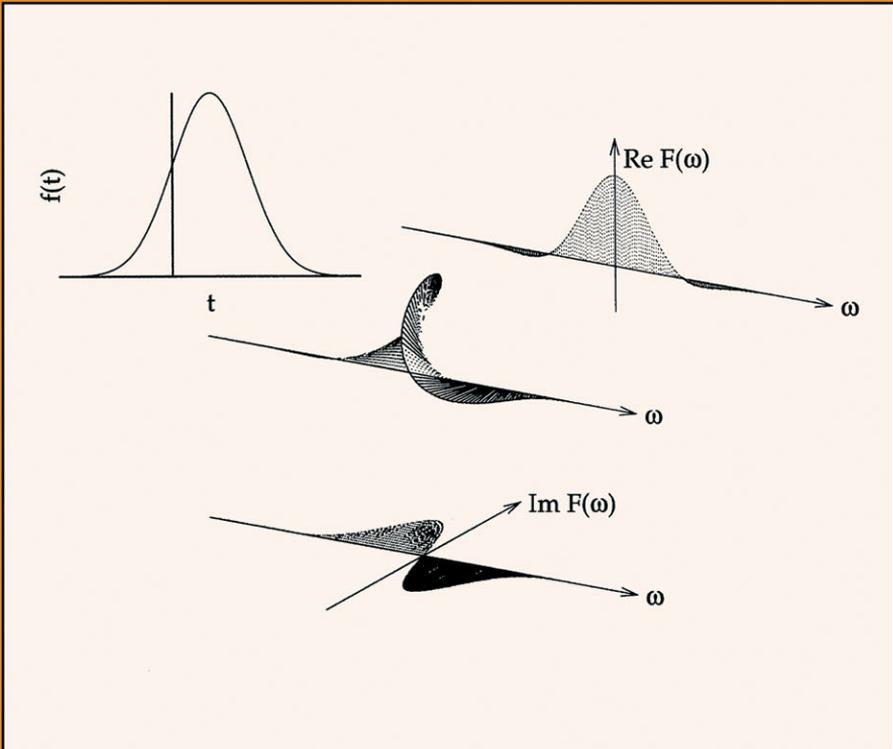


Data analysis  
and signal processing  
in chromatography

ATTILA FELINGER



DATA HANDLING IN SCIENCE AND TECHNOLOGY — VOLUME 21

Data Analysis and Signal Processing  
in Chromatography

This Page Intentionally Left Blank

DATA HANDLING IN SCIENCE AND TECHNOLOGY — VOLUME 21

**Advisory Editors:** B.G.M. Vandeginste and S.C. Rutan

# Data Analysis and Signal Processing in Chromatography

**Attila Felinger**

*Department of Analytical Chemistry, University of Veszprém, Egyetem utca 10, H-8200*

*Veszprém, Hungary*



**1998**

**ELSEVIER**

**Amsterdam — Lausanne — New York — Oxford — Shannon — Singapore — Tokyo**

ELSEVIER SCIENCE B.V.  
Sara Burgerhartstraat 25  
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Library of Congress Cataloging in Publication Data  
A catalog record from the Library of Congress has been applied for.

ISBN 0-444-82066-3

© 1998 Elsevier Science B.V. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher, Elsevier Science B.V., Copyright & Permissions Department, P.O. Box 521, 1000 AM Amsterdam, The Netherlands.

Special regulations for readers in the USA – This publication has been registered with the Copyright Clearance Center Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923. Information can be obtained from the CCC about conditions under which photocopies of parts of this publication may be made in the USA. All other copyright questions, including photocopying outside of the USA, should be referred to the publisher.

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

∞ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper)

Printed in The Netherlands

# Contents

---

<b>Preface</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Mathematical background</b>	<b>5</b>
2.1 Random variables . . . . .	5
2.1.1 Discrete variables . . . . .	6
2.1.2 Continuous variables . . . . .	7
2.1.3 Expectations . . . . .	9
2.1.4 Characteristic function . . . . .	10
2.1.5 Cumulants . . . . .	11
2.1.6 Moments . . . . .	12
2.1.7 Correlation . . . . .	14
2.2 Random processes . . . . .	17
2.2.1 Autocorrelation . . . . .	17
2.2.2 Cross-correlation . . . . .	18
2.3 Fourier analysis . . . . .	19
2.3.1 Fourier series . . . . .	20
2.3.2 Fourier integral . . . . .	22
2.3.3 Fourier transform properties . . . . .	24
2.3.4 Unit impulse and unit step . . . . .	34
2.3.5 Discrete Fourier transform . . . . .	37
2.3.6 Fast Fourier transform . . . . .	40
<b>3 Models of chromatography</b>	<b>43</b>
3.1 Plate models . . . . .	43
3.1.1 Martin-Synge plate model . . . . .	44
3.1.2 Craig plate model . . . . .	45
3.2 Mass balance model . . . . .	46
3.2.1 Equilibrium model . . . . .	48

3.2.2	Kinetic model . . . . .	49
3.3	General rate model . . . . .	51
3.4	Statistical model . . . . .	52
3.4.1	Random walk model . . . . .	52
3.4.2	Stochastic model . . . . .	53
3.4.3	Characteristic function method . . . . .	55
3.5	Asymmetrical peak shapes . . . . .	62
3.6	Nonlinear chromatography . . . . .	62
3.6.1	Houghton model . . . . .	63
3.6.2	Haarhoff-van der Linde model . . . . .	64
3.6.3	Thomas model . . . . .	67
3.7	Extra-column effects . . . . .	67
3.8	Exponentially modified Gaussian model . . . . .	69
<b>4</b>	<b>Data acquisition</b> . . . . .	<b>79</b>
4.1	Sampling theorem . . . . .	80
4.2	Errors of digitalization . . . . .	82
4.2.1	Aliasing . . . . .	82
4.2.2	Truncation error . . . . .	84
4.3	Recovery of the analog signal . . . . .	84
4.4	Practical issues of sampling analytical signals . . . . .	87
4.5	Quantization . . . . .	90
4.6	A/D converters . . . . .	93
<b>5</b>	<b>Peak shape analysis</b> . . . . .	<b>97</b>
5.1	Empirical peak shape models . . . . .	98
5.1.1	BiGaussian model . . . . .	98
5.1.2	Generalized exponential (GEX) function . . . . .	99
5.1.3	Weibull function . . . . .	100
5.1.4	Chesler-Cram model . . . . .	101
5.1.5	Lognormal function . . . . .	102
5.1.6	Fraser-Suzuki model . . . . .	104
5.1.7	Gram-Charlier and Edgeworth-Cramér series . . . . .	104
5.1.8	Other asymmetrical peak shapes . . . . .	110
5.2	Curve fitting to the empirical models . . . . .	113
5.3	Characterization of peak asymmetry . . . . .	114
5.3.1	Asymmetry factor . . . . .	114
5.3.2	Tailing factor . . . . .	117
5.4	Comparison of peak shapes . . . . .	119
5.4.1	Lévy distance . . . . .	119
5.4.2	Distribution function method . . . . .	119

<b>6</b>	<b>Noise</b>	<b>125</b>
6.1	Noise types and sources . . . . .	125
6.1.1	Thermal noise . . . . .	126
6.1.2	Shot noise . . . . .	126
6.1.3	Flicker noise . . . . .	127
6.1.4	Burst noise . . . . .	128
6.1.5	Environmental noise . . . . .	128
6.2	The observed noise . . . . .	129
6.3	Noise in analytical signals . . . . .	131
6.4	Noise profile of chromatographic detectors . . . . .	131
6.5	Signal-to-noise ratio . . . . .	137
6.6	Computer simulation of noise . . . . .	139
<b>7</b>	<b>Signal enhancement</b>	<b>143</b>
7.1	Digital filters . . . . .	143
7.2	Filtering in time domain . . . . .	145
7.2.1	Moving window techniques . . . . .	145
7.2.2	Recursive filtering . . . . .	160
7.3	Ensemble averaging . . . . .	162
7.4	Correlation techniques . . . . .	164
7.4.1	Correlation chromatography . . . . .	166
7.5	Fourier-domain filtering . . . . .	167
7.5.1	Smoothing windows . . . . .	169
7.5.2	Determination of the cut-off frequency . . . . .	171
7.5.3	Optimum filtering . . . . .	173
7.6	Kalman filter . . . . .	176
7.6.1	Smoothing with the linear Kalman filter . . . . .	178
<b>8</b>	<b>Peak detection</b>	<b>183</b>
8.1	Matched filters . . . . .	183
8.2	Peak detection by derivatives . . . . .	185
<b>9</b>	<b>Quantitative analysis</b>	<b>191</b>
9.1	Graphical integration of stand-alone peaks . . . . .	191
9.1.1	The "peak height $\times$ width at half height" method . . . . .	192
9.1.2	The Condal-Bosch method . . . . .	193
9.1.3	Triangulation . . . . .	194
9.1.4	Integration of asymmetrical peaks . . . . .	194
9.2	Graphical integration of overlapping peaks . . . . .	197
9.2.1	Overlapping asymmetrical peaks . . . . .	199
9.3	Numerical measurement of peak height and peak area . . . . .	202
9.3.1	Peak height and retention time measurement . . . . .	202
9.3.2	Area measurement . . . . .	202
9.4	Accuracy and precision of the measured values . . . . .	203

<b>10 Resolution enhancement</b>	<b>211</b>
10.1 Peak resolution	212
10.1.1 Shoulder limit	216
10.2 Resolution of overlapping peaks	219
10.2.1 Regression analysis	219
10.2.2 Curve fitting	222
10.2.3 Nonlinear regression	225
10.2.4 Effect of noise	228
10.2.5 Kalman filtering	229
10.2.6 Parameter estimation in frequency domain	233
10.3 Peak sharpening	237
10.3.1 Fourier deconvolution	241
10.3.2 Iterative deconvolution	248
10.3.3 Decomposition of superposition of density functions	253
10.4 Moment analysis of overlapping peaks	254
10.5 Deconvolution using natural computation	256
<b>11 Multivariate signal resolution</b>	<b>261</b>
11.1 Factor analysis	262
11.1.1 Abstract factor analysis	262
11.1.2 Principal component analysis	264
11.1.3 Singular value decomposition	265
11.1.4 Target transformation	266
11.1.5 Self modeling curve resolution	266
11.1.6 Evolutionary factor analysis	268
11.1.7 Rank annihilation	270
11.1.8 Alternating regression	272
11.1.9 Further applications of factor analysis in chromatography	272
11.2 Spectral and chromatographic ratioing	272
11.2.1 Spectral ratioing	272
11.2.2 Spectral suppression	274
11.2.3 Chromatographic ratioing	274
<b>12 Transform techniques in chromatography</b>	<b>279</b>
12.1 Fourier transform	279
12.1.1 Numerical aspects of Fourier transform	280
12.1.2 Interpolation	282
12.1.3 Differentiation, integration	284
12.1.4 Power spectrum, autocovariance	286
12.1.5 Data compression	286
12.2 Other transforms	287
12.2.1 Laplace transform	287
12.2.2 Hartley transform	287
12.2.3 Wavelet transform	289

<b>13 Information theory</b>	<b>299</b>
13.1 Entropy and information . . . . .	299
13.2 Entropy of random variables . . . . .	300
13.3 Information transfer through channels . . . . .	301
13.4 Information content of quantitative analysis . . . . .	302
13.4.1 Informing power . . . . .	302
13.4.2 Function of mutual information . . . . .	304
13.5 Maximum entropy method . . . . .	304
<b>14 Quality assurance and validation</b>	<b>307</b>
14.1 Errors associated with analytical measurements . . . . .	307
14.1.1 Hypothesis tests . . . . .	308
14.1.2 Random errors . . . . .	309
14.1.3 Systematic error . . . . .	311
14.1.4 Error propagation . . . . .	314
14.1.5 Requirement for normality . . . . .	315
14.2 Calibration . . . . .	316
14.2.1 Weighted regression . . . . .	319
14.2.2 Linear regression when both variables are subject to errors . . . . .	321
14.2.3 Residual analysis . . . . .	321
14.3 Limit of detection . . . . .	321
14.4 Validation of new analytical methods . . . . .	322
14.5 Validation of the chromatographic system . . . . .	323
14.5.1 Validation of the chromatographic instrument . . . . .	324
14.5.2 Validation of the chromatographic method . . . . .	327
<b>15 Statistical theory of peak overlap</b>	<b>331</b>
15.1 Peak resolution in multicomponent chromatograms . . . . .	333
15.2 Determination limit . . . . .	338
15.3 Combinatorial analysis of peak overlap . . . . .	340
15.4 Random retention pattern . . . . .	341
15.4.1 Poisson process . . . . .	344
15.4.2 Varying peak density . . . . .	350
15.4.3 Isocratic multicomponent separations . . . . .	352
15.4.4 Statistical model of overlap for limited number of components . . . . .	354
15.4.5 Uncertainty in statistical model of overlap . . . . .	356
15.4.6 Other retention patterns . . . . .	357
15.4.7 Effect of peak amplitudes . . . . .	360
15.4.8 Poorly resolved multicomponent separations . . . . .	364
15.5 Superposition of retention patterns . . . . .	366
15.6 Recognition of the retention pattern . . . . .	371
15.7 Estimating the number of compound families . . . . .	374
15.8 Multidimensional separations . . . . .	374
15.9 Order-disorder in multicomponent mixtures . . . . .	378

---

<b>16 Fourier analysis of multicomponent chromatograms</b>	<b>385</b>
16.1 Power spectrum of multicomponent chromatograms . . . . .	385
16.2 Power spectrum of a Poissonian chromatogram . . . . .	389
16.3 Power spectrum for constant peak width . . . . .	390
16.4 Power spectrum for varying peak width . . . . .	391
16.5 Autocovariance function of multicomponent chromatograms . . . . .	393
16.5.1 Autocovariance function of a Poissonian chromatogram	394
16.6 Parameter estimation . . . . .	398
16.7 Determination of the retention pattern by Fourier analysis . .	400
16.8 Simplified Fourier analysis . . . . .	403
16.9 Determination of the proper sample size by Fourier analysis .	406
<b>Index</b>	<b>410</b>

# Preface

---

Nowadays substantial time of the analytical chemists is spent in front of their computer comprehending the data their instruments produced. A continuously broadening spectrum of mathematical and statistical methods is applied for data evaluation in analytical chemistry.

Numerical analysis of chromatograms is in the focal point of this volume, but most of the chemometric methods described here are not specific to chromatography and related separation techniques.

I have been guided by the goal of giving explanation of those problems that are cardinal in the numerical treatment of chromatograms. Thus, most emphasis is given to peak shape analysis, signal enhancement, and resolution enhancement to strengthen both quantitative and qualitative analyses.

This book was typeset by the author with the public-domain  $\text{\LaTeX}$   $\text{\LaTeX}$  typesetting package. Most figures were produced by the also public-domain *gnuplot* package. I appreciate that the developers of  $\text{\TeX}$  and  $\text{\LaTeX}$  packages made these outstanding pieces of software publicly accessible. It was a delight for me that the anonymous  $\text{\TeX}$  community out there offered solutions to all the problems I struggled with when "setting the type."

I thank FRANCESCO DONDI for his encouragement to launch the project of writing this book; we started to establish the outline of this volume together, and he always followed the development of this work.

I am greatly indebted to GEORGES GUIOCHON for his valuable comments and help; I appreciate that during my stayings in his laboratory at the University of Tennessee, I could spend countless hours in the fabulous Hodges Library collecting material for this book.

My endeavor to write a book on the signal processing techniques applied in chromatography started as a two-year project—four years ago. On one hand, my pace let me include some very fresh developments published in this field, therefore the reader may even benefit from this. On the other hand, I couldn't agree more with the following lines of A. A. Milne:

The fact is this is more difficult  
than I thought,  
I ought—  
*(Very good indeed)*  
I ought  
To begin again  
But it is easier  
To stop.  
(The House at Pooh Corner)

Veszprém, March, 1998

Attila Felinger

# 1

---

## Introduction

If we believe that *The Purpose of Computing is Insight, Not Numbers*, then it follows that the man who is to get the insight must understand the computing.  
... It is sometimes suggested that the motto be revised to,  
The Purpose of Computing Numbers is Not Yet in Sight.  
(R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Dover, 1973)

Signal processing and data analysis are of increasing importance in analytical chemistry. During the past decades, chemometric techniques encountered an enormous development. The computer-connected modern analytical instruments available nowadays offer an enormous, easy-access possibility to enhance the information available from the zillions of recorded data.

Although commercial software packages offer a variety of data manipulation algorithms, users often regard these software packages as black boxes. Because vendors do not feel like supplying detailed information on how the software packages do whatever they do, the analyst has to trust that the manipulated data are reliable. Therefore, it is essential that one realize how information beyond conventional results can be deduced from chromatograms and other instrumental signals.

The goal of this work is to summarize what types of data and signal analysis techniques can be applied to chromatograms, and related instrumental signals. In this respect, not the full spectrum of chemometrics is discussed in detail here, while some other techniques that do not really fit into the definition of chemometrics<sup>1</sup>, but are important in chromatographic data analysis are also addressed.

---

<sup>1</sup> Note that there exist several definitions for chemometrics. At birth, the following definition was given by Wold "The art of extracting chemically relevant information from data produced in chemical experiments is given the name of *chemometrics*" [1].

According to Massart [2], chemometrics is the chemical discipline that uses mathematical, statistical and other methods employing formal logic (*i*) to design or select optimal measurement procedures and experiments, and (*ii*) to provide maximum rel-

The different topics of data analysis and signal processing are discussed in this book according to the following structure.

In Chapter 2, some preliminary mathematics, necessary to conceive the fundamentals of data analysis, is summarized.

The physical-chemical background of the chromatographic process is explained in Chapter 3, by means of different models of chromatography. It is important to understand what the most common peak shapes and the reasons of peak asymmetry are in order to exploit information by peak shape analysis.

Chapter 4 discusses how the analog signal supplied by the detector can be translated into a digital signal that can be stored and processed by a computer.

Besides the physical models of chromatography, a multitude of empirical peak shape models are available; those models, peak shape analysis, and the characterization of peak asymmetry are the subject of Chapter 5.

In Chapter 6, the origin and characteristics of the baseline noise, as well as different definitions for signal-to-noise ratio are discussed.

Chapter 7 enlists the methods of signal enhancement: filtering in time and in Fourier domains are both very significant. Different techniques are compared for filtering high- and low-frequency noise.

Detection of chromatographic peaks, *i.e.* the location of peaks; determination of the precise starting point and end point of each peak must be emphasized especially in the case of tailing peaks. This is the topic of Chapter 8.

The simple methods of peak area determination are based on graphical means. The discussion of the different graphical and numerical integration strategies gives hints on using the integrator software. The precision and accuracy of quantitative analysis is also addressed in Chapter 9.

The recognition of overlapping peaks, the limit resolutions (shoulder and detectability limits) are important in order to identify fused peaks. Resolution techniques as curve fitting or deconvolution of band broadening effects are described in Chapter 10. Those methods can simply be used to isolate the contribution of different phenomena to peak broadening and distortion. Deconvolution of overlapping peaks are more advanced numerical techniques to quantify overlapping peaks. Fourier domain deconvolution, constrained iterative relaxation, curve fitting, Kalman filtering and other methods are explained.

Multivariate techniques of signal processing have been rapidly developing since the availability of diode array detectors and other hyphenated techniques. Software packages using them can be found in many commercial instruments. Multivariate techniques are rather useful in detecting peak homogeneity or minor impurities. In Chapter 11, factor analysis and other multivariate methods are addressed.

Chapter 12 reveals some numerical tricks for those who are interested in numerical errantry with Fourier transforms. Transform techniques other than

---

evant chemical information by analyzing chemical data.

The International Chemometrics Society offers the following definition: "Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods."

Fourier are also discussed in that chapter.

The application of information theory and the concept of entropy are the subject of Chapter 13.

Quality issues are addressed in Chapter 14, with the description of the errors of analysis. Validation of the chromatographic instrument and method is also summarized.

The last part of the book is devoted to multicomponent chromatograms. By the help of the statistical theory of peak overlap, the probability of having singlets, doublets, etc. can be judged, which is very important during method development. On the other hand, from a single chromatogram, we can estimate the number of detectable components. These novel techniques are the subject of Chapter 15.

Further methods for the analysis of multicomponent chromatograms—by Fourier analysis—are described in Chapter 16. Utilizing the power spectrum of multicomponent chromatograms—besides the number of components—the peak shape parameters, and also the retention pattern can be determined. This latter is very useful in determining the amount of order and disorder in the sample.

## Bibliography

- [1] WOLD, S., Chemometrics; What Do We Mean with It, and What Do We Want from It?, *Chemometr. Intell. Lab. Syst.* **1995**, *30*, 109-115.
- [2] MASSART, D. L.; VANDEGINSTE, B. G. M.; DEMING, S. N.; MICHOTTE, Y.; KAUFMAN, L. *Chemometrics: a Textbook*, Elsevier: Amsterdam, 1988.

This Page Intentionally Left Blank

# 2

---

## Mathematical background

There is a famous formula—perhaps the most compact and famous of all formulas—developed by Euler from a discovery of De Moivre:  $e^{i\pi} + 1 = 0$ . . . It appeals equally to the mystic, the scientist, the philosopher, the mathematician.  
(Edward Kasner and James Newman, *Mathematics and the Imagination*, 1940)

In this chapter, the fundamental mathematics of signal and data analysis is considered. It is beyond the scope of this volume to give an extensive explanation of the complete mathematical background of signal and data analysis. Detailed description of the mathematical treatment can be found in mathematical and statistical textbooks<sup>1</sup>. Only a summary of the necessary mathematical principles is addressed here.

Most analytical signals are of transient type, or they can simply be transformed into a transient signal by integration or differentiation. Therefore, various aspects of probability theory are widely applied to describe peak shapes, statistical moments are used to characterize peak asymmetry. Accordingly, the properties of random variables and random signals, as well as the fundamentals of Fourier analysis are the subject of this chapter.

### 2.1 Random variables

The value of a random variable cannot be predicted, only a probability of the event can be given that the random variable takes a value. To characterize a random variable, first we have to know what values the random variable can take. Furthermore, we have to know how often a given value is taken by the

---

<sup>1</sup>See, for instance, references [1-5] for excellent introduction into probability theory. Detailed explanation of the Fourier methods can be found in [6-9]. Random processes and stochastic methods are addressed, for example, in [10-13].

random variable. Distribution functions describe what the probability of a given value is.

$P(X = a)$  denotes the probability that the random variable  $X$  takes the numerical value  $a$ . The probability *distribution function* of  $X$  describes the following probability

$$F(x) = P(X \leq x) \quad (2.1)$$

$F(x)$  is also known as cumulative distribution function. *In order to completely describe a random variable, we have to know its probability distribution function.*

The probability that  $a \leq X \leq b$  is calculated by

$$P(a \leq X \leq b) = F(b) - F(a) \quad (2.2)$$

Since probabilities are nonnegative numbers, if  $a < b$  the following inequality will hold

$$F(a) \leq F(b) \quad (2.3)$$

Accordingly, the distribution function of a random variable is a nondecreasing function. Furthermore, the following limits can be established for a distribution function

$$0 \leq F(x) \leq 1 \quad (2.4)$$

being  $F(-\infty) = 0$  and  $F(\infty) = 1$ .

If the random variable can only take a finite number of numerical values, it is a *discrete random variable*, otherwise it is a *continuous random variable*. A discrete random variable is for instance the number of detectable peaks in a multicomponent separation. The area of a given peak in repeated analyses is a continuous random variable.

### 2.1.1 Discrete variables

Let us assume that a discrete random variable  $X$  can take the  $x_1, x_2, \dots, x_n$  values with the probability of  $p_1, p_2, \dots, p_n$ , i.e.

$$P(X = x_i) = p_i \quad i = 1, 2, \dots, n \quad (2.5)$$

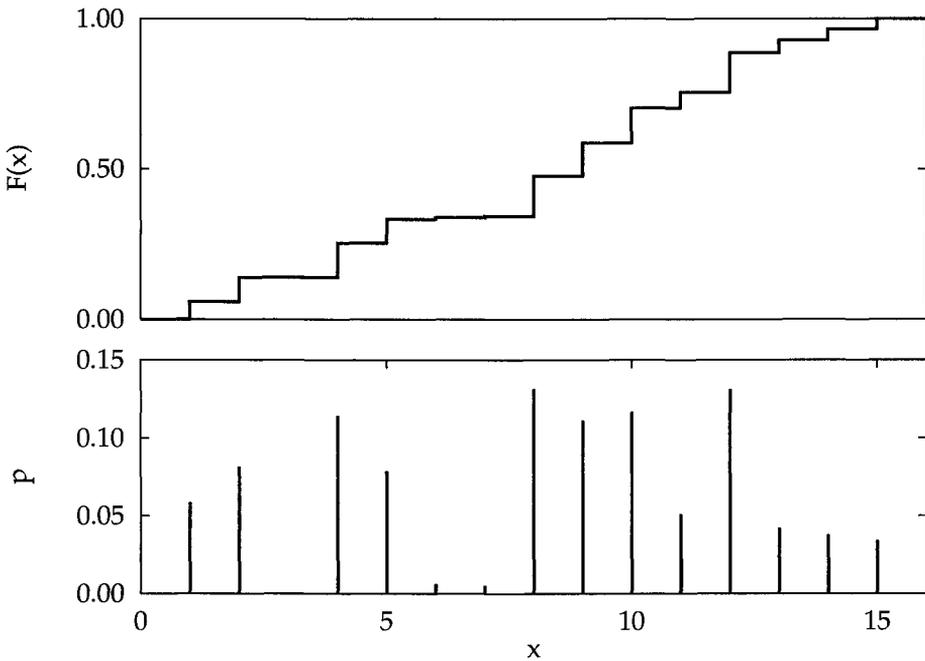
The distribution function of a discrete variable is defined as

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i \quad (2.6)$$

From the above equation follows that the total probability for all the  $n$  discrete values is unity

$$\sum_{i=1}^n p_i = 1 \quad (2.7)$$

The distribution function of a discrete random variable is a step function. It is constant for any interval not containing an  $x_i$ , and the height of the step at  $x_i$  is  $p_i$ . Figure 2.1 shows the distribution function and the frequency function of a discrete variable.



**FIGURE 2.1** Distribution function (upper) and frequency function (lower) of a discrete variable.

### 2.1.2 Continuous variables

A random variable is continuous if the probability distribution can be described by a function  $f(x)$  which is everywhere continuous. The probability of the event that the random variable takes a value within the infinitesimal narrow interval  $(x, x + dx)$  is  $f(x) dx$ .  $f(x)$  is the *frequency function*, or *probability density function* of the random variable.

The probability that the random variable  $X$  takes a value between  $a \leq X \leq b$  is

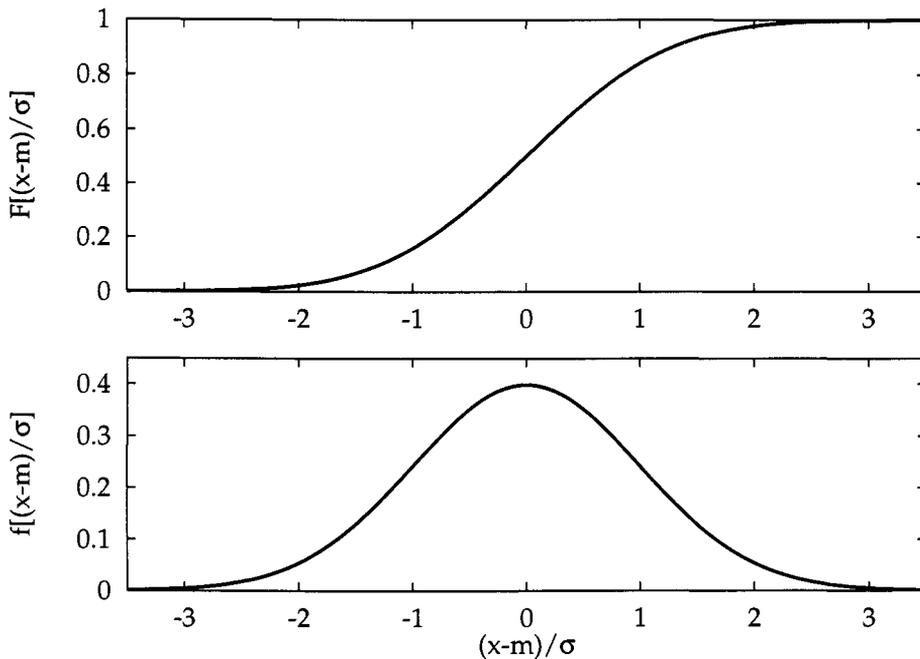
$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x) dx \quad (2.8)$$

From the above equation it follows that

$$F(b) = \int_{-\infty}^b f(x) dx \quad (2.9)$$

From which by differentiation we get

$$\frac{dF(x)}{dx} = f(x) \quad (2.10)$$



**FIGURE 2.2** Distribution function (upper) and frequency function (lower) of the normal distribution.

The probability density function is the derivative of the cumulative distribution function. The probability  $P(-\infty < X < \infty)$ , on the other hand, can be calculated as

$$P(-\infty < X < \infty) = F(\infty) - F(-\infty) = \int_{-\infty}^{\infty} f(x) dx = 1 \quad (2.11)$$

### Normal distribution

The normal distribution plays a central role in probability theory. Its probability density function is the Gaussian curve, which is the cardinal peak shape model in chromatography. For these reasons we shall demonstrate the properties of the continuous random variables by the example of the normal distribution.

The distribution function of a normal random variable is

$$F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \frac{x - m}{\sqrt{2}\sigma} \quad (2.12)$$

By differentiating Equation 2.12, the probability density function is obtained

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (2.13)$$

### 2.1.3 Expectations

The *mathematical expectation*, or mean value of the random variable  $X$  is denoted by  $E\{X\}$ . For a discrete random variable  $X$ , which can take the values  $x_1, x_2, \dots, x_n$  with the probability of  $p_1, p_2, \dots, p_n$ , it is calculated as

$$E\{X\} = \sum_{i=1}^n x_i p_i \quad (2.14)$$

For a continuous random variable, the expectation is defined by the integral

$$E\{X\} = \int_{-\infty}^{\infty} x f(x) dx \quad (2.15)$$

where  $f(x)$  is the probability density function of the random variable  $X$ .

For the expected value of the normal distribution, Equation 2.15 yields  $E\{X\} = m$ .

If  $X$  is a random variable,  $g(X)$  is a random variable, too. The expectation of  $g(X)$  is defined by the following equations for discrete and continuous random variables, respectively

$$E\{g(X)\} = \sum_{i=1}^n p_i g(x_i) \quad (2.16)$$

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (2.17)$$

A special expectation is the *variance*, for the calculation of which  $g(X) = (X - E\{X\})^2$  is used in Equation 2.17

$$D^2(X) = \int_{-\infty}^{\infty} (x - E\{X\})^2 f(x) dx \quad (2.18)$$

The variance can also be expressed as

$$D^2(X) = E\{X^2\} - (E\{X\})^2 \quad (2.19)$$

The variance is often denoted by  $\sigma^2$ . The square root of the variance is  $D\{X\}$ , or  $\sigma$ , which is known as *standard deviation*.

When a random variable  $X$  is the linear combination of several random variables

$$X = a_1 X_1 + a_2 X_2 + \dots + a_n X_n \quad (2.20)$$