894128,
2.2119, -0.816799, 0.344593, -0.698824, -0.470619, 0.502274,
0.940286, -1.94194, 0.897492, 1.39516, -1.53023, -0.126247,
0.8947, -0.958154, 0.199293, 0.66053, -1.34534, 1.75322,
-0.338096, -2.11878, 3.2534, -1.21718, -0.405543, 0.413332,
-0.375367, -2.4344, 4.33465, -4.15149, 2.44168, -1.26502,
1.6717, 0.914773, -2.10289, 0.713696, -0.939291, 0.124809,
-0.515525, 0.914653, 0.102627, 0.567457, 0.766725, -1.09135,
0.278225, -2.12101, 1.92608, 1.28077, -0.336511, -1.8577,
0.656761, 1.33076, -1.24178, -0.317488, -0.0655308, 0.540343,
0.00291415, -0.714359, 1.1559, -1.01383, 0.619388, 1.85065,
-3.39279, 2.73494, -1.73749, 0.369481, 0.452425, 0.801605,
-1.06127, 1.04946, -1.34485, 0.351694, -0.0323086, 0.0127435,
-1.72899, 0.569055, 1.27245, -1.53539, 2.53497, -1.98056,
1.01728, 0.252221, -0.123346, -0.963119, 1.52522, -2.59951,
2.70631, -0.853903, 0.17498, -1.08285, -0.805603, 3.50613,
-4.1166, 2.18343, -2.56471, 2.55596, 0.624361, -2.45507,
2.09628, -0.794994, -0.201666, 0.713224, 0.803646, -1.89323,
1.88523, -1.57122, 0.958893, -2.30286, 2.02618, 0.408114,
-1.40083}
```

This sample path is then plotted in Fig. 10.4. To visualize its progression in time better, the discrete plot points are joined.

```
In[5]:= ListPlot[xn, PlotJoined -> True, Frame -> True,
 GridLines -> Automatic,  PlotStyle -> {Thickness[0.005]}]

Out[5]=
```

10.8 Problems and Exercises

1 Verify the polarization formulas preceding the "isometric" formula (10.4.17)

2 Given a discrete-time stationary signal $X(n)$ with cumulative power spectrum $\mathcal{S}_X(f)$, find the cumulative power spectrum for the filtered signal $Y(n) = \sum_{k=-\infty}^{\infty} c_k X(n-k)$. Follow calculations in Example 10.5.2. Repeat the calculation in case when $X(n)$ has the power spectral density.

3 Extend the spectral representation (and the algorithm based on it) in the case of discrete-time signal $X(t_n)$, with $t_n = n \cdot \Delta t$, extending procedures described above in the case $\Delta t = 1$.

4 Find the theoretical spectral representation for continuous-time stationary signals taking $\Delta t \to 0$ in Problem 3.

5 Use the simulation algorithm described in Sect. 10.5 to produce sample trajectories of complex-valued signals with the following spectral density functions defined on the interval $0 \le f \le 1$. Plot the spectral densities first.

(a)

$$S(f) = \frac{1}{\sqrt{f(1-f)}}.$$

What is special about this spectrum? Check that the power of the corresponding signal is finite.

(b)

$$S(f) = 2/3,$$

(c)

$$S(f) = |\cos(\pi f)|,$$

(d)

$$S(f) = 1 - |f|,$$

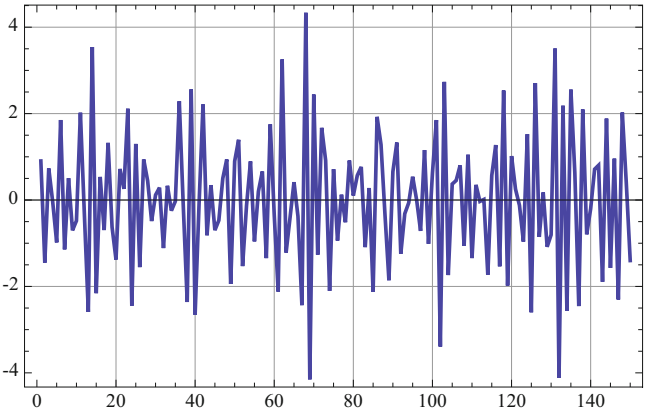


Fig. 10.4 A sample of the Gaussian stationary signal with the spectral density $S_X(f) = 100f^2$, $-1/2 \le f \le 1/2$

(e)

$$S(f) = |\sin(8\pi f)|.$$

In the simulations start with: (A) the white noise having the $N(0, 1)$ distributions, (B) the white noise having the $U(-1, 1)$ distributions normalized to have variance 1.

6 Use the simulation algorithm described in Sect. 10.6 to produce sample trajectories of real-valued signals with the following spectral density functions defined on the interval $-1/2 \le f \le +1/2$. Plot the spectral densities first.

(a)

$$S(f) = \frac{1}{\sqrt{(1/2 + f)(1/2 - f)}}$$

(b)

$$S(f) = 2/3$$

(c)

$$S(f) = \cos(\pi f)$$

(d)

$$S(f) = 1 - f$$

(e)

$$S(f) = |\sin(8\pi f)|$$

In the simulations start with: (A) the white noise having the $N(0, 1)$ distributions, (B) the white noise having the $U(-1, 1)$ distributions normalized to have variance 1.

7 Produce plots of several sample paths of the cumulative discrete-time white noise defined in Sect. 10.3. Use: (A) the white noise having the $N(0, 1)$ distributions, (B) the white noise having the $U(-1, 1)$ distributions normalized to have variance 1.

8* Verify that the additivity property (10.4.7) of any continuous function forces its linear form (10.4.8). Start with checking the property for the integers, then move on to rational numbers, and finally extend the result to all real numbers using the continuity assumption.

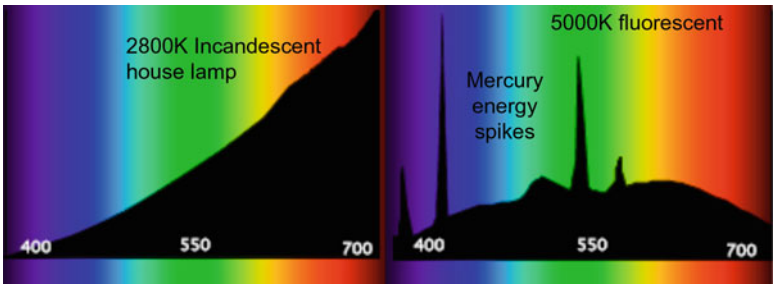


Fig. 10.5 Experimental power spectral densities $S(\lambda)$ of the light emitted by an incandescent light at 2800°K (left picture), and a fluorescent lamp at 5000°K (right picture). The horizontal scale shows the wave length λ in nanometers

9 Figure 10.5 shows experimental power spectral densities $S(\lambda)$ of the light emitted by an incandescent lamp at 2800°K (on the left) and a fluorescent lamp at 5000°K (on the right). The horizontal scale shows the wave length λ in nanometers.

(a) Produce an approximate mathematical formulas for $S(\lambda)$ representing the above two power spectral densities. Assume an arbitrary vertical scale of the experimental spectra (say, from 0 to 1). Plot them on top of pictures in Fig. 10.5 to verify your fit.

(b) Remembering the relationship $f[1/s] \cdot \lambda[m] = c[m/s]$, between the frequency, wave length, and the speed of the traveling wave, and knowing the speed of light, $c[m/s] = 3.0 \cdot 10^8$, convert the approximate mathematical formulas from part (a) to formulas representing the two spectral densities as functions of the frequency f. Plot them.

(c) Use the numerical algorithm from Sect. 10.7 to produce several sample paths of stationary signals with the power spectral densities from (b). Start with the white noise having the $N(0, 1)$ distributions. Plot them.

(d) Do the literature search to comment on whether the selection of the Gaussian distribution in (c) was appropriate for the physical phenomenon under consideration.

Chapter 11
Prediction Theory for Stationary Random Signals

Abstract Prediction (or forecasting) of future values of the stationary random signals based on the known past depends on the functional analytic tools from Hilbert spaces. Essentially, the optimal predictor is an orthogonal projection of the future values of the signal onto the space spanned by the past values. The chapter presents the relevant Wold decomposition theorem, and an application of the Spectral Representation to the solution of the optimal prediction problem.

11.1 The Wold Decomposition Theorem and Optimal Predictors

In this chapter we will consider prediction problems for discrete time weakly stationary random signals $(X_n), n = \ldots, -2, -1, 0, 1, 2, \ldots$. The assumption is that the second moments are finite, the mean value $\mathbf{E}X_n = 0$, and the span of the "past" of the process in the Hilbert space L_2 will be denoted

$$\mathcal{M}_0 = \overline{\text{span}}\{X_n, n \leq 0\}$$

The optimal predictor $\hat{X}_m$ of the values of the process at time $m > 0$ (in the future) based on the knowledge of the past of the process is, obviously, the orthogonal projection

$$\hat{X}_m := \text{Pred}_0 X_m = \text{Proj}_{\mathcal{M}_0} X_m.$$

In what follows we shall also need the special notation for the following spaces:

$$\mathcal{M}_n = \overline{\text{span}}\{X_k, k \leq n\}, \qquad \mathcal{M}_{-\infty} = \bigcap_n \mathcal{M}_n.$$

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7_11

We also need to distinguish between two important categories of time series (X_n):

Definition 11.1.1

(a) The process (X_n) is said to be *deterministic* (or, *singular*) if $\mathcal{M}_{-\infty} = \mathcal{M}_{+\infty}$, or, equivalently, in view of the stationarity assumption, if $\mathcal{M}_k = \mathcal{M}_{k+1}$ for all k. In this case the perfect linear prediction is possible because the error

$$\mathbf{E}(\hat{X}_m - X_m)^2 = \|\hat{X}_m - X_m\|_2^2 = 0.$$

(b) The process (X_n) is said to be *regular* if $\mathcal{M}_{-\infty} = \{0\}$. In this case

$$\mathbf{E}(\hat{X}_m - X_m)^2 = \|\hat{X}_m - X_m\|_2^2 > 0.$$

In general,

$$\{0\} \neq \mathcal{M}_{-\infty} \neq \mathcal{M}_{+\infty},$$

so the process is neither deterministic nor regular. However, nondeterministic processes can be decomposed into a regular and deterministic part:

Wold's Decomposition Theorem *If the process (X_n) is regular, then*

$$X_n = Z_n + Y_n, \qquad n = \ldots, -2, -1, 0, 1, 2, \ldots,$$

where (Z_n) is regular, and (Y_n) is deterministic, and, moreover, the two components are orthogonal to each other,

$$(Z_n) \perp (Y_n).$$

The regular process (Z_n) can be expressed in the form

$$Z_n = \sum_{k=0}^{\infty} \gamma_k W_{n-k},$$

where both (W_n) and (Y_n) have zero mean, (W_n) form an uncorrelated sequence with constant variance σ^2, $\gamma_0 = 0$, and $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$. The decomposition is unique.

Proof Let

$$W_k = X_k - \hat{X}_k, \quad k = n, n-1, n-2, \ldots$$

Since $W_k \perp \mathcal{M}_{k-1}$, we see right away that the sequence (W_k) is uncorrelated, that is $\mathbf{E}W_k W_l = 0$, for $l < k$. Define the coefficients γ_k as follows:

$$\gamma_k = \frac{\mathbf{E}X_n W_{n-k}}{\sigma^2}, \quad k = 1, 2, \ldots.$$

Now, we have the obvious inequality

$$0 \leq \mathbf{E}\left(X_n - \sum_{k=0}^{m} \gamma_k W_{n-k}\right)^2 = \mathbf{E}X_n^2 - \sigma^2 \sum_{k=0}^{m} \gamma_k^2,$$

which implies that $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$, and that $\sum_{k=0}^{\infty} \gamma_k W_{n-k}$ converges in L^2 to a random quantity in the subspace spanned by the sequence $W_n, W_{n-1}, W_{n-2}, \ldots$.

Now, the sequence (Y_n) can be defined by the equality,

$$Y_n = X_n - \sum_{k=0}^{\infty} \gamma_k W_{n-k},$$

so that

$$\mathbf{E}Y_n W_l = \mathbf{E}X_n W_l - \sigma^2 \gamma_{n-l} = 0, \quad \text{for,} \quad l \leq n,$$

and $\mathbf{E}Y_n W_l = 0$, for $l > n$, because W_l orthogonal to the subspace $\mathcal{M}_n \ni Y_n$. Therefore $W_n \in \mathcal{M}_{n-1}$, and by induction, $W_n \in \mathcal{M}_k$, for all $k \leq n$, so that

$$\mathcal{M}_{-\infty} = \bigcap_{k=0}^{\infty} \mathcal{M}_{n-s}.$$

To finish the proof of the theorem let us make two observations.

(i) If $\mathcal{M}_W^{\perp}$ is the subspace orthogonal to $\mathcal{M}_W^{\perp}$, the subspace spanned by (W_n), then $\mathcal{M}_{-\infty} = \mathcal{M}_W^{\perp}$. Indeed, if $Xin\mathcal{M}_{-\infty}$, then $X \in \mathcal{M}_n$, and is orthogonal to W_{n+1}, for every n. Hence, $X \in \mathcal{M}_W^{\perp}$. Conversely, if $X \in \mathcal{M}_W^{\perp}$, then $X \in \mathcal{M}_n$, for some n. Since $X \perp W_n$ we have $X \in \mathcal{M}_{n-1}$, and, by induction, $X \in \mathcal{M}_k$, for all $k \leq n$. Moreover, $X \in \mathcal{M}_k$, for $k > n$, because $\mathcal{M}_n \subset \mathcal{M}_k$. So the first observation is verified.

(ii) Since $Z_n = \sum_{k=0}^{\infty} \gamma_k W_{n-k}$, the subspace $\mathcal{M}_n^Z$ spanned by $Z_n, Z_{n-1}, \ldots,$ is contained in the subspace $\mathcal{M}_n^W$ spanned by $W_n, W_{n-1}, \ldots$. Conversely, if $W_n \in \mathcal{M}_n = \mathcal{M}_n^Z \oplus \mathcal{M}_n^Y$, and $W_n \perp \mathcal{M}_n^Y$, then $W_n \in \mathcal{M}_n^Z$. So $\mathcal{M}_n^W = \mathcal{M}_n^Z$.

Now we are ready to complete the proof of the Decomposition Theorem. Since, for every n, $Y_n \in \mathcal{M}_{-\infty} =\supseteq \mathcal{M}_n^Y$, the condition $X \in \mathcal{M}_{-\infty}$ implies that $X \in \mathcal{M}_n$ because $X \perp \mathcal{M}_n^W = \mathcal{M}_n^Z$. Thus $X \in \mathcal{M}_n Y$. This proves that $|calM_n^Y = \mathcal{M}_{-\infty}$, and the sequence (Y_n) is deterministic.

Now, since $Z_n = W_n + \sum_{k=1}^{\infty} \gamma_k W_{n-k}$, and $W_n \perp \sum_{k=1}^{\infty} \gamma_k W_{n-k} \in \mathcal{M}_{n-1}^W$, the error $\mathbf{E}(Z_n - \hat{Z}_n)^2 = \sigma^2 > 0$, so that the sequence (Z_n) is regular. ∎

Since

$$X_n = Z_n + Y_n = \sum_{k=0}^{\infty} \gamma_k W_{n-k} + Y_n = W_n + \sum_{k=1}^{\infty} \gamma_k W_{n-k} + Y_n,$$

and

$$W_n \perp \sum_{k=1}^{\infty} \gamma_k W_{n-k} + Y_n,$$

the best predictor for X_n is the orthogonal projection of X_n onto $\mathcal{M}_{n-1}$, which is

$$\hat{X}_n = \sum_{k=1}^{\infty} \gamma_k W_{n-k} + Y_n.$$

The square of its error

$$\|X_n - \hat{X}_n\|_{L^2}^2 = \mathbf{E}(X_n - \hat{X}_n)^2 = \mathbf{E}W_n^2 = \sigma^2,$$

because $\gamma_0 = 1$.

11.2 Application of the Spectral Representation to the Solution of the Prediction Problem

In this section we will consider the case of discrete time stationary signal $X(n)$, and assume that $\mathbf{E}X(n) = 0$. The spectral representation theorem of Sect. 10.4 gives rise to a linear isometry

$$L^2([0,1], dC_{\mathcal{W}}) \ni g \longrightarrow \int_0^1 g(f) d\mathcal{W}(f) \in L^2(\Omega, \mathcal{F}, P),$$

which simply extends the representation,

$$X(n) = \int_0^1 e^{j2\pi nf} d\mathcal{W}(f),$$

where the cumulative control function

$$C_{\mathcal{W}}(f) = \mathbf{E}[\mathcal{W}(f)]^2 = \mathcal{S}_X(f),$$

where $\mathcal{S}_X(f)$ is the cumulative spectral function of the process $X(n)$. Obviously, in the particular case $g(f) = e^{j2\pi nf}$ the isometry is the mapping,

$$e^{j2\pi nf} \longrightarrow X(n).$$

So, the optimal prediction of the value of the signal at the future time $m > 0$, based on the past values $X(n)$, $n \leq 0$, is reduced to finding the function,

$$g(f) \in \overline{\text{span}}_{L^2(dS)}(e^{j2\pi nf}, n \leq 0),$$

such that the error of the prediction is minimal, that is

$$\|e^{j2\pi mf} - g(f)\|_{L^2(dS)} = \min_h \|e^{j2\pi mf} - h(f)\|_{L^2(dS)},$$

where $h \in \text{span}_{L^2(d\mathcal{S})}(e^{j2\pi nf}, n \leq 0)$. Or, equivalently, the optimal choice of g has to be an orthogonal projection in L^2, that is

$$e^{j2\pi mf} - g(f) \perp \overline{\text{span}}_{L^2(d\mathcal{S})}(e^{j2\pi nf}, n \leq 0),$$

that is

$$\int_0^1 \left[e^{j2\pi mf} - g(f)\right] e^{-j2\pi nf} d\mathcal{S}(f) = 0, \quad \text{for} \quad n = 0, -1, -2, \ldots.$$

Remark 11.2.1 Observe that if the cumulative spectral function $\mathcal{S}_X(f)$ does not increase (or, its spectral density $S_X(f) = 0$) over the interval $[a, b] \subset [0, 1]$ of length greater than 1/2, then the signal $X(n)$ is singular.

Indeed, let $e^{-j2\pi f}$ be in the arc of the unit circle in the complex plane corresponding to $f \ni [a, b]$, and let $e^{j2\pi f_0}$ be the midpoint of the arc. Then, for large enough N,

$$\left| e^{j2\pi f_0} - \frac{e^{-j2\pi f}}{N} \right| < 1,$$

because of the above length assumption, so, also,

$$\left| 1 - \frac{e^{-j2\pi f}}{N e^{j2\pi f_0}} \right| < \frac{1}{|e^{j2\pi f_0}|} = 1.$$

Hence, we get the following uniformly convergent expansion on the complement of the interval $[a, b]$:

$$e^{j2\pi f} = \frac{1}{e^{-j2\pi f}} = \frac{1}{Ne_0^{j2\pi f}} \cdot \frac{1}{e^{-j2\pi f}/(Ne^{j\pi f_0})}$$

$$= \frac{1}{Ne^{j2\pi f_0}} \cdot \frac{1}{1-(1-e^{-j2\pi f}/(Ne^{j2\pi f_0}))}$$

$$\frac{1}{Ne^{j2\pi f_0}} \cdot \sum_{n=0}^{\infty}(1-e^{-jf2\pi}/(Ne^{j2\pi f_0}))^n \in \overline{\text{span}}_{L^2(dS_X)}(e^{j2\pi fn}, n \le 0) = \mathcal{M}_0,$$

which completes the justification of the above statement. On the other hand, on the set $[a, b]$, where the spectral density is 0, the approximation is trivial.

Remark 11.2.2 It turns out that the Wold decomposition is equivalent to decomposition of the spectral measure into the absolutely continuous (with density) and singular components[1]

In the reminder of this section we will just consider the absolutely continuous case when

$$\mathcal{S}(f) = S(f)df$$

with the spectral density $S(f)$ satisfying the condition,

$$0 < C_1 \le S(f) \le C_2 < \infty, \tag{11.2.1}$$

in which case $L^2(d\mathcal{S}) = L^2(df)$ and the convergences in those two spaces are equivalent.

In this case the best predictor $g(f)$ satisfies the following two conditions:

$$\int_0^1 [e^{j2\pi mf} - g(f)]S(f)e^{-j2\pi nf}df = 0, \quad \text{for} \quad n \le 0, \tag{11.2.2}$$

and

$$[e^{j2\pi mf} - g(f)]S(f) \in \overline{\text{span}}_{L^2(S(f)df)}(e^{j2\pi nf}, n \ge 0) \equiv \mathcal{M}_{>0}. \tag{11.2.3}$$

Now, assume that we can factor the spectral density,

$$S(f) = S_1(f) \cdot S_1^*(f),$$

[1]For more details see, U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, Almqvist and Wiksell, Stockholm 1956, and P. Bremaud, *Fourier Analysis and Stochastic Processes*, Springer 2014.

with both

$$S_1(f),\, S_1^{-1}(f) \in \text{span}_{\mathcal{C}}(e^{j2\pi nf},\, n \le 0) =: \mathcal{C}_{\le 0}$$

where $\mathcal{C}$ denotes the space of continuous functions. Then the condition (11.2.3) can be rewritten in the form

$$[e^{j2\pi mf} - g(f)]S_1(f)S_1^*(f) \in \overline{\text{span}}_{L^2(S(f)df)}(e^{j2\pi nf},\, n \ge 0) \equiv \mathcal{M}_{>0}. \tag{11.2.4}$$

with

$$(S_1^{-1}(f))^* \in \text{span}_{\mathcal{C}}(e^{j2\pi nf},\, n \ge 0).$$

Hence,

$$h(f) := [e^{j2\pi mf} - g(f)]S_1(f) \in \overline{\text{span}}_{L^2(S(f)df)}(e^{j2\pi nf},\, n > 0),$$

and the condition for the best linear prediction can be reformulated as follows:

$$e^{j2\pi mf}S_1(f) = g(f)S_1(f) + h(f), \qquad g \in \mathcal{M}_{\le 0},\ h \in \mathcal{M}_{>0}. \tag{11.2.5}$$

Since $S_1,\, S_1^{-1} \in \mathcal{C}_{\le 0}$,

$$g \in \mathcal{M}_{\le 0} \iff gS_1 \in \mathcal{M}_{\le 0},$$

so, what needs to be done at this point is to split the Fourier series of $e^{j2\pi mf}S_1(f)$ into the $\mathcal{M}_{\le 0}$, and $\mathcal{M}_{>0}$ parts.

Given the expansion

$$S_1(f) = c_0 + c_{-1}e^{-j2\pi f} + c_{-2}e^{-j2\pi 2f} + \ldots$$

we can write (11.2.5) with

$$h(f) = c_0 e^{j2\pi mf} + c_{-1}e^{k2\pi(m-1)f} + \ldots + c_{-m+1}e^{j2\pi f},$$

and

$$g(f)S_1(f) = c_{-m} + c_{-m-1}e^{-j2\pi f} + c_{-m-2}e^{-j2\pi 2f} + \ldots.$$

Hence,

$$g(f) = \left[c_{-m} + c_{-m-1}e^{-j2\pi f} + c_{-m-2}e^{-j2\pi 2f} + \ldots\right] \cdot S_1^{-1}(f),$$

which expands as follows:

$$g(f) = b_0 + b_{-1}e^{-j2\pi f} + b_{-2}e^{-j2\pi 2f} + \ldots,$$

with the predictor

$$\hat{X}_m = b_0 X_0 + b_{-1} X_{-1} + b_{-2} X_{-2} + \dots$$

The prediction error can then be calculated as follows:

$$\|\hat{X}_m - X_m\|^2_{L^2(S(f)df)} = \int_0^1 |e^{j2\pi mf} - g(f)|^2 S(f) df$$

$$= \int_0^1 |e^{j2\pi mf} - g(f) S_1(f)|^2 df = \int_0^1 |h(f)|^2 df \tag{11.2.6}$$

$$= \int_0^1 |c_0 e^{j2\pi mf} + c_{-1} e^{k2\pi(m-1)f} + \dots + c_{-m+1} e^{j2\pi f}|^2 df = |c_0|^2 + \dots |c_{-m+1}|^2.$$

When $m \to \infty$,

$$\sum_{n=0}^{\infty} |c_{-n}|^2 = \int_0^1 |S_1(f)|^2 df = \int_0^1 S(f) df = \mathbf{E}|X_k|^2, \qquad \forall k,$$

so that the signal $(X(k))$ is regular.

Remark 11.2.3 Let us take a look at the one step predictor $\hat{X}_1$ in the case $\log S(f)$ satisfies some smoothness conditions to permit the following expansion of its logarithm, $\log S(f)$:

$$\left(\dots + a_{-2} e^{-j2\pi 2f} + a_1 e^{-j2\pi} + \frac{a_0}{2}\right) + \left(\frac{a_0}{2} + a_1 e^{+j2\pi} + a_2 e^{+j2\pi 2f} + \dots\right).$$

Substituting

$$S_1(f) = \exp\left(\dots + a_{-2} e^{-j2\pi 2f} + a_1 e^{-j2\pi} + \frac{a_0}{2}\right),$$

we see that both S_1 and S_1^{-1} are functions from $\mathcal{C}_{\leq 0}$. Using the standard expansion $e^z = 1 + z + z^2/2 + \dots$, one obtains the equality

$$c_0 = 1 \frac{a_0}{2} + \frac{(a_0/2}{2!} + \dots = e^{a_0/2}.$$

Hence, the one step error

$$\|\hat{X}_1 - X_1\|^2_{L^2(S(f)df)} = |c_0|^2 = e^{a_0} = \exp\left(\int_0^1 \log S(f) df\right).$$

Notice that, in general, this error is nonzero if, and only if,

$$\int_0^1 \log S(f)df > -\infty,$$

which is the general condition for the regularity of the random stationary signal X_n.[2]

11.3 Examples of Linear Prediction for Stationary Time Series

In this section we will consider a simple example of stationary time series where the calculation of the optimal predictor is not very difficult.

Let $X(t)$ be a stationary time series, $t = \ldots, -1, 0, 1, \ldots$, with the autocovariance function

$$\gamma_X(t) = a^{|t|}, \qquad -1 < a < 1.$$

The corresponding spectral density, assuming the representation $\gamma_X(t) = \int S_X(f)e^{-jft}df$, is

$$S_x(f) = \frac{1 - a^2}{2\pi(e^{jf} - a)(e^{-jf} - a)},$$

which can be rewritten in the form

$$S_X(f) = \hat{S}_X(e^{jf}),$$

where

$$\hat{S}_X(z) = \frac{(1 - a^2)z}{2\pi(z - a)(1 - az)}.$$

Finding the optimal predictor m steps ahead requires finding a function

$$\Phi_m(f) = a_1e^{-jf} + a_2e^{-j2f} + a_3e^{-j3f},$$

satisfying the condition

$$\int_{-\pi}^{\pi} e^{jkf}[e^{jmf} - \Phi_m(f)]S_X(f)df = 0, \qquad k = 1, 2, 3, \ldots$$

[2] Again, see, Grenander and Rosenblatt, and Bremaud's books cited on page 284, for more details.

In other words, the Fourier expansion of the function

$$\Psi_m(f) = [e^{jmf} - \Phi_m(f)]S_X(f) = \sum_{k=0}^{\infty} c_k e^{jkf}$$

contains only nonnegative powers of e^{jf}.

In the case of rational $\hat{S}_X(z)$, the function

$$\hat{\Phi}_m(z) = \sum_{k=1}^{\infty} a_k z^{-k}$$

is an analytic function of z for $|z| \geq 1$, with $\hat{\Phi}_m(\infty) = 0$, and

$$\hat{\Psi}_m(z) = [z^m - \hat{\Psi}_m(z)]\hat{S}_X(z),$$

is analytic for $|z| \leq 1$.

So, if in our case we are attempting to make a prediction one time step ahead, that is, assuming $m = 0$, we need to find a function $\hat{\Phi}_0(z)$ with no singularities for $|z| \geq 1$, vanishing at infinity, and such that the function

$$\hat{\Psi}_0(z) = \frac{(1-a^2)[1 - \hat{\Phi}_0(z)]]z}{2\pi(z-a)(1-az)}$$

has no singularities for $|z| \leq 1$. Since $|a| < 1$ we must have $\hat{\Phi}_0(a) = 1$. The above formula implies that $\hat{\Phi}_0(z)$ has no singularities other that a simple pole at $z = 0$. Thus,

$$\hat{\Phi}_0(z) = g_0(z)z^{-1},$$

where $g_0(z)$ is analytic in the whole complex plane, and $g_0(a) = a$. So the only function satisfying the above conditions is

$$\hat{\Phi}_0(a) = az^{-1}, \qquad with \qquad \Phi(f) = ae^{-jf}.$$

Therefore the optimal predictor for $X(t)$ is $aX(t-1)$. So, in this case the best predictor just depends on the value of the process one step back and does not depend on the whole past of the process.[3]

[3]For more details and analysis of more complicated rational spectral densities see *An Introduction to the Theory of Random Stationary Functions*, by A.M. Yaglom, Dover Publications. New York, 1973.

11.4 Problems and Exercises

1 Verify that in the case considered in Remark 11.2.1 the best predictor one time-step ahead $\hat{X}_1$ is expressed by the formula

$$\hat{X}_1 = \sum_{n=0}^{\infty} (Ne^{jf_0})^{-n-1} \sum_{k=0}^{n} \binom{n}{k} (e^{jf_0})^{n-k} (-1)^k X_{-k}.$$

2 Prove that if the spectral density $S(f)$ is satisfying the condition (11.2.1),

$$0 < C_1 \leq S(f) \leq C_2 < \infty,$$

then $L^2(dS) = L^2(df)$, and the convergences in those two spaces are equivalent.

3 Show that in the case analyzed in Sect. 11.3 the optimal prediction m time steps ahead, that is at time $t + m$, also depends only on the single value of the process in the past and is of the form

$$a^{m+1} X(t-1).$$

4 Show that in the case of the spectral density of the form

$$S_x(f) = \frac{1}{|e^{jf} - a_1|^2 |e^{jf} - a_2|^2}, \qquad |a_1|, |a_2| < 1,$$

the optimal prediction one time step ahead depends only on the two values of the process in the past, and is of the form

$$(a_1 + a_2) X(t-1) + a_1 a_2 X(t-2).$$

Solutions to Selected Problems and Exercises

Chapter 1

Problem 1. Find the real and imaginary parts of (a) $(j+3)/(j-3)$ and (b) $(1+j\sqrt{2})^3$.

Solution

(a)

$$\frac{j+3}{j-3} = \frac{(j+3)(-j-3)}{(j-3)(-j-3)} = \frac{1-3j-3j-9}{1^2+3^2} = -\frac{4}{5} - \frac{3}{5}j$$

(b)

$$(1+j\sqrt{2})^3 = 1^3 + 3\cdot 1^2(j\sqrt{2}) + 3\cdot 1(j\sqrt{2})^2 + (j\sqrt{2})^3 = -5 + \sqrt{2}j$$

Problem 2. Find the moduli $|z|$ and arguments θ of complex numbers
(a) $z = -2j$; (b) $z = 3+4j$.

Solution

(a) $\mid z \mid = \sqrt{(-2)^2+0} = 2$, $\tan\theta = \infty \Rightarrow \theta = -\pi/2$ (You have to be careful with the coordinate angle, here $cos\theta = 0$, $\sin\theta < 0$).
(b) $\mid z \mid = \sqrt{9+16} = 5$, $\tan\theta = 4/3 \Rightarrow \theta = \arctan 4/3$.

Problem 3. Find the real and imaginary components of complex numbers (a) $z = 5\,e^{j\pi/4}$; (b) $z = -2\,e^{j(8\pi+1.27)}$.

Solution

(a) $z = 5e^{j\pi/4} = 5\cos(\pi/4) + j\sin(\pi/4) = \frac{5\sqrt{2}}{2} + j\frac{5\sqrt{2}}{2} \Rightarrow Rez = \frac{5\sqrt{2}}{2}$, $Imz = \frac{5\sqrt{2}}{2}$

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7

(b) $z = -2e^{j(8\pi+1.27)} = -2\cos(1.27) - 2j\sin(1.27) \Rightarrow Rez = -2\cos(1.27)$, $Imz = -2\sin(1.27)$

Problem 4. Show that

$$\frac{5}{(1-j)(2-j)(3-j)} = \frac{j}{2}.$$

Solution

$$\frac{5}{(1-j)(2-j)(3-j)} = \frac{5}{(1-3j)(3-j)} = -\frac{5}{10j} = \frac{j}{2}$$

Problem 5. Sketch sets of points in complex plane (x, y), $z = x + jy$, such that (a) $|z - 1 + j| = 1$; (b) $z^2 + (z^*)^2 = 2$.

Solution

(a)

$$\{(x, y) : |z - 1 + j| = 1\} = \{(x, y) : |x + jy - 1 + j| = 1\}$$

$$= \{(x, y) : |(x - 1) + j(y + 1)| = 1\} = \{(x, y) : (x - 1)^2 + (y + 1)^2 = 1^2\}$$

So the set is a circle with radius 1 and center at $(1, -1)$.

(b)

$$\{(x, y) : z^2 + (z^*)^2 = 2\} = \{(x, y) : (x + jy)^2 + (x - jy)^2 = 2\}$$

$$= \{(x, y) : x^2 + 2jxy - y^2 + x^2 - 2jxy - y^2 = 2\} = \{(x, y) : x^2 - y^2 = 1\}$$

So the set is a hyperbola (sketch it, please).

Problem 6. Using de Moivre's formulas find $(-2j)^{1/2}$. Is this complex number uniquely defined?

Solution

$$(-2j)^{1/2} = \sqrt{2}\left(e^{j(\frac{3\pi}{2}+2\pi k)}\right)^{1/2} = \sqrt{2}e^{j(\frac{3\pi}{4}+\pi k)}, \qquad k = 0, 1, 2, \ldots$$

$$= \begin{cases} \sqrt{2}e^{j(\frac{3\pi}{4})}, & \text{for } k = 0, 2, 4, \ldots; \\ \sqrt{2}e^{j(\frac{3\pi}{4}+\pi)}, & \text{for } k = 1, 3, 5, \ldots; \end{cases}$$

$$= \begin{cases} \sqrt{2}\left(\cos(\frac{3\pi}{4}) + j\sin(\frac{3\pi}{4})\right), & \text{for } k = 0, 2, 4, \ldots; \\ \sqrt{2}\left(\cos(\frac{7\pi}{4}) + j\sin(\frac{7\pi}{4})\right), & \text{for } k = 1, 3, 5, \ldots; \end{cases}$$

Problem 10. Using de Moivre's formula derive the complex exponential representation (1.4.5) of the signal $x(t)$ given by the cosine series representation $x(t) = \sum_{m=1}^{M} c_m \cos(2\pi m f_0 t + \theta_m)$.

Solution

$$x(t) = c_0 + \sum_{m=1}^{M} c_m \cos(2\pi m f_0 t + \theta_m)$$

$$= c_0 e^{j2\pi 0 f_0 t} + \sum_{m=1}^{M} c_m \frac{1}{2}\left(e^{j(2\pi m f_0 t + \theta_m)} + e^{-j(2\pi m f_0 t + \theta_m)}\right)$$

$$= c_0 e^{j2\pi 0 f_0 t} + \sum_{m=1}^{M} \frac{c_m}{2} e^{j(2\pi m f_0 t + \theta_m)} + \sum_{m=1}^{M} \frac{c_m}{2} e^{-j(2\pi m f_0 t + \theta_m)}$$

$$= \sum_{m=-1}^{-M} \left(\frac{c_{-m}}{2} e^{-j\theta_{-m}}\right) e^{j2\pi m f_0 t} + c_0 e^{j2\pi 0 f_0 t} + \sum_{m=1}^{M} \left(\frac{c_m}{2} e^{j\theta_m}\right) e^{j2\pi m f_0 t}$$

Problem 12. Using a computing platform such as *Mathematica*, *Maple*, or *Matlab* produce plots of the signals

$$x_M(t) = \frac{\pi}{4} + \sum_{m=1}^{M} \left[\frac{(-1)^m - 1}{\pi m^2} \cos mt - \frac{(-1)^m}{m} \sin mt\right],$$

for $M = 0, 1, 2, 3, \ldots, 9$, and $-2\pi < t < 2\pi$. Then produce their plots in the frequency-domain representation. Calculate their power (again, using *Mathematica*, *Maple*, or *Matlab*, if you wish). Produce plots showing how power is distributed over different frequencies for each of them. Write down your observations. What is likely to happen with the plots of these signals as we take more and more terms of the above series, that is, as $M \to \infty$? Is there a limit signal $x_\infty(t) = \lim_{M\to\infty} x_M(t)$? What could it be?

Partial Solution Sample *Mathematica* code for the plot:

```
M = 9;

Plot[
Sum[
    (((-1)^m - 1)/(Pi*m^2))*Cos[m*t]
              - (((-1)^m)/m)*Sin[m*t],
          {m, M}],
            {t, -2*Pi, 2*Pi}]
```

Sample power calculation:

```
M2 = 2;

N[Integrate[(1/(2*Pi))*
   Abs[Pi/4 +
      Sum[(((-1)^m - 1)/(Pi*m^2))*Cos[m*u]
        - (((-1)^m)/m)*
         Sin[m*u], {m, M2}]]^2, {u, 0, 2*Pi}], 5]

1.4445
```

Problem 13. Use the analog-to-digital conversion formula (1.1.1) to digitize signals from Problem 13 for a variety of sampling periods and resolutions. Plot the results.

Solution We provide a sample *Mathematica* code:

```
M=9;
x[t_]:=Sum[(((-1)^m-1)/(Pi*m^2))*Cos[m*t]
         -(((-1)^m)/m)*Sin[m*t],{m,M}]
T=0.1;
R=0.05;

xDigital=Table[R*Floor[x[m T]/R],{m,1,50}];

ListPlot[xDigital]
```

Problem 14. Use your computing platform to produce a discrete-time signal consisting of a string of random numbers uniformly distributed on the interval [0,1]. For example, in *Mathematica*, the command

```
Table[Random[], {20}]
```

will produce the following string of 20 random numbers between 0 and 1:

```
{0.175245, 0.552172, 0.471142, 0.910891, 0.219577,
0.198173, 0.667358, 0.226071, 0.151935, 0.42048,
0.264864, 0.330096, 0.346093, 0.673217, 0.409135,
0.265374, 0.732021, 0.887106, 0.697428, 0.7723}
```

Use the "random numbers" string as additive noise to produce random versions of the digitized signals from Problem 14. Follow the example described in Fig. 1.3. Experiment with different string length and various noise amplitudes. Then center the noise around zero and repeat your experiments.

Solution We provide a sample *Mathematica* code:

```
M=9;
x[t_]:=Sum[(((-1)^m-1)/(Pi*m^2))*Cos[m*t]
         -(((-1)^m)/m)*Sin[m*t],{m,M}]
```

```
T=0.1;
R=0.05;
xDigital=Table[R*Floor[x[m T]/R],{m,1,50}];
ListPlot[xDigital]
Noise=Table[Random[],{50}];
noisysig = Table[Noise[[t]] + xDigital[[t]],
{t, 1, 50}];
ListPlot[noisysig]
Centernoise = Table[Random[] - 0.5, {50}];
noisysig1 = Table[Centernoise[[t]] + xDigital[[t]],
{t, 1, 50}];
ListPlot[noisysig1]
```

Chapter 2

Problem 1. Prove that the system of real harmonic oscillations

$$\sin(2\pi m f_0 t), \quad \cos(2\pi m f_0 t), \quad m = 1, 2, \ldots,$$

forms an orthogonal system. Is the system normalized? Is the system complete? Use the above information to derive formulas for coefficients in the Fourier expansions in terms of sines and cosines. Model this derivation on calculations in Sect. 2.1.

Solution First of all, we have to compute the scalar products:

$$\frac{1}{P}\int_0^P \sin(2\pi m f_0 t)\cos(2\pi n f_0 t)dt,$$

$$\frac{1}{P}\int_0^P \sin(2\pi m f_0 t)\sin(2\pi n f_0 t)dt,$$

$$\frac{1}{P}\int_0^P \cos(2\pi m f_0 t)\cos(2\pi n f_0 t)dt.$$

Using the trigonometric formulas listed in Sect. 1.2 we obtain

$$\begin{aligned}&\frac{1}{P}\int_0^P \sin(2\pi m t/P)\cos(2\pi n t/P)\,dt\\ =\ &\frac{1}{2P}\int_0^P \left(\sin(2\pi(m-n)t/P)+\sin(2\pi(m+n)t/P)\right)\,dt = 0 \ \text{ for all } m, n;\end{aligned}$$

$$\frac{1}{P}\int_0^P \cos(2\pi mt/P)\cos(2\pi nt/P)\,dt$$
$$= \frac{1}{2P}\int_0^P \left(\cos(2\pi(m-n)t/P) + \cos(2\pi(m+n)t/P)\right)\,dt$$
$$= \begin{cases} \frac{1}{2}, & \text{if } m = n; \\ 0, & \text{if } m \neq n. \end{cases}$$

Similarly,

$$\frac{1}{P}\int_0^P \sin(2\pi mt/P)\sin(2\pi nt/P)\,dt = \begin{cases} \frac{1}{2}, & m = n; \\ 0, & m \neq n. \end{cases}$$

Therefore we conclude that the given system is orthogonal but not normalized. It can be normalized by multiplying each sine and cosine by $1/\sqrt{2}$. It is not complete but it becomes complete if we add the function identically equal to 1 to it; it is obviously orthogonal to all the sines and cosines.

Using the orthogonality property of the above real trigonometric system we arrive at the following Fourier expansion for a periodic signal $x(t)$:

$$x(t) = a_0 + \sum_{m=1}^{\infty}[a_m \cos(2\pi m f_0 t) + b_m \sin(2\pi m f_0 t)],$$

with coefficients

$$a_0 = \frac{1}{P}\int_0^P x(t)\,dt$$
$$a_m = \frac{2}{P}\int_0^P x(t)\cos(2\pi mt/P)\,dt,$$
$$b_m = \frac{2}{P}\int_0^P x(t)\sin(2\pi mt/P)\,dt.$$

for $m = 1, 2, \ldots$

2. Using the results from Problem 1 find formulas for amplitudes c_m and phases θ_m in the expansion of a periodic signal $x(t)$ in terms of only cosines, $x(t) = \sum_{m=0}^{\infty} c_m \cos(2\pi m f_0 t + \theta_m)$.

Solution Obviously, $c_0 = a_0$. To find the connection between a_m, b_m and c_m, and θ_m we have to solve the following system:

$$a_m = c_m \cos\theta_m, \qquad b_m = -c_m \sin\theta_m.$$

This gives us:

$$\theta_m = \arctan\left(-\frac{b_m}{a_m}\right), \qquad c_m = \sqrt{a_m^2 + b_m^2}.$$

Problem 9. Find the complex and real Fourier series for the periodic signal $x(t) = |\sin t|$. Produce graphs comparing the signal $x(t)$ and its finite Fourier sums of order 1, 3, and 6.

Solution The first observation is that $x(t)$ has period π. So,

$$z_m = \frac{1}{\pi}\int_0^\pi |\sin t| e^{-2jmt}\, dt = \frac{1}{\pi}\int_0^\pi \sin t \cdot e^{-2jmt}\, dt$$

$$= \frac{1}{\pi}\int_0^\pi \frac{e^{jt} - e^{-jt}}{2j}\cdot e^{-2jmt}\, dt = \frac{1}{2j\pi}\int_0^\pi (e^{jt(1-2m)} - e^{-jt(1+2m)})\, dt$$

$$= \frac{1}{2j\pi}\left(\frac{e^{j\pi(1-2m)} - 1}{j(1-2m)} - \frac{e^{-j\pi(1+2m)}}{-j(1+2m)}\right) dt = \frac{2}{\pi(1-4m^2)},$$

because $e^{j\pi} = e^{-j\pi} = -1$, and $e^{-2jm\pi} = 1$, for all m. Therefore, the sought complex Fourier expansion is

$$x(t) = \frac{2}{\pi}\sum_{m=-\infty}^{\infty} \frac{1}{1-4m^2}\cdot e^{j2mt}.$$

We observe that for any $m = \ldots -1, 0, 1, \ldots$ we have $z_{-m} = z_m$. Pairing up complex exponentials with the exponents of opposite signs, and using de Moivre's formula, we arrive at the real Fourier expansion that contains only cosine functions:

$$x(t) = \frac{2}{\pi}\left(1 + 2\sum_{m=1}^{\infty}\frac{\cos(2mt)}{1-4m^2}\right).$$

In particular, the partial sums of order 1, and 3, are

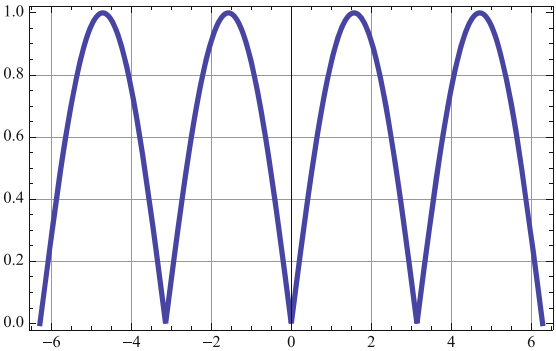
$$s_1(t) = \frac{2}{\pi}\left(1 - \frac{2\cos 2t}{3}\right)$$

$$s_3(t) = \frac{2}{\pi}\left(1 - \frac{2\cos 2t}{3} - \frac{2\cos 4t}{15} - \frac{2\cos 6t}{35}\right).$$

A *Mathematica* code and the output showing $x(t)$, $s_1(t)$, and $s_6(t)$ is shown on the next page.

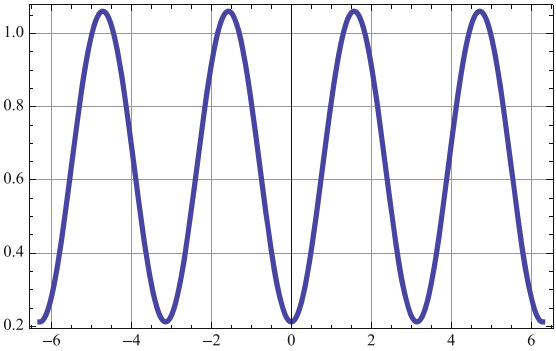
```
x[t_] := Abs[Sin[t]]

p1 = Plot[x[t], {t, -2 * Pi, 2 * Pi}, Frame -> True,
   GridLines -> Automatic, PlotStyle -> {Thickness[0.01]}]
```

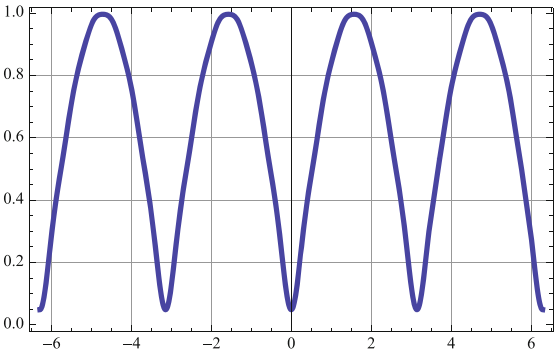



```
sum[t_, M_] := (2 / Pi) * (1 + Sum[(2 / (1 - 4 * m^2)) * Cos[2 * m * t], {m, 1, M}])

s1 = Plot[ sum[t, 1], {t, -2 * Pi, 2 * Pi}, Frame -> True,
   GridLines -> Automatic, PlotStyle -> {Thickness[0.01]}]
```




```
s3 = Plot[ sum[t, 6], {t, -2 * Pi, 2 * Pi}, Frame -> True,
   GridLines -> Automatic, PlotStyle -> {Thickness[0.01]}]
```



Problem 13.

(a) The nonperiodic signal $x(t)$ is defined as equal to 1/2 on the interval $[-1, +1]$, and 0 elsewhere. Plot it and calculate its Fourier transform $X(f)$. Plot the latter.
(b) The nonperiodic signal $y(t)$ is defined as equal to $(t + 2)/4$ on the interval $[-2, 0]$, $(-t + 2)/4$ on the interval $[0, 2]$, and 0 elsewhere. Plot it and calculate its Fourier transform $Y(f)$. Plot the latter.
(c) Compare the Fourier transforms $X(f)$ and $Y(f)$. What conclusion do you draw about the relationship of the original signals $x(t)$ and $y(t)$?

Solution

(a) The Fourier transform of $x(t)$ is

$$X(f) = \int_{-1}^{+1} \frac{1}{2} e^{-j2\pi ft}\, dt = \frac{e^{-j2\pi f} - e^{j2\pi f}}{-4j\pi f} = \frac{\sin 2\pi f}{2\pi f}.$$

(b) Integrating by parts, the Fourier transform of $y(t)$ is

$$\begin{aligned} Y(f) &= \int_{-2}^{0} ((t+2)/4) e^{-j2\pi ft}\, dt + \int_{0}^{+2} ((-t+2)/4) e^{-j2\pi ft}\, dt \\ &= \frac{1}{4}\left(\frac{1}{-j2\pi f} \cdot 2 - \frac{1}{(-j2\pi f)^2}(1 - e^{j2\pi f 2}) \right) \\ &\quad + \frac{1}{4}\left(\frac{-1}{-j2\pi f} \cdot 2 - \frac{1}{(-j2\pi f)^2}(e^{-j2\pi f 2} - 1) \right) \\ &= \frac{1}{4} \frac{1}{(-j2\pi f)^2} \left(-(1 - e^{j2\pi f 2}) - (e^{-j2\pi f 2} - 1) \right) \\ &= \frac{1}{4} \frac{1}{(j2\pi f)^2} \left(e^{j2\pi f} - e^{-j2\pi f} \right)^2 = \left(\frac{\sin 2\pi f}{2\pi f} \right)^2 \end{aligned}$$

(c) So we have that $Y(f) = X^2(f)$. This means that the signal $y(t)$ is the convolution of the signal $x(t)$ with itself: $y(t) = (x * x)(t)$.

Problem 18. Utilize the Fourier transform (in the space variable z) to find a solution of the diffusion (heat) partial differential equation

$$\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial z^2},$$

for a function $u(t, z)$ satisfying the initial condition $u(0, z) = \delta(z)$. The solution of the above equation is often used to describe the temporal evolution of the density of a diffusing substance.

Solution Let us denote the Fourier transform (in z) of $u(t, z)$ by

$$U(t, f) = \int_{-\infty}^{\infty} u(t, z) e^{-j2\pi f z}\, dz.$$

Then for the second derivative

$$\frac{\partial^2 u(t, z)}{\partial z^2} \qquad \longmapsto (j2\pi f)^2 U(t, f) = -4\pi^2 f^2 U(t, f).$$

So taking the Fourier transform of both sides of the diffusion equation gives the equation

$$\frac{\partial}{\partial t} U(t, f) = -4\pi^2 f^2 \sigma U(t, f),$$

which is now just an ordinary differential linear equation in the variable t, which has the obvious exponential (in t) solution

$$U(t, f) = C e^{-4\pi^2 f^2 \sigma t},$$

where C is a constant to be matched later to the initial condition $u(0, z) = \delta(z)$. Taking the inverse Fourier transform gives

$$u(t, z) = \frac{1}{\sqrt{4\pi\sigma t}} e^{-\frac{z^2}{4\sigma t}}$$

Indeed, by completing the square,

$$\int_{-\infty}^{\infty} U(t, f) e^{j2\pi f x}\, df = C \int_{-\infty}^{\infty} e^{-4\pi^2 f^2 \sigma t} e^{j2\pi f x}\, df$$

$$= C e^{\frac{-x^2}{4\sigma t}} \int_{-\infty}^{\infty} e^{-4\pi^2 \sigma t (f - jx/(4\pi\sigma))^2}\, df$$

with the last (Gaussian) integral being equal to $1/\sqrt{4\pi\sigma t}$. A verification of the initial condition gives $C = 1$.

Chapter 4

2. Calculate the probability that a random quantity uniformly distributed over the interval $[0, 3]$ takes values between 1 and 3. Do the same calculation for the exponentially distributed random quantity with parameter $\mu = 1.5$, and the Gaussian random quantity with parameters $\mu = 1.5, \sigma^2 = 1$.

Solution

a) Since X has uniform distribution on the interval [0,3], then the value of p.d.f. is 1/3 between 0 and 3, and 0 elsewhere.

$$\mathbf{P}\{1 \le X \le 3\} = (3-1) \cdot 1/3 = 2/3$$

b)

$$\int_1^3 \frac{1}{\mu} e^{-x/\mu} dx = \frac{2}{3} \int_1^3 e^{-2x/3} dx = -1(e^{-2\cdot 3/3} - e^{-2/3}) = 0.378$$

c) We can solve this problem using the table for the c.d.f. of the standard normal random quantity:

$$\mathbf{P}(1 \le X \le 3) = \mathbf{P}(1 - 1.5 \le X - \mu \le 3 - 1.5) = \mathbf{P}(-0.5 \le Z \le 1.5)$$

$$= \Phi(1.5) - \Phi(-0.5) = .9332 - (1 - \Phi(0.5)) = .9332 - 1 + .6915 = .6247$$

4. The p.d.f. of a random variable X is expressed by the quadratic function $f_X(x) = ax(1-x)$, for $0 < x < 1$, and is zero outside the unit interval. Find a from the normalization condition and then calculate $F_X(x)$, $\mathbf{E}X$, $\mathrm{Var}(X)$, $\mathrm{Std}(X)$, the n-th central moment, and $\mathbf{P}(0.4 < X < 0.9)$. Graph $f_X(x)$, and $F_X(x)$.

Solution

a) We know that for the p.d.f. of any random quantity we have

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

So,

$$1 = \int_0^1 ax(1-x)\, dx = \frac{a}{6}.$$

Thus the constant $a = 6$.

b) To find the c.d.f. we will use the definition:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy$$

In our case, when $0 < x < 1$

$$F_X(x) = \int_0^x 6y(1-y)\, dy = x^2(3-2x)$$

Finally,

$$F_X(x) = \begin{cases} 0, & x < 0; \\ x^2(3-2x), & 0 \le x < 1; \\ 1, & x \ge 1. \end{cases}$$

c)

$$\mathbf{E}X = \int_0^1 6x^2(1-x)\,dx = \frac{1}{2}$$

$$\mathrm{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}X)^2 = \int_0^1 6x^3(1-x)\,dx - \frac{1}{4} = \frac{3}{10} - \frac{1}{4} = 0.05$$

$$\mathrm{Std}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{0.05} \approx 0.224$$

d) The n-th central moment:

$$\int_0^1 (x-.5)^n 6x(1-x)\,dx = 6\int_0^1 x(1-x)\sum_{k=0}^{n} x^k \left(-\frac{1}{2}\right)^{n-k} dx$$

$$= 6\sum_{k=0}^{n} \binom{n}{k}\left(-\frac{1}{2}\right)^{n-k} \int_0^1 x^{k+1}(1-x)dx = 6\sum_{k=0}^{n} \binom{n}{k}\left(-\frac{1}{2}\right)^{n-k} \frac{1}{6+5k+k^2}$$

e)

$$\mathbf{P}(0.4 < X < 0.9) = \int_{0.4}^{0.9} 6x(1-x)\,dx = 0.62$$

6. Find the c.d.f and p.d.f. of the random quantity $Y = \tan X$, where X is uniformly distributed over the interval $(-\pi/2, \pi/2)$. Find a physical (geometric) interpretation of this result.

Solution The p.d.f. $f_X(x)$ is equal to $1/\pi$ for $x \in (-\pi/2, \pi/2)$ and 0 elsewhere. So, c.d.f.

$$F_X(x) = \begin{cases} 0, & \text{for } x \le -\pi/2; \\ (1/\pi)(x+\pi/2), & \text{for } x \in (-\pi/2, \pi/2); \\ 1, & \text{for } x \ge \pi/2. \end{cases}$$

Hence,

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(\tan X \le y) = \mathbf{P}(X \le \arctan(y))$$

$$= F_X(\arctan(y)) = \frac{1}{\pi}(\arctan(y) + \pi/2).$$

The p.d.f.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \frac{1}{\pi}(\arctan(y) + \pi/2) = \frac{1}{\pi(1+y^2)}.$$

This p.d.f. is often called Cauchy probability density function.

A physical interpretation: A particle is emitted from the origin of the (x, y)-plane with the uniform distribution of directions in the half-plane, $y > 0$. The p.d.f. of the random quantity Y describes the probability distribution of locations of the particles when they hit the vertical screen located at $x = 1$.

13. A random quantity X has an even p.d.f. $f_X(x)$ of the triangular shape shown in Sect. 4.7.

(a) How many parameters do you need to describe this p.d.f.? Find an explicit analytic formula for p.d.f. $f_X(x)$ and c.d.f. $F_X(x)$. Graph both of them.
(b) Find the expectation and variance of X.
(c) Let $Y = X^3$. Find the p.d.f. $f_Y(y)$ and graph it.

Solution

(a) Notice that the triangle is symmetric about the line $x = 0$. Let us assume that the vertices of the triangle have the following coordinates $A(a, 0)$, $B(-a, 0)$, $C(0, c)$. Then the pdf is represented by the equations $y = -\frac{c}{a}x + c$, in the interval $[0, a]$, and $y = \frac{c}{a}x + c$, in the interval $[-a, 0]$. So, we need at most 2 parameters.

Next, the normalization condition says that area under the pdf is one. So, necessarily $ac = 1 \Rightarrow c = 1/a$. Therefore, actually, one parameter suffices and our one-parameter family of pdf's has the following analytic description:

$$f_X(x) = \begin{cases} 0, & \text{for } x < -a; \\ \frac{x}{a^2} + \frac{1}{a}, & \text{for } -a \le x < 0; \\ -\frac{x}{a^2} + \frac{1}{a}, & \text{for } 0 \le x < a; \\ 0, & \text{for } x \ge a. \end{cases}$$

The corresponding c.d.f is as follows: If $x < -a$, then $F_X(x) = 0$; If $-a \le x < 0$, then $F_X(x) = \int_{-a}^{x} (\frac{t}{a^2} + \frac{1}{a})\, dt = \frac{x^2}{2a^2} + \frac{x}{a} + \frac{1}{2}$; If $0 \le x < a$, then $F_X(x) = \frac{1}{2} + \int_0^x (-\frac{t}{a^2} + \frac{1}{a})\, dt = \frac{1}{2} - \frac{x^2}{2a^2} + \frac{x}{a}$; If $x > a$ then $F(x) = 1$.

(b) Find the expectation and variance of X.

$$EX = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_{-a}^{0} x \left(\frac{x}{a^2} + \frac{1}{a}\right) dx + \int_{0}^{a} x \left(-\frac{x}{a^2} + \frac{1}{a}\right) dx = 0.$$

Of course, the above result can be obtained without any integration by observing that the p.d.f. is an even function, symmetric about the origin.

$$\mathrm{Var}X = \int_{-\infty}^{\infty} x^2\, f_X(x)$$

$$dx = \int_{-a}^{0} x^2 \left(\frac{x}{a^2} + \frac{1}{a}\right) dx + \int_{0}^{a} x^2 \left(-\frac{x}{a^2} + \frac{1}{a}\right) dx = \frac{a^2}{6}.$$

(c) The function $y = g(x) = x^3$ is monotone, therefore there exists an inverse function which in this case is $x = g^{-1}(y) = y^{1/3}$. The derivative $g'(x) = 3x^2$, and $g'(g^{-1}(y) = 3y^{2/3}$. Then, see (4.1.12),

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y)} = \begin{cases} 0, & y < (-a)^3 \\ \left(\frac{y^{1/3}}{a^2} + \frac{1}{a}\right) \frac{1}{3y^{2/3}} & (-a)^3 \le y < 0 \\ \left(-\frac{y^{1/3}}{a^2} + \frac{1}{a}\right) \frac{1}{3y^{2/3}} & 0 < y < a^3 \\ 0 & y \ge a^3 \end{cases}$$

Here is the needed *Mathematica* code producing the desired plots:

```
(*pdf, a=2*)
H[x_] := If[x < 0, 0, 1]
f[a_, b_, x_] := H[x - a] - H[x - b];
ff[x_, a_] := (x/a^2 + 1/a)*f[-a, 0, x] +
  (-x/a^2 + 1/2)*f[0, a, x]
Plot[ff[x, 2], {x, -3, 3}]
```

```
F[x_, a_] := (x^2/(2*a^2) + x/a + 1/2)*f[-a, 0, x] +
  (1/2 - x^2/(2*a^2) + x/a)*f[0, a, x]
Plot[F[x, 2], {x, -4, 4}]
```

15. Verify the Cauchy-Schwartz Inequality (3.3.18). *Hint:* Take $Z = (X - \mathbf{E}X)/\sigma(X)$ and $W = (Y - \mathbf{E}Y/\sigma(Y)$, and consider the discriminant of the expression $\mathbf{E}(Z + xW)^2$. The latter is quadratic in variable x and necessarily always nonnegative, so it can have at most one root.

Solution The quadratic form in x,

$$0 \le \mathbf{E}(Z + xW)^2 = EZ^2 + 2xE(ZW) + x^2 EW^2 = p(x)$$

is nonnegative for any x. Thus the quadratic equation $p(x) = 0$ has at most one solution (root). Therefore the discriminant of this equation must be nonpositive, that is,

$$(2E(ZW))^2 - 4EW^2EZ^2 \leq 0,$$

which gives the basic form of the Cauchy-Schwarz inequality

$$|E(ZW)| \leq \sqrt{EW^2} \cdot \sqrt{EZ^2}.$$

Finally, substitute for Z and W as indicated in the above hint to obtain the desired result.

24. Complete the following sketch of the proof of the Central Limit Theorem from Sect. 3.5. Start with a simplifying observation (based on Problem 23) that it is sufficient to consider random quantities $X_n, n = 1, 2, \ldots$, with expectations equal to 0, and variances 1.

(a) Define $\mathcal{F}_X(u)$ as the inverse Fourier transform of the distribution of X:

$$\mathcal{F}_X(u) = \mathbf{E}e^{juX} = \int_{\infty}^{\infty} e^{jux}\, dF_X(x).$$

Find $\mathcal{F}'_X(0)$ and $\mathcal{F}''_X(0)$. In statistical literature $\mathcal{F}_X(u)$ is called the characteristic function of the random quantity X. Essentially, it completely determines the probability distribution of X via the Fourier transform (inverse of the inverse Fourier transform).

(b) Calculate $\mathcal{F}_X(u)$ for the Gaussian $N(0, 1)$ random quantity. Note the fact that its functional shape is the same as that of the $N(0, 1)$ p.d.f. This fact is the crucial reason for the validity of CLT.

(c) Prove that, for independent random quantities X and Y,

$$\mathcal{F}_{X+Y}(u) = \mathcal{F}_X(u) \cdot \mathcal{F}_Y(u).$$

(d) Utilizing (c), calculate

$$\mathcal{F}_{\sqrt{n}(\bar{X}-\mu_X)/\mathrm{Std(X)}}(u).$$

Then find its limit as $n \to \infty$. Compare it with the characteristic of the Gaussian $N(0, 1)$ random quantity. (Hint: it is easier to work here with the logarithm of the above transform.)

Solution Indeed, $(X_k - \mathbf{E}X_k)/\mathrm{Std}(X_k)$ has expectation 0 and variance 1, so it is enough to consider the problem for such random quantities. Then,

(a)

$$\mathcal{F}'_X(0) = \frac{d}{du}\mathbf{E}e^{juX}\Big|_{u=0} = j\mathbf{E}Xe^{juX}\Big|_{u=0} = j\mathbf{E}X = 0$$

$$\mathcal{F}''_X(0) = \frac{d}{du}j\mathbf{E}Xe^{juX}\Big|_{u=0} = j^2\mathbf{E}X^2e^{juX}\Big|_{u=0} = -\mathbf{E}X^2 = -1$$

(b) If Z is an $N(0, 1)$ random quantity, then

$$\mathcal{F}_Z(u) = \int_{-\infty}^{\infty} e^{jux}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = e^{-u^2/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2jux+(ju)^2)}\frac{1}{\sqrt{2\pi}}\,dx$$

$$= e^{-u^2/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-ju)^2}\frac{1}{\sqrt{2\pi}}\,dx = e^{-u^2/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\frac{1}{\sqrt{2\pi}}\,dz = e^{-u^2/2}$$

by changing the variable $x - ju \mapsto z$ in the penultimate integral and because the Gaussian density in the last integral integrates to 1.

(c) Indeed, if X and Y are independent, then

$$\mathcal{F}_{X+Y}(u) = \mathbf{E}e^{ju(X+Y)} = \mathbf{E}(e^{juX}\cdot e^{juY}) = \mathbf{E}e^{juX}\cdot\mathbf{E}e^{juY} = \mathcal{F}_X(u)\cdot\mathcal{F}_Y(u)$$

because the expectation of a product of independent random quantities is the product of their expectations.

(d) Observe first that

$$\frac{\sqrt{n}(\bar{X}-\mu_X)}{\mathrm{Std(X)}} = \frac{1}{\sqrt{n}}(Y_1+\ldots+Y_n),$$

where

$$Y_1 = \frac{X_1-\mu_X}{\mathrm{Std(X)}},\ldots,Y_n = \frac{X_n-\mu_X}{\mathrm{Std(X)}}$$

so that, in particular, $Y_1,\ldots,Y_n$, are independent, identically distributed with $\mathbf{E}Y_1 = 0$ and $\mathbf{E}Y_1^2 = 1$. Hence, using (a)–(c)

$$\mathcal{F}_{\sqrt{n}(\bar{X}-\mu_X)/\mathrm{Std(X)}}(u) = \mathcal{F}_{(Y_1/\sqrt{n}+\cdots+Y_n/\sqrt{n})}(u) = \mathcal{F}_{(Y_1/\sqrt{n})}(u)\cdot\ldots\cdot\mathcal{F}_{(Y_n/\sqrt{n})}(u)$$

$$= [\mathcal{F}_{Y_1}(u/\sqrt{n})]^n.$$

Now, for each fixed but arbitrary u, instead of calculating the limit $n \to \infty$ of the above characteristic functions, it will be easier to calculate the limit of their logarithm. Indeed, in view of de l'Hospital's rule applied twice (differentiating with respect to n; explain why this is OK):

$$\lim_{n\to\infty} \log \mathcal{F}_{\sqrt{n}(\bar{X}-\mu_X)/\mathrm{Std(X)}}(u) = \lim_{n\to\infty} \log[\mathcal{F}_{Y_1}(u/\sqrt{n})]^n$$

$$= \lim_{n\to\infty} \frac{\log \mathcal{F}_{Y_1}(u/\sqrt{n})}{1/n} = \lim_{n\to\infty} \frac{(1/\mathcal{F}_{Y_1}(u/\sqrt{n})) \cdot \mathcal{F}'_{Y_1}(u/\sqrt{n}) \cdot (-\frac{1}{2}u/n^{3/2})}{-1/n^2}$$

$$= \frac{1}{2}u \lim_{n\to\infty} \frac{1 \cdot \mathcal{F}'_{Y_1}(u/\sqrt{n})}{1/n^{1/2}} = \frac{1}{2}u \lim_{n\to\infty} \frac{\mathcal{F}''_{Y_1}(u/\sqrt{n}) \cdot (-\frac{1}{2}u/n^{3/2})}{-\frac{1}{2} \cdot 1/n^{3/2}} = -\frac{1}{2}u^2,$$

because $\mathcal{F}'_{Y_1}(0) = 0$, and $\mathcal{F}''_{Y_1}(0) = -1$, see Part (a). So for the characteristic functions themselves

$$\lim_{n\to\infty} \mathcal{F}_{\sqrt{n}(\bar{X}-\mu_X)/\mathrm{Std(X)}}(u) = e^{-u^2/2},$$

and we recognize the above limit as the characteristic function of the $N(0, 1)$ random quantity, see Part (b).

The above proof glosses over the issue of whether indeed convergence of characteristic functions implies convergence of c.d.f.s of the corresponding random quantities. The relevant Continuity Theorem can be found in any of the mathematical probability theory textbooks listed in the Bibliographical Comments at the end of this volume.

Chapter 5

Problem 1. Consider a random signal

$$X(t) = \sum_{k=0}^{n} A_k \cos\Big(2\pi k f_k(t + \Theta_k)\Big),$$

where $A_0, \Theta_1, \ldots, A_n, \Theta_n$ are independent random variables of finite variance, and $\Theta_1, \ldots, \Theta_n$ are independent, independent of As and uniformly distributed on the time interval $[0, P = 1/f_0]$. Is this signal stationary? Find its mean, and autocorrelation functions.

Solution The mean value of the signal (we use the independence conditions)

$$EX(t) = E\Big(A_1 \cos 2\pi f_0(t + \Theta_1)\Big) + \ldots + E\Big(A_n \cos 2\pi n f_0(t + \Theta_n)\Big)$$

$$= EA_1 \cdot \int_0^P \cos 2\pi f_0(t + \theta_1)\frac{d\theta_1}{P} + \ldots + EA_n \cdot \int_0^P \cos 2\pi n f_0(t + \theta_n)\frac{d\theta_n}{P} = 0.$$

The mean value doesn't depend on time t; thus the first requirement of stationarity is satisfied.

The autocorrelation function

$$\gamma_X(t, t+\tau) = E[X(t)X(t+\tau)]$$

$$= E\left(\sum_{i=1}^{n} A_i \cos(2\pi i f_0(t+\Theta_i)) \cdot \sum_{k=1}^{n} A_k \cos(2\pi k f_0(t+\tau+\Theta_k))\right)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{n} E(A_i A_k) \cdot E\Big(\cos(2\pi i f_0(t+\Theta_i)) \cdot \cos(2\pi k f_0(t+\tau+\Theta_k))\Big)$$

$$= \sum_{i=1}^{n} \frac{EA_i^2}{2} \cos(2\pi i f_0 \tau),$$

because all the cross-terms are zero. The autocorrelation function is thus depending only on τ (and not on t), so that the second condition of stationarity is also satisfied.

Problem 2. Consider a random signal

$$X(t) = A_1 \cos 2\pi f_0(t+\Theta_0),$$

where A_1, Θ_0, are independent random variables, and Θ_0 is uniformly distributed on the time interval $[0, P/3 = 1/(3f_0)]$. Is this signal stationary? Is the signal $Y(t) = X(t) - \mathbf{E}X(t)$ stationary. Find its mean, and autocorrelation functions.

Solution The mean value of the signal

$$\mathbf{E}X(t) = \mathbf{E}\Big(A\cos 2\pi f_0(t+\Theta)\Big) = \mathbf{E}A \cdot \int_0^{P/3} \cos(2\pi f_0(t+\theta))\,\frac{d\theta}{P/3}$$

$$= \frac{3\mathbf{E}A}{2\pi} \sin(2\pi f_0(t+\theta))\,\Big|_{\theta=0}^{P/3} = \frac{3\mathbf{E}A}{2\pi}\Big(\sin(2\pi f_0(t+P/3)) - \sin(2\pi f_0 t)\Big).$$

Since

$$\sin p - \sin q = 2\cos\frac{p+q}{2}\sin\frac{p-q}{2},$$

we finally get

$$\mathbf{E}X(t) = \mathbf{E}A\frac{3\sqrt{3}}{2\pi}\cos\Big(2\pi f_0 t + \frac{\pi}{3}\Big),$$

which clearly depends on t in an essential way. Thus the signal is not stationary.

The signal $Y(t) = X(t) - \mathbf{E}X(t)$ has obviously mean zero. Its autocorrelation function

$$\gamma_Y(t,s) = \mathbf{E}[X(t)X(s)] - \mathbf{E}X(t)\mathbf{E}X(s)$$

$$= \mathbf{E}A^2 \int_0^{P/3} \cos 2\pi f_0(t+\theta)\cos 2\pi f_0(s+\theta)\frac{3}{P}\,d\theta - \mathbf{E}X(t)\mathbf{E}X(s),$$

with $\mathbf{E}X(t)$ already calculated above. Since $\cos\alpha\cos\beta = [\cos(\alpha+\beta) + \cos(\alpha-\beta)]/2$, the integral in the first term is

$$\cos 2\pi f_0(t-s) + \frac{3}{4\pi}\Big(\sin(2\pi f_0(t+s+\frac{2}{3f_0})) - \sin(2\pi f_0(t+s))\Big).$$

Now, $\gamma_Y(t,s)$ can be easily calculated. Simplify the expression (and plot the ACF) before you decide the stationarity issue for $Y(t)$.

Problem 8. Show that if $X_1, X_2, \ldots, X_n$, are independent, exponentially distributed random quantities with identical p.d.f.s e^{-x}, $x \geq 0$, then their sum $Y_n = X_1 + X_2 + \cdots + X_n$ has the p.d.f. $e^{-y}y^{n-1}/(n-1)!$, $y \geq 0$. Use the technique of characteristic functions (Fourier transforms) from Chap. 3. The random quantity Y_n is said to have the Gamma probability distribution with parameter n. Thus the Gamma distribution with parameter 1 is just the standard exponential distribution, see Example 4.1.4. Produce plots of Gamma p.d.f.s with parameters $n = 2, 5, 20$, and 50. Comment on what you observe as n increases.

Solution The characteristic function (see, Chap. 3) for each of X_i's is

$$\mathcal{F}_X(u) = \mathbf{E}e^{juX} = \int_0^\infty e^{jux}e^{-x}\,dx = \frac{1}{1-ju}.$$

In view of the independence of X_i's the characteristic function of Y_n is necessarily the n-th power of the common characteristic function of X_i's:

$$\mathcal{F}_{Y_n}(u) = \mathbf{E}e^{ju(X_1+\ldots+X_n)} = \mathbf{E}e^{juX_1}\cdot\ldots\cdot\mathbf{E}e^{juX_n} = \frac{1}{(1-ju)^n}.$$

So it suffices to verify that the characteristic function of the p.d.f. $f_n(u) = e^{-y}y^{n-1}/(n-1)!$, $y \geq 0$ is also of the form $(1-ju)^{-n}$. Indeed, integrating by parts, we obtain

$$\int_0^\infty e^{juy}e^{-y}\frac{y^{n-1}}{(n-1)!}\,dy = \frac{e^{(ju-1)y}}{ju-1}\cdot\frac{y^{n-1}}{(n-1)!}\Big|_{y=0}^\infty + \frac{1}{1-ju}\int_0^\infty e^{(ju-1)y}\frac{y^{n-2}}{(n-2)!}\,dy.$$

The first term on the right side is zero so that we get the recursive formula

$$\mathcal{F}_{f_n}(u) = \frac{1}{1-ju}\mathcal{F}_{f_{n-1}}(u),$$

which gives the desired result since $\mathcal{F}_{f_1}(u) = \mathcal{F}_X(u) = (1-ju)^{-1}$.

Chapter 6

Problem 5. A stationary signal $X(t)$ has the autocorrelation function

$$\gamma_X(\tau) = 16e^{-5|\tau|}\cos 20\pi\tau + 8\cos 10\pi\tau$$

(a) Find the variance of the signal
(b) Find the power spectrum density of this signal
(c) Find the value of the spectral density at zero frequency

Solution

(a)

$$\sigma^2 = \gamma_X(0) = 16 + 8 = 24$$

(b) Let us denote the operation of Fourier transform by $\mathcal{F}$. Then, writing, perhaps, a little informally,

$$S_X(f) = \int_{-\infty}^{\infty} \gamma_X(\tau)e^{-j2\pi f\tau}d\tau = (\mathcal{F}\gamma_X)(f)$$

$$= \mathcal{F}\Big(16e^{-5|\tau|}\cdot\cos(20\pi\tau) + 8\cos(10\pi\tau)\Big)(f)$$

$$= 16\cdot\Big(\mathcal{F}(e^{-5|\tau|}) * \mathcal{F}(\cos(20\pi\tau))\Big)(f) + 8\cdot\mathcal{F}(\cos(10\pi\tau))(f).$$

But

$$\mathcal{F}(e^{-5|\tau|})(f) = \frac{2\cdot 5}{5^2 + (2\pi f)^2} = \frac{10}{25 + (2\pi f)^2},$$

and

$$\mathcal{F}(\cos 20\pi\tau)(f) = \frac{\delta(f+10) + \delta(f-10)}{2},$$

so that

$$\Big(\mathcal{F}(e^{-5|\tau|}) * \mathcal{F}(\cos(20\pi\tau))\Big)(f)$$

$$= \int_{-\infty}^{\infty} \frac{10}{25 + (2\pi f)^2} * \frac{\delta(f-s+10) + \delta(f-s-10)}{2}ds$$

$$= 5\left[\int_{-\infty}^{\infty} \frac{\delta(s-(f+10))}{25+(2\pi f)^2} ds + \int_{-\infty}^{\infty} \frac{\delta(s-(f-10))}{25+(2\pi f)^2} ds\right]$$

$$= 5\left[\frac{1}{25+4\pi^2(f+10)^2} + \frac{1}{25+4\pi^2(f-10)^2}\right],$$

because we know that $\int \delta(f-f_0)X(f)\,df = X(f_0)$. Since $\mathcal{F}(\cos 10\pi\tau)(f) = \delta(f+5)/2 + \delta(f-5)/2$,

$$S_X(f) = \frac{80}{25+4\pi^2(f+10)^2} + \frac{80}{25+4\pi^2(f-10)^2} + 4\delta(f+5) + 4\delta(f-5).$$

Another way to proceed would be to write $e^{-5|\tau|} \cdot \cos(20\pi\tau)$ as $e^{-5\tau} \cdot (e^{j(20\pi\tau)} - e^{-j(20\pi\tau)})/2$, for $\tau > 0$ (similarly for negative τ's), and do the integration directly in terms of just exponential functions (but it was more fun to do convolutions with the Dirac-delta impulses, wasn't it?).

(c)

$$S_X(0) = \frac{80}{25+4\pi^2 100} + \frac{80}{25+4\pi^2 100} + 4\delta(5) + 4\delta(-5) = \frac{160}{25+400\pi^2}.$$

Problem 9. Verify positive-definiteness (see, Remark 5.2.1) of autocorrelation functions of stationary signals directly from their definition.

Solution Let N be an arbitrary positive integer, $t_1, \ldots, t_N \in \mathbf{R}$, and $z_1, \ldots, z_N \in \mathbf{C}$. Then, in view of stationarity of $X(t)$,

$$\sum_{n=1}^{N}\sum_{k=1}^{N} \gamma(t_n - t_k) z_n z_k^* = \sum_{n=1}^{N}\sum_{k=1}^{N} \mathbf{E}[X^*(t)X(t+(t_n-t_k)]z_n z_k^*$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{N} \mathbf{E}[X^*(t+t_k)X(t+t_n)]z_n z_k^* = \mathbf{E}\sum_{n=1}^{N}\sum_{k=1}^{N} (z_k X(t+t_k))^* \cdot (z_n X(t+t_n))$$

$$= \mathbf{E}\left|\sum_{n=1}^{N} z_n X(t+t_n)\right|^2 \geq 0.$$

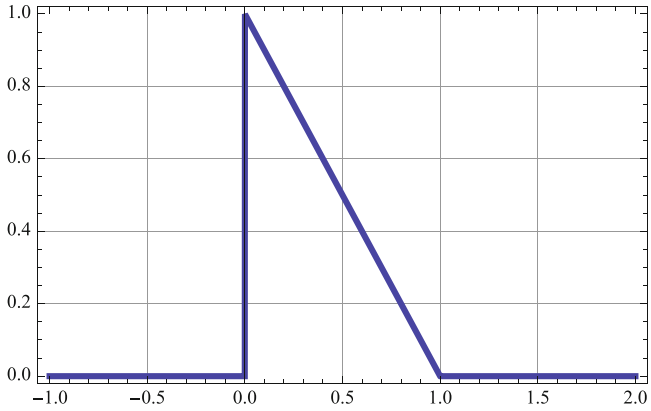
Chapter 7

Problem 1. The impulse response function of a linear system is $h(t) = 1 - t$, for $0 \leq t \leq 1$, and 0 elsewhere.

(a) Produce a graph of $h(t)$.
(b) Assume that the input is the standard white noise. Find the autocovariance function of the output.
(c) Find the power transfer function of the system, its equivalent-noise bandwidth and half-power bandwidth.
(d) Assume that the input has the autocovariance function $\gamma_X(t) = 3/(1 + 4t^2)$. Find the power spectrum of the output signal.
(e) Assume that the input has the autocovariance function $\gamma_X(t) = \exp(-4|t|)$. Find the power spectrum of the output signal.
(f) Assume that the input has the autocovariance function $\gamma_X(t) = 1-|t|$ for $|t| < 1$ and 0 elsewhere. Find the power spectrum of the output signal.

Solution

(a)



(b) With $\gamma_X(\tau) = \delta(\tau)$, the autocovariance function of the output is

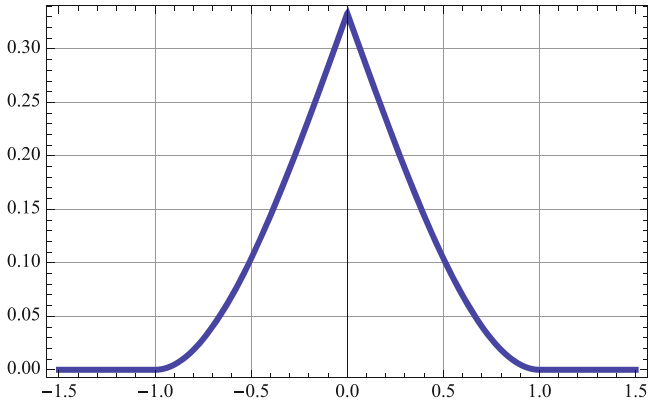
$$\gamma_Y(\tau) = \int_0^\infty \int_0^\infty \gamma_X(\tau - u + s)h(s)h(u)\, ds\, du =$$

$$= \int_0^1 \int_0^1 \delta(s - (u - \tau))(1 - s)(1 - u)\, ds\, du.$$

As long $0 < u - \tau < 1$, which implies that $-1 < \tau < 1$, the inner integral is

$$\int_0^1 \delta(s - (u - \tau))(1 - s)\, ds = 1 - (u - \tau),$$

and otherwise it is zero.



So, for $0 < \tau < 1$,

$$\gamma_Y(\tau) = \int_\tau^1 (1 - (u - \tau))(1 - u)\, du = \frac{1}{6}(\tau - 1)^2(\tau + 2),$$

and, in view of evenness of ACF,

$$\gamma_Y(\tau) = \frac{1}{6}(|\tau| - 1)^2(|\tau| + 2) \quad \text{for } -1 < \tau < 1,$$

for $-1 < \tau < 1$, and it is zero outside the interval $[-1, 1]$, see the figure above.

(c) The transfer function of the system is

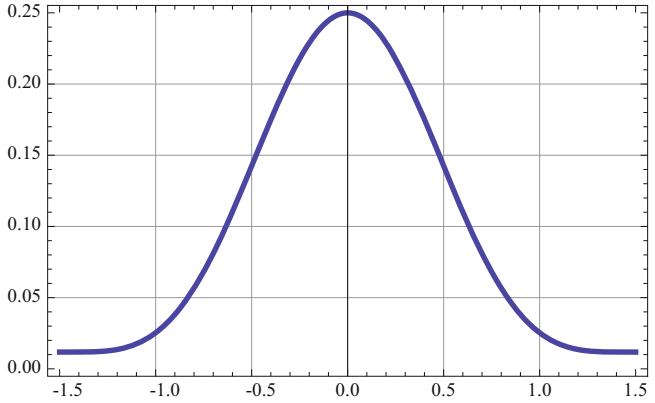
$$H(f) = \int_0^1 (1 - t)e^{-2\pi jft} dt = \frac{\sin^2(\pi f)}{2\pi^2 f^2} - j\frac{2\pi f - \sin(2\pi f)}{4\pi^2 f^2}.$$

Therefore, the power transfer function is

$$|H(f)|^2 = H(f)H^*(f) = \left(\frac{\sin^2(\pi f)}{2\pi^2 f^2}\right)^2 + \left(\frac{2\pi f - \sin(2\pi f)}{4\pi^2 f^2}\right)^2$$

$$= \frac{-1 + \cos 2\pi f + 2\pi f \sin 2\pi f - 2\pi^2 f^2}{8\pi^4 f^4}.$$

It is shown in the figure below.



To find the value of the power transfer function at $f = 0$ one can apply L'Hospital's rule, differentiating numerator and denominator of $|H(f)|^2$ three times which yields $|H(0)|^2 = 1/4$. Thus the equivalent noise bandwidth is

$$BW_n = \frac{1}{2|H(0)|^2} \int_0^1 (1-t)^2 \, dt = 2/3.$$

Checking the above plot of the power transfer function one finds that, approximately, the half power bandwidth is $BW_{1/2} = 0.553$.

(d) The power spectrum of the output signal is given by

$$S_Y(f) = S_X(f)|H(f)|^2,$$

where $S_X(f)$ is the power spectrum of the input signal. In our case,

$$S_X(f) = \int_{-\infty}^{\infty} \frac{3}{1+4t^2} \cdot \cos(2\pi f t) \, dt = \frac{3\pi}{2} e^{-\pi|f|}.$$

Therefore,

$$S_Y(f) = \frac{3\pi}{2} e^{-\pi|f|} \cdot \frac{-1 + \cos 2\pi f + 2\pi f \sin 2\pi f - 2\pi^2 f^2}{8\pi^4 f^4}.$$

(e) In this case, similarly,

$$S_X(f) = \int_{-\infty}^{\infty} e^{-4|t|} \cdot \cos(2\pi f t) \, dt = \frac{2}{4 + \pi^2 f^2}$$

and

$$S_Y(f) = \frac{2}{4+\pi^2 f^2} \cdot \frac{-1+\cos 2\pi f + 2\pi f \sin 2\pi f - 2\pi^2 f^2}{8\pi^4 f^4}.$$

(f) Finally, here

$$S_X(f) = \frac{(\sin \pi f)^2}{\pi^2 f^2},$$

and

$$S_Y(f) = \frac{(\sin \pi f)^2}{\pi^2 f^2} \cdot \frac{-1+\cos 2\pi f + 2\pi f \sin 2\pi f - 2\pi^2 f^2}{8\pi^4 f^4}.$$

Problem 5. Consider the circuit shown in Fig. 7.10. Assume that the input, $X(t)$, is the standard white noise.

(a) Find the power spectra $S_Y(f)$ and $S_Z(f)$ of the outputs $Y(t)$ and $Z(t)$.
(b) Find the crosscorrelation between those two outputs,

$$\gamma_{YZ}(\tau) = \mathbf{E}\Big(Z(t)Y(t+\tau)\Big).$$

Solution

(a) Note that $X(t) = Y(t)+Z(t)$. The impulse response function for the "Z" circuit is

$$h_Z(t) = \frac{1}{RC} e^{-t/RC},$$

and

$$Y(t) = X(t) - \int_0^\infty h_Z(s) X(t-s)\, ds.$$

So the impulse response function for the "Y" circuit is

$$h_Y(t) = \delta(t) - \int_0^\infty \frac{1}{RC} e^{-s/RC} \delta(t-s)\, ds =$$

$$= \delta(t) - \frac{1}{RC} e^{-t/RC}, \qquad t \geq 0.$$

The Fourier transform of $h_Y(t)$ will give us the transfer function

$$H_Y(f) = \int_0^\infty (\delta(t) - \frac{1}{RC} e^{-t/RC}) e^{-2\pi j f t}\, dt = \frac{2\pi j RCf}{1+2\pi j RCf}.$$

For the standard white noise input $X(t)$, the power spectrum of the output is equal to the power transfer function of the system. Indeed,

$$S_Y(f) = 1 \cdot |H_Y(f)|^2 = \frac{4\pi^2 R^2 C^2 f^2}{1 + 4\pi^2 R^2 C^2 f^2}.$$

The calculation of $S_X(f)$ has been done before as the "Z" circuit represents the standard RC filter.

(b)

$$\begin{aligned}
\gamma_{yz}(\tau) &= E(Y(t)Z(t+\tau)) \\
&= E\left[\int_{-\infty}^{\infty} X(t-s)h_Y(s)\,ds \int_{-\infty}^{\infty} X(t+\tau-u)h_Z(u)\,du\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} EX(t-s)X(t+\tau-u)h_Y(s)h_Z(u)\,ds\,du \\
&= \int_{0}^{\infty}\int_{0}^{\infty} \delta(\tau-u+s)(\delta(s) - \frac{1}{RC}e^{-s/RC})\frac{1}{RC}e^{-u/RC}\,du\,ds \\
&= \int_{0}^{\infty} (\delta(s) - \frac{1}{RC}e^{-s/RC})\frac{1}{RC}e^{-(\tau+s)/RC}\,ds \\
&= \int_{0}^{\infty} \delta(s)\frac{1}{RC}e^{-(\tau+s)/RC}ds - \int_{0}^{\infty} \frac{1}{RC}e^{-s/RC}\frac{1}{RC}e^{-(\tau+s)/RC}\,ds \\
&= \frac{1}{RC}e^{-\tau/RC} - \frac{1}{2RC}e^{-\tau/RC} = \frac{1}{2RC}e^{-\tau/RC}.
\end{aligned}$$

Chapter 8

Problem 2. A signal of the form $x(t) = 5e^{-(t+2)}u(t)$ is to be detected in the presence of white noise with a flat power spectrum of 0.25 V^2/Hz using a matched filter.

(a) For $t_0 = 2$ find the value of the impulse response of the matched filter at $t = 0, 2, 4$.
(b) Find the maximum output signal-to-noise ratio that can be achieved if $t_0 = \infty$.
(c) Find the detection time t_0 that should be used to achieve output signal-to-noise ratio that is equal to 95% of the maximum signal-to-noise ratio discovered in part (b).
(d) The signal $x(t) = 5e^{-(t+2)}u(t)$ is combined with white noise having a power spectrum of 2 V^2/Hz. Find the value of RC such that the signal/noise at the output of the RC filter is maximal at $t = 0.01$ s.

Solution

(a) The impulse response function for the matched filter is of the form

$$h(s) = 5\exp[-(t_0 - s + 2)] \cdot u(t_0 - s) = 5e^{-(4-s)}u(2 - s),$$

where t_0 is the detection time and $u(t)$ is the usual unit step function. Therefore,

$$h(0) = 5e^{-4}, \quad h(2) = 5e^{-2}, \qquad h(4) = 0.$$

(b) The maximum signal-to-noise ration at detection time t_0 is

$$\frac{\mathcal{S}}{\mathcal{N}}_{\max}(t_0) = \frac{\int_0^\infty x^2(t_0 - s)\,ds}{N_0} = \frac{\int_0^{t_0} 25e^{-2(t_0-s+2)}\,ds}{0.25} = 50e^{-4}(1 - e^{-2t_0}).$$

So

$$\frac{\mathcal{S}}{\mathcal{N}}_{\max}(t_0 = 0) = 50e^{-4}.$$

(c) The sought detection time t_0 can thus be found by solving numerically the equation

$$50e^{-4}(1 - e^{-2t_0}) = 0.95 \cdot 50e^{-4},$$

which yields, approximately, $t_0 = -\log 0.05/2 \approx 1.5$.

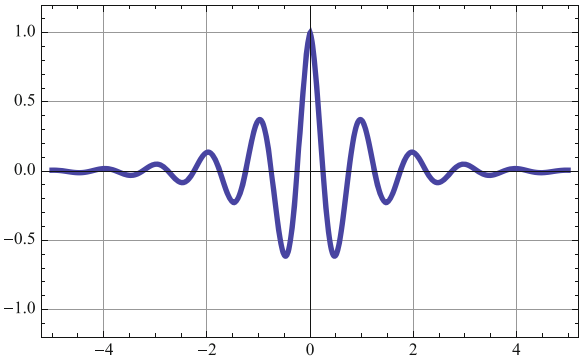
Chapter 9

Problem 1. A zero-mean Gaussian random signal has the autocovariance function of the form

$$\gamma_X(\tau) = e^{-0.1|\tau|}\cos 2\pi\tau.$$

Plot it. Find the power spectrum $S_X(f)$. Write the covariance matrix for the signal sampled at four time instants separated by 0.5 s. Finds its inverse (numerically; use any of the familiar computing platforms, such as *Mathematica, Matlab*, etc.).

Solution We will use *Mathematica* to produce plots and do symbolic calculations although it is fairly easy to calculate $S_X(f)$ by direct integration. The plot of $\gamma_X(\tau)$ is below.



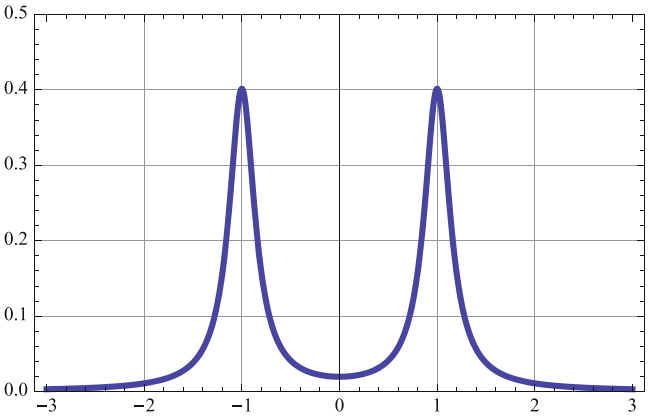
The power spectrum $S_X(f)$ is the Fourier Transform of the ACF, so

```
In[1]:= GX[t_] := Exp[- Abs[t]]*Cos[2*Pi*t];

In[2]:= FourierTransform [GX[t], t, 2*Pi*f]
```

$$\text{Out[2]} = \frac{1}{\sqrt{2\pi}\left(1+4\,(-1+f)^2\pi^2\right)} + \frac{1}{\sqrt{2\pi}\left(1+4\,(1+f)^2\pi^2\right)}$$

Note that the Fourier Transform in *Mathematica* is defined as a function of the angular velocity variable $\omega = 2\pi f$; hence the above substitution. The plot of the power spectrum is shown below.



Problem 3. Find the joint p.d.f. of the signal from Problem 1 at $t_1 = 1, t_2 = 1.5, t_3 = 2$, and $t_4 = 2.5$. Write the integral formula for

$$P(-2 \le X(1) \le 2, -1 \le X(1.5) \le 4, -1 \le X(2) \le 1, 0 \le X(2.5) \le 3).$$

Evaluate the above probability numerically.

Solution Again, we use *Mathematica* to carry out all the numerical calculations. First, we calculate the relevant covariance matrix

```
In[3]:= CovGX = N[{{GX[0], GX[0.5], GX[1], GX[1.5]},
    {GX[0.5], GX[0], GX[0.5], GX[1]},
    {GX[1], GX[0.5], GX[0], GX[0.5]},
    {GX[1.5], GX[1], GX[0.5], GX[0]}}] // MatrixForm
```

$$\text{Out[3]} = \begin{pmatrix} 1. & -0.606531 & 0.367879 & -0.22313 \\ -0.606531 & 1. & -0.606531 & 0.367879 \\ 0.367879 & -0.606531 & 1. & -0.606531 \\ -0.22313 & 0.367879 & -0.606531 & 1. \end{pmatrix}$$

Its determinant, and its inverse are:

```
In[4]:=Det [CovGX]
Out[4]= 0.25258

In[5]:= ICovGX = Inverse[CovGX] // MatrixForm
Out[5]=
```

$$\begin{pmatrix} 1.58198 & 0.959517 & -6.73384 \times 10^{-17} & -1.11022 \times 10^{-16} \\ 0.959517 & 2.16395 & 0.959517 & -2.63452 \times 10^{-16} \\ -1.11022 \times 10^{-16} & 0.959517 & 2.16395 & 0.959517 \\ -5.55112 \times 10^{-17} & -2.22045 \times 10^{-16} & 0.959517 & 1.58198 \end{pmatrix}$$

Thus the corresponding 4D Gaussian p.d.f. is

```
In[6]:=  f[x1, x2, x3, x4]= (1/((2*Pi)^2
        *Sqrt[Det[CovGX]])) *
     Exp[-(1/2)*
     Transpose[{{x1},{x2},{x3},{x4}}]. ICovGX.
        {x1,x2,x3,x4}]
Out[6]=  0.05 * E^( -0.79  x1^2 - 1.08  x2^2
- 0.96 x2 x3 -1.08  x3^2 +  x1 (-0.96 x2
+ 8.92*10^-17 x3 + 8.33*10^-17 x4) +
 2.43*10^-16 x2 x4 - 0.96 x3 x4 - 0.79  x4^2
```

Note the quadratic form in four variables, `x1, x2, x3, x4,` in the exponent. The calculations of the sought probability requires evaluation of the 4D integral,

$$P\Big(-2 \le X(1) \le 2, -1 \le X(1.5) \le 4, -1 \le X(2) \le 1, 0 \le X(2.5) \le 3\Big) =$$

$$\int_{-2}^{2}\int_{-1}^{4}\int_{-1}^{1}\int_{0}^{3} f(x_1, x_2, x_3, x_4)\, dx_1\, dx_2\, dx_3\, dx_4,$$

which can be done only numerically:

```
In[7]:= NIntegrate[ f[x1, x2, x3, x4],
            {x1, -2, 2}, {x2, -1, 4}, {x3, -1, 1},
            {x4, 0, 3}]
Out[7]= {0.298126}
```

Problem 4. Show that if a 2D Gaussian random vector $\vec{Y} = (Y_1, Y_2)$ has uncorrelated components Y_1, Y_2, then those components are statistically independent random quantities.

Solution Recall the p.d.f. of a general zero-mean 2D Gaussian random vector (Y_1, Y_2) (see, (8.2.9)):

$$f_{\vec{\mathbf{Y}}}(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{y_1^2}{\sigma_1^2} - 2\rho\frac{y_1 y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2}\right)\right].$$

If the two components are uncorrelated, then $\rho = 0$, and the formula takes the following simplified shape,

$$f_{\vec{\mathbf{Y}}}(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2} \cdot \exp\left[-\frac{1}{2}\left(\frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2}\right)\right],$$

and it factors into the product of the marginal densities of the two components of random vector $\vec{Y}$:

$$f_{\vec{\mathbf{Y}}}(y_1, y_2) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{y_1^2}{\sigma_1^2}\right)\right] \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{y_2^2}{\sigma_2^2}\right)\right],$$

$$= f_{Y_1}(y_1) \cdot f_{Y_2}(y_2),$$

which proves the statistical independence of Y_1 and Y_2.

Chapter 10

Problem 8. Verify that the additivity property (10.3.7) of any function forces its linear form (10.3.8).

Solution Our assumption is that a function $\mathcal{C}(v)$ satisfies the functional equation

$$\mathcal{C}(v+w) = \mathcal{C}(v) + \mathcal{C}(w) \tag{S.9.1}$$

for any real numbers v, w. We will also assume that is it continuous although the proof is also possible (but harder) under a weaker assumption of measurability. Taking $v = 0, w = 0$, gives

$$\mathcal{C}(0) = \mathcal{C}(0) + \mathcal{C}(0) = 2\mathcal{C}(0),$$

which implies that $\mathcal{C}(0) = 0$. Furthermore, taking $w = -v$, we get

$$\mathcal{C}(0) = \mathcal{C}(v) + \mathcal{C}(-v) = 0,$$

so that $\mathcal{C}(v)$ is necessarily an odd function.

Now, iterating (S.9.1) n times we get that, for any real number v

$$\mathcal{C}(nv) = n \cdot \mathcal{C}(v)$$

and choosing $v = 1/n$, we see that $\mathcal{C}(1) = n\mathcal{C}(1/n)$ for any positive integer n. Replacing n by m in the last equality and combining it with the preceding equality with $v = 1/m$, we get that, for any positive integers n, m,

$$\mathcal{C}\left(\frac{n}{m}\right) = \frac{n}{m} \cdot \mathcal{C}(1).$$

Finally, since any real number can be approximated by the rational numbers of the form n/m, and since $\mathcal{C}$ was assumed to be continuous, we get that for any real number

$$\mathcal{C}(v) = v \cdot \mathcal{C}(1),$$

that is, $\mathcal{C}(v)$ is necessarily a linear function.

Bibliographical Comments

The classic modern treatise on the theory of Fourier series and integrals which influenced much of the harmonic analysis research in the second half of the twentieth century is

[1] A. Zygmund, *Trigonometric Series,* Cambridge University Press, 1959.

More modest in scope, but perhaps also more usable for the intended reader of this text, are

[2] H. Dym and H. McKean, *Fourier Series and Integrals*, Academic Press, New York 1972
[3] T.W. Körner, *Fourier Analysis*, Cambridge University Press, 1988,
[4] E.M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press 2003.
[5] P.P.G. Dyke, *An Introduction to Laplace Transforms and Fourier Series*, Springer-Verlag, New York 1991.

The above four books are now available in paperback.

The Schwartz distributions (generalized functions), such as the Dirac-delta impulse and its derivatives, with special emphasis on their applications in engineering and the physical sciences, are explained in

[6] F. Constantinescu, *Distributions and Their Applications in Physics*, Pergamonn Press, Oxford 1980,
[7] T. Schucker, *Distributions, Fourier Transforms and Some of Their Applications to Physics*, World Scientific, Singapore 1991,
[8] A.I. Saichev and W.A. Woyczyński, *Distributions in the Physical and Engineering Sciences, Volume 1. Distributional and Fractal Calculus, Integral Transforms and Wavelets*, Birkhäuser-Boston 1997.

The second volume of the above text,

[9] A.I. Saichev and W.A. Woyczyński, *Distributions in the Physical and Engineering Sciences, Volume 2. Linear, Nonlinear, Fractal and Random Dynamics in Continuous Media*, Birkhäuser-Boston 2005,

is scheduled to appear later this year.

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7

Good elementary introductions to probability theory, and accessible reads for the engineering and the physical sciences audience, are

[11] S.M. Ross, Introduction to Probability Models, Academic Press, Burlington, MA, 2003.
[10] J. Pitman, Probability, Springer-Verlag, New York 1993.

On the other hand,

[12] M. Denker and W.A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity, and Chaotic Behavior in Engineering and Science,* Birkhäuser-Boston, 1998,

deal with a broader issue of how randomness appears in diverse models of natural phenomena and with the fundamental question of the meaning of randomness itself.

A more ambitious, mathematically rigorous treatments of probability theory, based on measure theory, can be found in

[13] P. Billingsley, *Probability and Measure*, J. Wiley, New York 1983,
[14] O. Kallenberg, *Foundations of Modern Probability*, Springer-Verlag, New York 1997, and
[15] M. Loève, *Probability Theory*, Van Nostrand, Princeton 1961.

All three also contain a substantial account of the theory of stochastic processes.

Readers more interested in the general issues of statistical inference and, in particular, parametric estimation, should consult

[16] G. Casella and R.L. Berger, *Statistical Inference*, Duxbury, Pacific Grove, CA, 2002, or
[17] D.C. Montgomery and G.C. Runger, *Applied Statistics and Probability for Engineers*, J. Wiley, New York 1994.

The classic texts on the general theory of stationary processes (signals) are

[18] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*, Dover Books 2004, and
[19] A.M. Yaglom, *Correlation Theory of Stationary and Related Random Functions, Volumes I and II*, Springer-Verlag, New York 1987.

However, the original

[20] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time series*, MIT Press and J. Wiley, New York 1950,

still reads very well.

Statistical tools in the spectral analysis of stationary discrete-time random signals (also known as *time series*) are explored in

[21] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*, J. Wiley, New York 1976, and
[22] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York 1991,

and difficult issues of analysis of nonlinear and nonstationary random signals are tackled in

[23] M.B. Priestley, *Non-linear an Non-stationary Time Series Analysis,* Academic Press, London, 1988, and

[24] W.J. Fitzgerald, R.L. Smith, A.T. Walden and P.C. Young, Eds., *Nonlinear and Nonstationary Signal Processing* , Cambridge University Press 2000.

The latter is a collection of articles, by different authors, on the current research issues in the area.

A more engineering approach to random signal analysis can be found in a large number of sources, including

[25] A. Papoulis, *Signal Analysis*, McGaw-Hill, New York 1977, and
[26] R.G. Brown and P.Y. Hwang, *Introduction to Random Signal Analysis and Kalman Filtering*. J. Wiley, New York 1992.

A general discussion of transmission of signals through linear systems can be found in

[27] M.J. Roberts, *Signals and Systems: Analysis of Signals Through Linear Systems*, McGraw-Hill, New York 2003, and
[28] B.D.O. Anderson, and J. B. Moore, *Optimal Filtering*, Dover Books, 2005.

Gaussian stochastic processes are thoroughly investigated in

[29] I.A. Ibragimov and Yu. A. Rozanov, *Gaussian Random Processes,* Springer-Verlag, New York 1978, and
[30] M.A. Lifshits, *Gaussian Random Functions*, Kluwer, 1995,

and for a review of the modern mathematical theory of, not necessarily second-order, and not necessarily Gaussian stochastic integrals we refer to

[31] S. Kwapien and W.A. Woyczyński, *Random Series and Stochastic Integrals: Single and Multiple*, Birkhäuser-Boston 1992.

Index

A
Additive noise, 23, 294
Additivity property of probabilities, 116
Adhesion model, 10
Analog-to-digital conversion, 4, 23, 294
Anderson, B.D.O., 325
Angular velocity, 26, 269, 318
Approximation of periodic signals
  by Césaro averages, 39
  at each time separately, 36–37
  Gibbs phenomenon, 39–41
  at jump points, 38
  mean-square error, 36
  in power, 36
  uniform, 37–39
ARMA system, 209–210
Autocorrelation function (ACF), 73–75, 153, 154, 169, 229, 308, 310
  as a positive-definite function, 250
Autocovariance function, 151–171, 177–185, 187, 190, 191, 194–202, 208, 212–214, 217–219, 221–223, 225, 237–242, 244–247, 250, 251, 254, 256, 263, 265, 287, 312, 317
  normalized, 153

B
Band-limited noise, 182, 183
Bandwidth
  equivalent noise, 191, 203, 206, 212–214, 312, 314
  of finite-time integrating circuit, 205
  half-power, 191, 203, 204, 206, 207, 212, 213, 312
Bayes formula, 122, 123
Berger, R.L., 324
Berry-Eseen theorem, 138
Billingsley, P., 254, 324
Binomial formula, 95
Bloomfield, P., 324
Branicky, M., xii
Brockwell, P.J., 169, 324
Brownian motion, 6, 256
Brown, R.G., 325
Burgers
  equation, 10
  turbulence, 10

C
Casella, G., 324
Cauchy criterion of convergence, 240
Cauchy-Schwartz inequality, 69, 125, 145, 222, 250, 304
Causal system, 195
Central limit theorem
  error of approximation in, 36, 136
  sketch of proof of, 135, 136, 146
Césaro average, 39
Chaotic behavior, 5, 96, 103, 139, 147, 252, 324
Chebyshev inequality, 243, 246
Circuit
  integrating, 196
  RC, 214
Complex exponentials, 19–22, 25, 26, 28–30, 32, 35, 41, 49, 260, 269, 297
  orthogonality of, 15

W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Statistics for Industry, Technology, and Engineering, https://doi.org/10.1007/978-3-030-20908-7

Complex numbers, 3, 14, 20, 22, 178, 250, 251, 259, 260, 265, 291
Computational complexity, 5, 52, 53
  of fast Fourier transform, 52
Computer algorithms, 249, 263–275
Conditional probability, 121–123
  reverse, 123
Confidence intervals
  for means, 136
  for variance, 136, 138
Constantinescu, F., 323
Control function cumulative, 255, 256, 258, 260, 262, 264, 282
Convergence in mean-square, 41
  Cauchy criterion for, 240
Cooley, J.W., 52
Correlation coefficient, 125–127, 129, 131, 132, 145, 153, 235, 236
Correlation of r.q.s, 125
Covariance, 125–127, 129, 131, 145, 152, 153, 156, 166, 168, 184, 191, 225, 229–246, 253, 257, 263, 317, 319
Covariance matrix, 234, 235, 237, 238, 245, 253, 257, 263, 317, 319
Cramer, H., 324
Crosscorrelation, 225, 315
Crosscovariance, 211, 213, 214, 241, 242, 246
Crutchfield, J.P., 9
Cumulative control function, 255, 256, 258, 260, 262, 264, 282
Cumulative distribution function (c.d.f.), 92, 94, 95, 100–102, 109, 112, 113, 135, 230
Cumulative power spectrum, 250–254, 275
Czekajewski, J., xii

D

Daubechies wavelet, 84–89
Davis, R.A., 169, 324
de Moivre formula, 12, 14, 19, 20, 297
Denker, M., 5, 96, 103, 139, 147, 169, 229, 252, 324
"devil's staircase", 103, 104
diffusion, 6, 7, 55, 299, 300
  equation, 300
Dirac-delta impulse
  calculus of, 50–51
  Fourier transform of, 47–51
  "probing" property of, 48, 50
Discrete Fourier transform (DFT), 51–53
Discrete-time sampling, 185, 206–212
Distributions in the sense of Schwartz, 50
Dyke, P.P.G., 323
Dym, H., 323

E

Edwards, R., xii
EEG signals, 154, 169, 170
Ergodic behavior, 9, 169
Ergodic signal, 166–170
Estimation
  of the autocorrelation, 169–170
  consistency of, 168
  of the mean, 167–168
  of parameters, 136–144, 166
  of power spectrum, 179–185
Expected value (expectation) of r.q., 113, 115, 135, 161, 176
  linear scaling of, 115

F

Fast Fourier transform (FFT), 51–53, 185, 208
  computational complexity of, 52
Filter
  causal, 196, 226
  matched, 227, 316, 317
  RC, 199–201, 203, 204, 214, 218, 219, 223, 227, 316
  Wiener, 224–227
Filtering noise out of signal, 164
Fitzgerald, W.J., 325
"Floor" function, 5
Folland, G.B., 253, 258
Fourier analysis, ix, 27, 37, 46, 178, 284, 323, 324
Fourier coefficient, 26–28, 34, 39, 41–43, 248, 251, 259, 263
Fourier expansion
  pure cosine, 30
  pure sine, 32, 39
Fourier, J.-B., 55
Fourier series, 25–56, 80, 179, 187, 210, 214, 225, 248, 285, 297, 323
  complex, 25–35
Fourier transform (FT)
  basic properties, 44–46
  computational complexity of, 52
  of convolution, 44, 46, 210, 225
  discrete, 51–53
  fast (FFT), 51–53
  inverse (IFT), 42, 43, 49, 71, 146, 177, 178, 181, 182, 214, 247, 248, 300, 305
  linearity of, 44, 46
  of nonintegrable signals, 42, 47–51

table of, 45, 46
table of properties of, 45, 46
Fractional dimension, 103
Frequency-domain description, 12–18
Frequency spectrum, 18
Fundamental frequency, 15, 16, 18, 21, 22, 25, 26, 41, 42, 181
Fundamental theorem of calculus, 48, 97

G
Gamma function, 139
Gauss, C.F., 52
Gibbs phenomenon, 39, 40

H
Haar wavelet, 78–82, 84, 89, 90
Harmonic analysis, xi, 323
Heat equation, 55, 299
Herglotz Theorem, 251–254
Histogram, 92
Hwang, P.Y., 325

I
Ibragimov, I.A., 325
Impulse response function
causal, 195
realizable, 195
Integrating circuit, 196, 205, 207
Inverse Fourier transform, 42, 43, 49, 71, 146, 177, 178, 181, 182, 214, 247, 248, 300, 305

K
Kallenberg, O., 160, 324
Kinetic energy, 106, 110
Kolmogorov's Theorem
on infinite sequences of r.q.s, 254
on sample path continuity, 242
Kronecker delta, 26, 156, 158
Kwapien, S., 325

L
Landau's asymptotic notation, 167
Laplace transform, 227, 323
Law of large numbers (LLN), xi, 133–136, 167
Leadbetter, M.R., 324
Least-squares fit, 129–132
Lévy process, 7
Lifshits, M.A., 325
Loève, M., 243, 324
Loparo, K., xii

M
Marginal probability distribution, 120, 121
Matching filter, 222–224
McKean, H., 323
Mean power, 36, 153, 175–177, 182, 183, 186, 190, 191, 203, 218, 264
Mexican hat wavelet, 68
Moments of r.q.s, 113–117
Montgomery, D.C., 324
Moore, J.B., 325
Morlet wavelet, 66, 68, 76
Moving average
autoregressive (ARMA), 209–210
to filter noise, 164
general, 163, 164
interpolated, 188
of white noise, 188

N
Nabla operator, 10
Noise
additive, 23, 294
white, 156–159, 161–165, 171–173, 182–184, 188, 189, 191, 197, 198, 200, 201, 203, 204, 210, 214, 219, 221–223, 226, 227, 247–249, 254, 255, 262, 263, 273, 274, 276–278, 315, 316
Normal equations, 130
Normalization condition, 71, 72, 95–98, 127, 144, 301, 303

O
Optimal filter, 225–227
Optimal predictor, 279–282, 287, 288
Orthonormal basis
of complex exponentials, 25, 29–30
in 3-D space, 25, 29–30
orthonormality, 26, 28, 29, 79, 84
of complex exponentials, 26

P
Papoulis, A., 227, 325
Parameter estimation, 108
Parseval's formula, 28, 29, 36, 44, 56, 58, 59, 90, 179, 186, 205, 226
extended, 28, 29

Passive tracer, 8
Period of the signal, 4
  infinite, 41
Periodogram, 179, 180, 189, 191
Petrov, V.V., 138
Piryatinska, A., ix, xviii, 154, 169, 180
Pitman, J., 324
Poisson distribution, 96, 144
Polarization identity, 56
Power spectral density, 177, 184, 186, 189, 263, 275, 278
Power spectrum
  cumulative, 251–254, 275
  of interpolated digital signal, 184–189
Power transfer function, 203–205, 207, 209, 210, 212–214, 312–314, 316
Priestley, M.B., 324
Probability density function (p.d.f.)
  joint of random vector, 118
  normalization condition for, 98
Probability distribution
  absolutely continuous, 96, 243
  Bernoulli, 94, 114, 124, 160
  binomial, 94–96, 144, 145
  chi-square, 108, 141–143, 146
    table of, 142
  conditional, 121–123
  continuous, 96–102, 243
  cumulative, 91, 94
  exponential, 98
  of function of r.q., 91, 117
  Gaussian (normal), 102, 116–117
    calculations with, 100
    mean and variance of, 116–117
    table of, 98–100, 102, 173, 309
  joint, 118, 121, 134, 237
  of kinetic energy, 106, 111
  marginal, 121
  mixed, 102
  n-point, 229
  Poisson, 76, 96, 98, 144, 145, 160, 161
  quantiles of, 112, 139–143
  singular, 93–113
  of square of Gaussian r.q., 102, 103, 105, 201, 229, 242
  Student-T, 138–140, 143, 146
    table of, 140
  uniform, 97–98, 115
Probability theory
  measure-theoretic, 324
  paradoxes in, 91
Pythagorean theorem, 29

Q
Quantiles of probability distribution, 139–143
  table of chi-square, 142

R
Random errors, 123
Random harmonic oscillations
  random interval, 136, 137
  superposition of, 16, 17, 155, 166, 175, 181, 247, 271, 273
Randomness, 3, 5, 9, 10, 324
  of signals, 5
Random numbers, 23, 172, 264, 294
Random phase, 155, 166, 180, 271
Random quantities (r.q.)
  absolute moments of, 114
  continuous, 92–113, 135
  correlation of, 153
  discrete, 135
  expectation of, 113–117
  function of, 108–109, 117
  linear transformation of Gaussian, 104
  moments of, 113–117
  singular, 93–113
  standard deviation of, 115, 116
  standardized, 116, 138
  statistical independence of, 119, 320
  variance of, 114, 233
Random switching signal, 160–162, 164, 184, 185
Random variable, 92, 133, 134, 138, 144, 170, 171, 247, 249, 301, 307, 308
Random vectors
  covariance matrix of, 234, 235, 238
  2-D, 235
  Gaussian, 232, 234–237, 246, 253, 257, 263, 320
  joint probability distribution of, 118, 121, 134
  linear transformation of, 230
  moments of, 124–129
Random walk, 254–256
RC filter, 199–201, 203, 204, 218, 219, 223, 227, 316
Rectangular waveform, 30–32, 34, 37, 38, 53
Regression line, 131, 132, 146
REM sleep, 154
Resolution (R), 3–5, 23, 67, 68, 78, 80–83, 86, 87, 90
Reverse conditional probability, 123
Roberts, M.I.,325

Ross, S.M., 324
Rozanov, Y.A., 325
Rudin, W., 240
Runger, G.C., 324

S
Saichev, A.I., 50, 78
Sample paths
  continuity with probability 1 of, 245
  differentiability of, 229, 239–243
  mean-square continuity of, 239–242
  mean-square differentiability of, 240, 241
Sampling period, 4, 23, 51, 209, 294
Scalar product, 26, 28, 29, 48, 222, 225, 259, 295
Scatter plot, 129, 146
Schucker, T., 323
Schwartz distributions, 50
Shakarchi, R., 323
Signals
  analog, 3–4, 11
  aperiodic, 4, 21, 41–44
  characteristics of, 11–12
  delta-correlated, 184
  description of, 3–23
  deterministic, 4, 9, 25–56, 185, 217
    spectral representation of, 25–56
  digital, 3, 11, 12, 185–189, 206–212
    discrete sampling of, 185, 189, 208
    interpolated, 185–190, 208
  Dirac-delta impulse, 47–51, 156, 162, 181–183, 194, 252, 311
  energy of, 11, 44, 106
  filtering of, 193, 226
  Gaussian, xvii, 45–46, 229–246, 256, 263, 273
    stationary, 237–239
  jointly stationary, 213
  nonintegrable, 42, 47–51
  periodic, 4, 11–19, 21, 25–44, 51, 53, 54, 225, 296, 297
  power of, 11–12, 15, 18, 20–22, 36, 154, 249
  random
    mean power of, 175
    switching, 160–162, 164–165
    types of, 3–10
  stationary
    discrete, 151, 251, 262, 269
    Gaussian, 173, 237, 244, 263
    power spectra of, xvii, 175–191
    second-order, weakly, 152, 153
    simulation of, 173, 259, 265, 273
    spectral representation of, 247–278
    strictly, 152, 172, 238
  stochastic, 151
  time average of, 11, 22, 168, 169
  transmission of binary, 123
Signal-to-noise ratio
  in matching filter, 222–224
  optimization of, 217–227
  in RC filter, 218, 219, 223, 227
Simulation
  of stationary signals, 259, 265, 273
  of white noise, 172–173
Smith, R.L., 325
Spectral representation theorem, xviii, 249, 259–263, 269, 282
Stability of fluctuations law, xvii, 133–136
Standard deviation of r.q., 115, 116, 131, 138
Stationary conditions, 5
Statistical independence, 117–132, 320
Statistical inference, 324
Stein, E.M., 323
Stochastic difference equation, 162–164, 168, 201, 210, 214
  consistency of mean estimation in, 168
Stochastic integration
  Gaussian, 256–258
  isometric property of, 258, 261
  for signals with uncorrelated increments, 254–263
Stochastic processes
  Gaussian, 258
  Lévy, 7
  stationary, 284
  Wiener, 6, 7, 256
System bandwidth, 202–206

T
Time-domain description, 12–13
Time series, xvii, xviii, 151, 163, 169, 227, 280, 284, 287–288, 324
  nonlinear, 324
  nonstationary, 324
Total probability formula, 122, 123
Trajectory, 10
Transfer function, 202–205, 207, 209, 210, 212–214, 225, 226, 312–316
Transmission
  of binary signals, 123
  in presence of random errors, 123
  through linear systems, xvii, 193
Trigonometric formulas, 13, 21, 295
Trigonometric series, 27, 323
Tukey, O.W., 52

U
Uncertainty principle, 57–90
Usoltsev, A., xviii

V
Variance of r.q.
invariance under translation of, 102, 116
quadratic scaling of, 115

W
Walden, A.T., 325
Waveform rectangular, 30–32, 34, 37, 38, 53
Wavelet transform, xvii, 57–90
White noise
band-limited, 182, 183
continuous-time, 162, 182–185
discrete time, 156–158, 162, 163, 171, 172, 184, 188, 189, 247, 262, 277
filtered, 262, 263
integrals, 249
interpolated, 188
moving average of, 157, 158, 262
simulation of, 172–173
Wiener filter
acausal, 224–226
causal, 226–227
Wiener-Hopf equation, 227
Wiener, N., 227, 324
Wiener process, 7, 259
n-point c.d.f. of, 229
Windowed Fourier transform, 57, 60–65, 89–90
Wold decomposition, 279–282, 284
Woyczynski, W.A., 5, 10, 50, 78, 96, 103, 139, 147, 169, 229, 252, 323–325

Y
Yaglom, A.M., 288, 324
Young, P.C., 325

Z
Zygmund, A., 27, 323