Dunedin
New Zealand

ISBN: 978-3-540-70712-7

Library of Congress Control Number: 2008936877

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

This fourth edition of *Stochastic Methods* is thoroughly revised and augmented, and has been completely reset. While keeping to the spirit of the book I wrote originally, I have reorganised the chapters of Fokker-Planck equations and those on approximation methods, and introduced new material on the white noise limit of driven stochastic systems, and on applications and validity of simulation methods based on the Poisson representation. Further, in response to the revolution in financial markets following from the discovery by Fischer Black and Myron Scholes of a reliable option pricing formula, I have written a chapter on the application of stochastic methods to financial markets. In doing this, I have not restricted myself to the geometric Brownian motion model, but have also attempted to give some flavour of the kinds of methods used to take account of the realities of financial markets. This means that I have also given a treatment of Lévy processes and their applications to finance, since these are central to most current thinking.

Since this book was written the rigorous mathematical formulation of stochastic processes has developed considerably, most particularly towards greater precision and generality, and this has been reflected in the way the subject is presented in modern applications, particularly in finance. Nevertheless, I have decided to adhere to my original decision, to use relatively simple language without excessive rigour; indeed I am not convinced that the increase in rigour and precision has been of significant help to those who want to use stochastic methods as a practical tool.

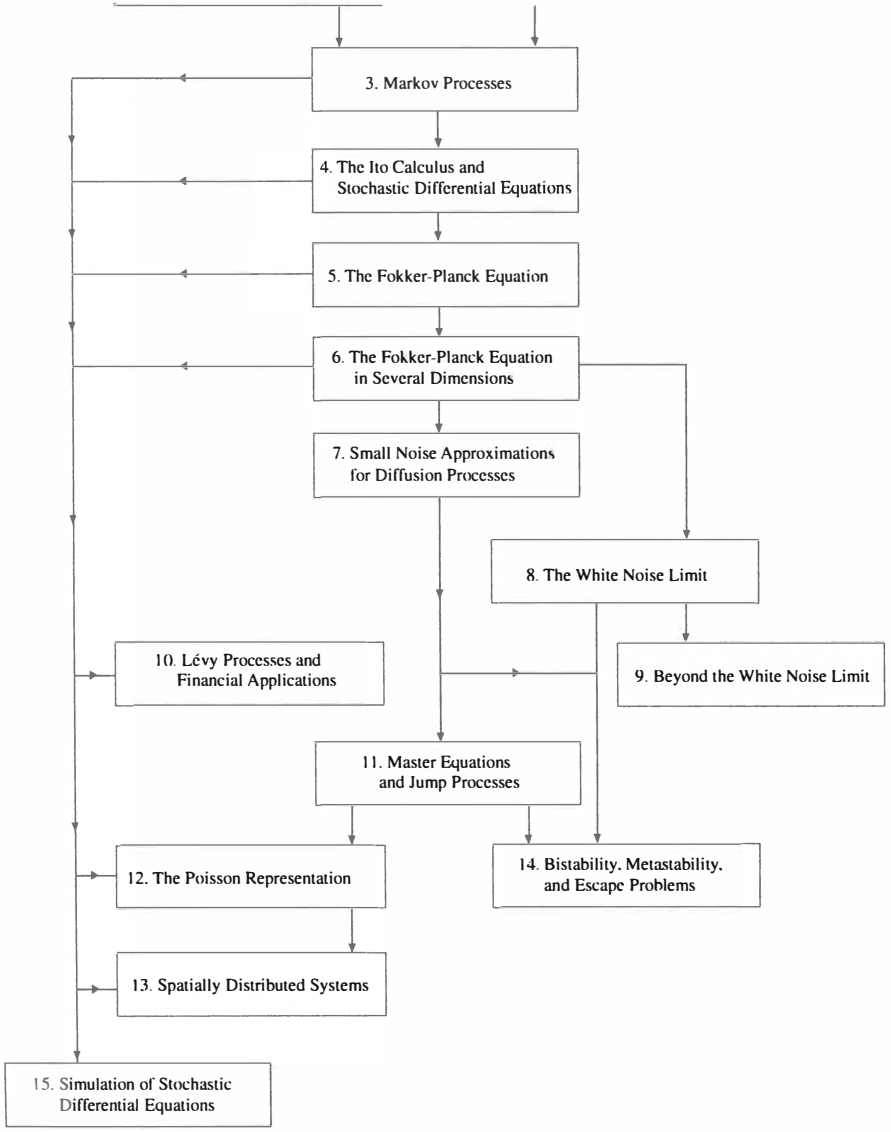
The new organisation of the material in the book is as in the figure on the next page. Instead of the original ten chapters, there are now fifteen. Some of the increase is a result of my decision to divide up some of the larger chapters into tighter and more logically structured smaller chapters, but Chapters 8 and 10 are completely new. The basic structure of the book is much the same, building on the basis of Ito stochastic differential equations, and then extending into Fokker-Planck equations and jump processes. I have put all of the work on the Poisson representation into a single chapter, and augmented this chapter with new material.

Stochastic Methods, although originally conceived as a book for physicists, chemists and similar scientists, has developed a readership with far more varied tastes, and this new edition is designed to cater better for the wider readership, as well as to those I originally had in mind. At the same time, I have tried hard to maintain "look and feel" of the original, and the same degree of accessibility.

University of Otago, New Zealand *C.W. Gardiner*
July, 2008

From the Preface to the First Edition

My intention in writing this book was to put down in relatively simple language and in a reasonably deductive form, all those formulae and methods which have been



ful perusal of these soon shows that their aim does not coincide with mine. There are purely theoretical and highly mathematical books, there are books related to electrical engineering or communication theory, and there are books for biologists—many of them very good, but none of them covering the kind of applications that appear nowadays so frequently in Statistical Physics, Physical Chemistry, Quantum Optics and Electronics, and a host of other theoretical subjects.

The main new point of view here is the amount of space which deals with methods of approximating problems, or transforming them for the purpose of approximating them. I am fully aware that many workers will not see their methods here. But my criterion here has been whether an approximation is *systematic*. Many approximations are based on unjustifiable or uncontrollable assumptions, and are justified *a posteriori*. Such approximations are not the subject of a systematic book—at least, not until they are properly formulated, and their range of validity controlled. In some cases I have been able to put certain approximations on a systematic basis, and they appear here—in other cases I have not.

A word on the background assumed. The reader must have a good knowledge of practical calculus including contour integration, matrix algebra, differential equations, both ordinary and partial, at the level expected of a first degree in applied mathematics, physics or theoretical chemistry.

I expect the readership to consist mainly of theoretical physicists and chemists, and thus the general standard is that of these people. This is not a rigorous book in the mathematical sense, but it contains results, all of which I am confident are provable rigorously, and whose proofs can be developed out of the demonstrations given. The organisation of the book is as in the following table, and might raise some eyebrows. For, after introducing the general properties of Markov processes, I have chosen to base the treatment on the conceptually difficult but intuitively appealing concept of the stochastic differential equation. I do this because of my own experience of the simplicity of stochastic differential equation methods, once one has become familiar with the Ito calculus, which I have presented in Chap. 4 in a rather straightforward manner, such as I have not seen in any previous text.

For the sake of compactness and simplicity I have normally presented only one way of formulating certain methods. For example, there are several different ways of formulating the adiabatic elimination results, though few have been used in this context. To have given a survey of all formulations would have required an enormous and almost unreadable book. However, where appropriate I have included specific references, and further relevant matter can be found in the general bibliography.

Hamilton, New Zealand
January, 1983

C.W. Gardiner

1.1 Motivation

Theoretical science up to the end of the nineteenth century can be viewed as the study of solutions of differential equations and the modelling of natural phenomena by deterministic solutions of these differential equations. It was at that time commonly thought that if all initial data could only be collected, one would be able to predict the future with certainty.

We now know this is not so, in at least two ways. Firstly, the advent of quantum mechanics within a quarter of a century gave rise to a new physics, and hence a new theoretical basis for all science, which had as an essential basis a purely statistical element. Secondly, more recently, the concept of chaos has arisen, in which even quite simple differential equation systems have the rather alarming property of giving rise to essentially unpredictable behaviour. To be sure, one can predict the future of such a system given its initial conditions, but any error in the initial conditions is so rapidly magnified that no practical predictability is left. In fact, the existence of chaos is really not surprising, since it agrees with more of our everyday experience than does pure predictability—but it is surprising perhaps that it has taken so long for the point to be made.

Chaos and quantum mechanics are not the subject of this chapter. Here I wish to give a semihistorical outline of how a phenomenological theory of fluctuating phenomena arose and what its essential points are. The very usefulness of predictable models indicates that life is not entirely chaos. But there is a limit to predictability, and what we shall be most concerned with in this book are models of limited predictability. The experience of careful measurements in science normally gives us data like that of Fig. 1.1, representing the growth of the number of molecules of a

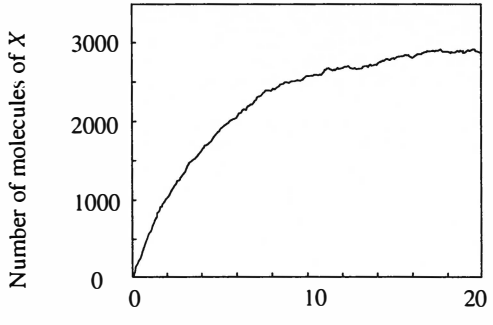


Fig. 1.1. Stochastic simulation of an isomerisation reaction $X \rightleftharpoons A$

around this motion, which are not.

1.2 Some Historical Examples

1.2.1 Brownian Motion

The observation that, when suspended in water, small pollen grains are found to be in a very animated and irregular state of motion, was first systematically investigated by Robert Brown in 1827 [1.1], and the observed phenomenon took the name *Brownian Motion* because of his fundamental pioneering work. Brown was botanist—indeed a very famous botanist—and he was examining pollen grains in order to elucidate the mechanism which by which the grains moved towards the ova when fertilising flowers. At first he thought this motion was a manifestation of life he was seeking, but when he found that this motion was present in apparently dead pollen, some over a century old, some even extracted from fossils, and then even in any suspension of fine particles—glass, minerals and even a fragment of the sphinx—he ruled out any specifically organic origin of this motion. The motion is illustrated in Fig. 1.2.

The riddle of Brownian motion was not quickly solved, and a satisfactory explanation did not come until 1905, when *Einstein* published an explanation under the rather modest title "Über die von der molekular-kinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen" (concerning the motion, as required by the molecular-kinetic theory of heat, of particles suspended in liquids at rest) [1.2]. The same explanation was independently developed by *Smoluchowski* [1.3], who was responsible for much of the later systematic development and for much of the experimental verification of Brownian motion theory.

There were two major points in Einstein's solution to the problem of Brownian motion.

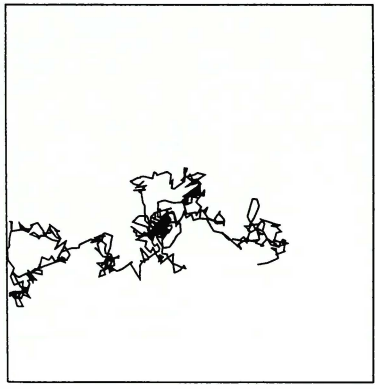


Fig. 1.2. Motion of a point undergoing Brownian motion

ii) The motion of these molecules is so complicated that its effect on the pollen grain can only be described probabilistically in terms of exceedingly frequent statistically independent impacts.

The existence of fluctuations like these ones calls out for a statistical explanation of this kind of phenomenon. Statistics had already been used by Maxwell and Boltzmann in their famous gas theories, but only as a description of possible states and the likelihood of their achievement and not as an intrinsic part of the time evolution of the system. *Rayleigh* [1.4] was in fact the first to consider a statistical description in this context, but for one reason or another, very little arose out of his work. For practical purposes, Einstein's explanation of the nature of Brownian motion must be regarded as the beginning of stochastic modelling of natural phenomena.

Einstein's reasoning is very clear and elegant. It contains all the basic concepts which will make up the subject matter of this book. Rather than paraphrase a classic piece of work, I shall simply give an extended excerpt from Einstein's paper (author's translation):

> "It must clearly be assumed that each individual particle executes a motion which is independent of the motions of all other particles; it will also be considered that the movements of one and the same particle in different time intervals are independent processes, as long as these time intervals are not chosen too small.
>
> "We introduce a time interval τ into consideration, which is very small compared to the observable time intervals, but nevertheless so large that in two successive time intervals τ, the motions executed by the particle can be thought of as events which are independent of each other.
>
> "Now let there be a total of n particles suspended in a liquid. In a time interval τ, the X-coordinates of the individual particles will increase by an amount Δ, where for each particle Δ has a different (positive or negative) value. There will be a certain *frequency law* for Δ; the number dn of the particles which experience a shift which is between Δ and $\Delta + d\Delta$ will be expressible by an equation of the form
>
> $$dn = n\phi(\Delta)d\Delta\,, \tag{1.2.1}$$
>
> where
>
> $$\int_{-\infty}^{\infty} \phi(\Delta)d\Delta = 1\,, \tag{1.2.2}$$
>
> and ϕ is only different from zero for very small values of Δ, and satisfies the condition
>
> $$\phi(\Delta) = \phi(-\Delta)\,. \tag{1.2.3}$$
>
> "We now investigate how the diffusion coefficient depends on ϕ. We shall once more restrict ourselves to the case where the number ν of particles per unit volume depends only on x and t.

"Let $\nu = f(x,t)$ be the number of particles per unit volume. We compute the distribution of particles at the time $t+\tau$ from the distribution at time t. From the definition of the function $\phi(\Delta)$, it is easy to find the number of particles which at time $t+\tau$ are found between two planes perpendicular to the x-axis and passing through points x and $x+dx$. One obtains

$$f(x,t+\tau)dx = dx \int_{-\infty}^{\infty} f(x+\Delta,t)\phi(\Delta)d\Delta \,. \tag{1.2.4}$$

But since τ is very small, we can set

$$f(x,t+\tau) = f(x,t) + \tau\frac{\partial f}{\partial t} \,. \tag{1.2.5}$$

Furthermore, we develop $f(x+\Delta,t)$ in powers of Δ:

$$f(x+\Delta,t) = f(x,t) + \Delta\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!}\frac{\partial^2 f(x,t)}{\partial x^2} + \cdots \,. \tag{1.2.6}$$

We can use this series under the integral, because only small values of Δ contribute to this equation. We obtain

$$f + \frac{\partial f}{\partial \tau}\tau = f\int_{-\infty}^{\infty}\phi(\Delta)d\Delta + \frac{\partial f}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta + \frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta \,. \tag{1.2.7}$$

Because $\phi(x) = \phi(-x)$, the second, fourth, etc., terms on the right-hand side vanish, while out of the 1st, 3rd, 5th, etc., terms, each one is very small compared with the previous. We obtain from this equation, by taking into consideration

$$\int_{-\infty}^{\infty}\phi(\Delta)d\Delta = 1 \,, \tag{1.2.8}$$

and setting

$$\frac{1}{\tau}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta = D \,, \tag{1.2.9}$$

and keeping only the 1st and third terms of the right-hand side,

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}\cdots \,. \tag{1.2.10}$$

This is already known as the differential equation of diffusion and it can be seen that D is the diffusion coefficient. ...

"The problem, which corresponds to the problem of diffusion from a single point (neglecting the interaction between the diffusing particles), is now completely determined mathematically: its solution is

$$f(x,t) = \frac{n}{\sqrt{4\pi D}} = \frac{e^{-x^2/4Dt}}{\sqrt{t}}\cdots . \tag{1.2.11}$$

"We now calculate, with the help of this equation, the displacement λ_x in the direction of the X-axis that a particle experiences on the average or, more exactly, the square root of the arithmetic mean of the square of the displacement in the direction of the X-axis; it is

$$\lambda_x = \sqrt{\bar{x}^2} = \sqrt{2Dt} \,." \tag{1.2.12}$$

Einstein's derivation is really based on a discrete time assumption, that impacts happen only at times $0, \tau, 2\tau, 3\tau \ldots$, and his resulting equation (1.2.10) for the distribution function $f(x,t)$ and its solution (1.2.11) are to be regarded as approximations, in which τ is considered so small that t may be considered as being continuous. Nevertheless, his description contains very many of the major concepts which have been developed more and more generally and rigorously since then, and which will be central to this book. For example:

i) *The Chapman-Kolmogorov Equation* occurs as Einstein's equation (1.2.4). It states that the probability of the particle being at point x at time $t+\tau$ is given by the sum of the probability of all possible "pushes" Δ from positions $x+\Delta$, multiplied by the probability of being at $x+\Delta$ at time t. This assumption is based on the independence of the push Δ of any previous history of the motion: it is only necessary to know the initial position of the particle at time t—not at any previous time. This is the *Markov postulate* and the Chapman Kolmogorov equation, of which (1.2.4) is a special form, is the central dynamical equation to all Markov processes. These will be studied in detail in Chap. 3.

ii) *The Fokker-Planck Equation*: Eq. (1.2.10) is the diffusion equation, a special case of the Fokker-Planck equation (also known as *Kolmogorov's equation*) which describes a large class of very interesting stochastic processes in which the system has a continuous sample path. In this case, that means that the pollen grain's position, if thought of as obeying a probabilistic law given by solving the diffusion equation (1.2.10), in which time t is continuous (not discrete, as assumed by Einstein), can be written $x(t)$, where $x(t)$ is a *continuous function of time*—but a random function. This leads us to consider the possibility of describing the dynamics of the system in some direct probabilistic way, so that we would have a *random* or *stochastic differential equation for the path*. This procedure was initiated by Langevin with the famous equation that to this day bears his name. We will discuss this in Sect. 1.2.2, and in detail in Chap. 4.

iii) *The Kramers-Moyal* and similar expansions are essentially the same as that used by Einstein to go from (1.2.4) (the Chapman-Kolmogorov equation) to the diffusion equation (1.2.10). The use of this type of approximation, which effectively replaces a process whose sample paths need not be continuous with one whose paths are continuous, is very common and convenient. Its use and validity will be discussed in Chap. 11.

1.2.2 Langevin's Equation

Some time after Einstein's original derivation, *Langevin* [1.5] presented a new method which was quite different from Einstein's and, according to him, "infinitely more simple." His reasoning was as follows.

From statistical mechanics, it was known that the mean kinetic energy of the Brownian particle should, in equilibrium, reach a value

$$\left\langle \tfrac{1}{2}mv^2 \right\rangle = \tfrac{1}{2}kT\,, \tag{1.2.13}$$

(T; absolute temperature, k; Boltzmann's constant). (Both Einstein and Smoluchowski had used this fact). Acting on the particle, of mass m there should be two forces:

i) A viscous drag: assuming this is given by the same formula as in macroscopic hydrodynamics, this is $-6\pi\eta a\,dx/dt$ where η is the viscosity and a the diameter of the particle, assumed spherical.

ii) Another *fluctuating force* X which represents the incessant impacts of the molecules of the liquid on the Brownian particle. All that is known about it is that fact, and that it should be positive and negative with equal probability. Thus, the equation of motion for the position of the particle is given by Newton's law as

$$m\frac{d^2x}{dt^2} = -6\pi\eta a\frac{dx}{dt} + X\,, \tag{1.2.14}$$

and multiplying by x, this can be written

$$\frac{m}{2}\frac{d^2}{dt^2}(x^2) - mv^2 = -3\pi\eta a\frac{d(x^2)}{dt} + Xx\,, \tag{1.2.15}$$

where $v = dx/dt$. We now average over a large number of different particles and use (1.2.13) to obtain an equation for $\langle x^2\rangle$:

$$\frac{m}{2}\frac{d^2\langle x^2\rangle}{dt^2} + 3\pi\eta a\frac{d\langle x^2\rangle}{dt} = kT\,, \tag{1.2.16}$$

where the term $\langle xX\rangle$ has been set equal to zero because (to quote Langevin) "of the irregularity of the quantity X". One then finds the general solution

$$\frac{d\langle x^2\rangle}{dt} = kT/(3\pi\eta a) + C\exp(-6\pi\eta at/m)\,, \tag{1.2.17}$$

where C is an arbitrary constant. Langevin estimated that the decaying exponential approaches zero with a time constant of the order of 10^{-8} s, which for any practical observation at that time, was essentially immediately. Thus, for practical purposes, we can neglect this term and integrate once more to get

$$\langle x^2\rangle - \langle x_0^2\rangle = [kT/(3\pi\eta a)]t\,. \tag{1.2.18}$$

This corresponds to (1.2.12) as deduced by Einstein, provided we identify

$$D = kT/(6\pi\eta a)\,. \tag{1.2.19}$$

a result which Einstein derived in the same paper but by independent means.

Langevin's equation was the first example of the *stochastic differential equation*—a differential equation with a random term X and hence whose solution is, in some sense, a random function. Each solution of Langevin's equation represents a different random trajectory and, using only rather simple properties of X (his fluctuating force), measurable results can be derived.

One question arises: Einstein explicitly required that (on a sufficiently large time scale) the change Δ be completely independent of the preceding value of Δ. Langevin did not mention such a concept explicitly, but it is there, implicitly, when one sets $\langle Xx\rangle$ equal to zero. The concept that X is extremely irregular *and* (which is not mentioned by Langevin, but is implicit) that X and x are *independent* of each other—that the irregularities in x as a function of time, do not somehow conspire to be always in the same direction as those of X, so that it would not be valid to set the average of the product equal to zero. These are really equivalent to Einstein's independence assumption. The method of Langevin equations is clearly very much more direct, at least at first glance, and gives a very natural way of generalising a dynamical equation to a probabilistic equation. An adequate mathematical grounding for the approach of Langevin, however, was not available until more than 40 years later, when Ito formulated his concepts of stochastic differential equations. And in this formulation, a precise statement of the independence of X and x led to the calculus of stochastic differentials, which now bears his name and which will be fully developed in Chap. 4.

As a physical subject, Brownian motion had its heyday in the first two decades of last century, when Smoluchowski in particular, and many others carried out extensive theoretical and experimental investigations, which showed complete agreement with the original formulation of the subject as initiated by himself and *Einstein*, see [1.6]. More recently, with the development of laser light scattering spectroscopy, Brownian motion has become very much more quantitatively measurable. The technique is to shine intense, coherent laser light into a small volume of liquid containing Brownian particles, and to study the fluctuations in the intensity of the scattered light, which are directly related to the motions of the Brownian particles. By these means it is possible to observe Brownian motion of much smaller particles than the traditional pollen, and to derive useful data about the sizes of viruses and macromolecules. With the preparation of more concentrated suspensions, interactions between the particles appear, generating interesting and quite complex problems related to macromolecular suspensions and colloids [1.7].

The general concept of fluctuations describable by such equations has developed very extensively in a very wide range of situations. The advantages of a continuous description turn out to be very significant, since only a very few parameters are required, i.e., essentially the coefficients of the derivatives in (1.2.7):

$$\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta, \quad \text{and} \quad \int_{-\infty}^{\infty}\Delta^2\phi(\Delta)d\Delta\,. \tag{1.2.20}$$

It is rare to find a problem which cannot be specified, in at least some degree of approximation, by such a system, and for qualitative simple analysis of problems it is normally quite sufficient to consider an appropriate Fokker-Planck equation, of a

form obtained by allowing both coefficients (1.2.20) to depend on x, and in a space of an appropriate number of dimensions.

1.3 The Stock Market

The equations of Brownian motion were in fact first derived by *Bachelier* in his doctoral thesis [1.8], in which he applied the ideas of probability to the pricing of shares and options in the stock market. He introduced the idea of the *relative value* $x = X - X_0$ of a share, that is the difference between its absolute value X and the *most probable value* X_0. He then considered the probability distribution $p_{x,t}$ of relative share prices x at time t, and then deduced the "law of composition" of these probabilities

$$p_{x,t_1+t_2} = \int p_{x,t_1} p_{z-x,t_2}\, dz\,. \tag{1.3.1}$$

This is the Chapman-Kolmogorov equation, that is, it is essentially Einstein's equation, (1.2.4), and the reasoning used to deduce it is basically the same as that of Einstein. Bachelier then sought a solution of the form

$$p = A\mathrm{e}^{-B^2x^2}, \tag{1.3.2}$$

and showed that A and B would be functions of time, concluding:

"The definitive expression for the probability is thus

$$p = \frac{1}{2\pi k\sqrt{t}}\,\mathrm{e}^{-\frac{x^2}{4\pi k^2 t}}\,. \tag{1.3.3}$$

The mathematical expectation

$$\int_0^\infty p\,x\,dx = k\sqrt{t}\,. \tag{1.3.4}$$

is proportional to the square root of the time."

Bachelier gave another derivation rather more similar to Einstein's, in which he divided time into discrete intervals, and considered discrete jumps in the share prices, arriving finally at the *heat equation*, (1.2.10) as the differential equation for the probability distribution. The thesis then considers applications of this probability law to a range of the kind of financial transactions current on the Paris stock exchange of the early 1900s. The value of the work lies in the ideas, rather than the actual results, since Bachelier's use of the Gaussian form for the distribution $p_{x,t}$ clearly has the defect that there is a finite probability that the stock price can become negative, a possibility that he considers, but prefers to treat as negligible.

1.3.1 Statistics of Returns

That the price changes x can have a Gaussian distribution is a reasonable result only if these changes are small compared with the mean price—but this must clearly break down with increasing time if $\langle x^2\rangle \sim t$. Bachelier's work did not generate much interest in finance circles until the 1960s, when *Samuelson* [1.9] decided to develop the approach further. Samuelson rather unfairly criticised Bachelier for "forgetting" that negative prices of shares were not permissible, and suggested a solution to this problem by proposing that changes in prices are most reasonably described as percentages. Explicitly, he proposes the correct quantity to consider is what has become known as the *return* on the share price, given by

$$r = \frac{x}{X}\,, \tag{1.3.5}$$

that is the fractional gain or loss in the share price. This leads to a formulation in which

$$p = \log X\,, \tag{1.3.6}$$

is regarded as the quantity that undergoes Brownian motion. This has the obvious advantage that $p \to -\infty$ means $X \to 0$, so the natural range $(0,\infty)$ of prices is recovered.

There is also a certain human logic in the description. Prices move as a result of judgments by buyers and sellers, to whom the natural measure of a price change is not the absolute size of the change, but the fractional change. The improvement over Bachelier's result is so significant, and the resulting description in terms of the logarithm of the price and the fractional price change so simple, that this is the preferred model to this day. Samuelson termed the process *geometric Brownian motion* or alternatively *economic Brownian motion.*

1.3.2 Financial Derivatives

In order to smooth the running of business, it is often helpful to fix in advance the price of a commodity which will be needed in the future—for example, the price of wheat which has not yet been grown and harvested is moderately uncertain. A baker could choose to pay a fixed sum now for the future delivery of wheat. Rather than deal with an individual grower, the baker can buy the ungrown wheat from a dealer in *wheat futures*, who charges a premium and arranges appropriate contracts with growers. However, the contract to deliver wheat at a certain price on a future date can itself become a tradable item. Having purchased such a contract, the baker can sell it to another baker, or indeed, to anyone else, who may buy it with the view to selling it at a future date, without ever having had anything to do with any wheat at all.

Such a contract is known as a *derivative* security. The wheat future exists only because there is a market for real wheat, but nevertheless can develop an existence of its own. Another kind of derivative is an *option*, in which one buys the *right* to purchase something at a future date at a definite price. If the market price on the date at which the option is exercised is larger than the option price, one exercises the option. If the market price turns out to be below the option price, one discards the option and pays the market price. Purchasing the option limits exposure to price

rises, transferring the risk to the seller of the option, who charges appropriately, and specializes in balancing risks. Options to purchase other securities, such as shares and stocks, are very common, and indeed there are options markets which trade under standardized conditions.

1.3.3 The Black-Scholes Formula

Although a description of market processes in terms of stochastic processes was well-known by the 1970s, it was not clear how it could be used as a tool for making investment decisions. The breakthrough came with the realization that a *portfolio* containing an appropriate mix of cash, stocks and options could be devised in which the short term fluctuations in the various values could be cancelled, and that this gave a relatively simple formula for valuing options—the *Black-Scholes Formula*—which would be of very significant value in making investment decisions. This formula has truly revolutionized the practice of finance; to quote Samuelson [1.10]

> "A great economist of an earlier generation said that, useful though economic theory is for understanding the world, no one would go to an economic theorist for advice on how to run a brewery or produce a mousetrap. Today that sage would have to change his tune: economic principles really do apply and woe the accountant or marketer who runs counter to economic law. Paradoxically, one of our most elegant and complex sectors of economic analysis—the modern theory of finance—is confirmed daily by millions of statistical observations. When today's associate professor of security analysis is asked 'Young man, if you're so smart why ain't you rich?', he replies by laughing all the way to the bank or his appointment as a high-paid consultant to Wall Street."

The derivation was given first in the paper of *Black* and *Scholes* [1.11], and a different derivation was given by *Merton* [1.12]. The formula depends critically on description of the *returns* on securities as a Brownian motion process, which is of limited accuracy. Nevertheless, the formula is sufficiently realistic to make investing in stocks and options a logical and rational process, justifying Samuelson's perhaps over-dramatised view of modern financial theory.

1.3.4 Heavy Tailed Distributions

There is, however, no doubt that the geometric Brownian motion model of financial markets is not exact, and even misses out very important features. One need only study the empirical values of the returns in stock market records (as well as other kinds of markets) and check what kinds of distributions are in practice observed. The results are not really in agreement with a Gaussian distribution of returns—rather, the observed distribution of returns is usually approximately Gaussian for small values of r, but the probability of large values of r is always observed to be significantly larger than the Gaussian prediction—the observed distributions are said to have *heavy tails*.

The field of *Continuous Time Finance* [1.10] is an impressive theoretical edifice built on this flawed foundation of Brownian motion, but so far it appears to be the most practical method of modelling financial markets. With modern electronic banking and transfer of funds, it is possible to trade over very short time intervals, during which perhaps, in spite of the overall increase of trading activity which results, a Brownian description is valid.

It is certainly sufficiently valued for its practitioners to be highly valued, as Samuelson notes. However, every so often one of these practitioners makes a spectacular loss, threatening financial institutions. While there is public alarm about billion dollar losses, those who acknowledge the significance of heavy tails are unsurprised.

1.4 Birth-Death Processes

A wide variety of phenomena can be modelled by a particular class of process called a birth-death process. The name obviously stems from the modelling of human or animal populations in which individuals are born, or die. One of the most entertaining models is that of the prey-predator system consisting of two kinds of animal, one of which preys on the other, which is itself supplied with an inexhaustible food supply. Thus letting X symbolise the prey, Y the predator, and A the food of the prey, the process under consideration might be

$$X + A \to 2X\,, \tag{1.4.1a}$$

$$X + Y \to 2Y\,, \tag{1.4.1b}$$

$$Y \to B\,, \tag{1.4.1c}$$

which have the following naive, but charming interpretation. The first equation symbolises the prey eating one unit of food, and reproducing immediately. The second equation symbolises a predator consuming a prey (who thereby dies—this is the only death mechanism considered for the prey) and immediately reproducing. The final equation symbolises the death of the predator by natural causes. It is easy to guess model differential equations for x and y, the numbers of X and Y. One might assume that the first reaction symbolises a rate of production of X proportional to the product of x and the amount of food; the second equation a production of Y (and an equal rate of consumption of X) proportional to xy, and the last equation a death rate of Y, in which the rate of death of Y is simply proportional to y; thus we might write

$$\frac{dx}{dt} = k_1 ax - k_2 xy\,, \tag{1.4.2a}$$

$$\frac{dy}{dt} = k_2 xy - k_3 y\,. \tag{1.4.2b}$$

The solutions of these equations, which were independently developed by *Lotka* [1.13] and *Volterra* [1.14] have very interesting oscillating solutions, as presented in Fig. 1.3a. These oscillations are qualitatively easily explicable. In the absence of significant numbers of predators, the prey population grows rapidly until the presence of so much prey for the predators to eat stimulates their rapid reproduction, at

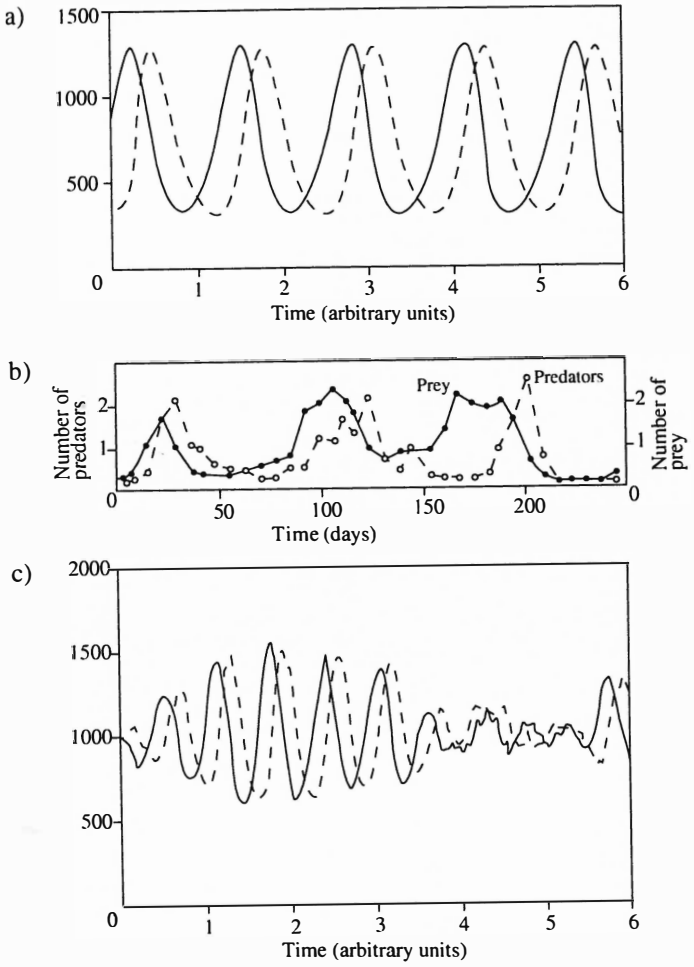


Fig. 1.3a–c. Time development in prey-predator systems. (**a**) Plot of solutions of the deterministic equations (1.4.2a, 1.4.2b) (x = solid line, y = dashed line). (**b**) Data for a real prey-predator system. Here the predator is a mite (Eotetranychus sexmaculatus—dashed line) which feeds on oranges, and the prey is another mite (Typhlodromus occidentalis). Data from [1.15, 1.16]. (**c**) Simulation of stochastic equations (1.4.3a–1.4.3d).

the same time reducing the number of prey which get eaten. Because a large number of prey have been eaten, there are no longer enough to maintain the population of predators, which then die out, returning us to our initial situation. The cycles repeat indefinitely and are indeed, at least qualitatively, a feature of many real prey-predator systems. An example is given in Fig. 1.3b.

Of course, the realistic systems do not follow the solutions of differential equations exactly—they fluctuate about such curves. One must include these fluctuations and the simplest way to do this is by means of a *birth-death master equation*. We assume a probability distribution, $P(x,y,t)$, for the number of individuals at a given time and ask for a probabilistic law corresponding to (1.4.2a, 1.4.2b). This is done by assuming that in an infinitesimal time Δt, the following *transition probability* laws holds.

$$\text{Prob}\,(x \to x+1; y \to y) = k_1 a x \Delta t\,, \tag{1.4.3a}$$

$$\text{Prob}\,(x \to x-1; y \to y+1) = k_2 x y \Delta t\,, \tag{1.4.3b}$$

$$\text{Prob}\,(x \to x; y \to y-1) = k_3 y \Delta t\,, \tag{1.4.3c}$$

$$\text{Prob}\,(x \to x; y \to y) = 1 - (k_1 a x + k_2 x y + k_3 y)\Delta t\,. \tag{1.4.3d}$$

Thus, we simply, for example, replace the simple rate laws by probability laws. We then employ what amounts to the same equation as Einstein and others used, i.e., the Chapman-Kolmogorov equation, namely, we write the probability at $t + \Delta t$ as a sum of terms, each of which represents the probability of a previous state multiplied by the probability of a transition to the state (x, y). Thus, we find by letting $\Delta t \to 0$:

$$\begin{aligned}\frac{P(x,y,t+\Delta t) - P(x,y,t)}{\Delta t} \to \frac{\partial P(x,y,t)}{\partial t} \\ = k_1 a (x-1) P(x-1,y,t) + k_2 (x+1)(y-1) \\ \times P(x+1, y-1, t) + k_3 (y+1) P(x, y+1, t) \\ - (k_1 a x + k_2 x y + k_3 y) P(x,y,t).\end{aligned} \tag{1.4.4}$$

In writing the assumed probability laws (1.4.3a–1.4.3d), we are assuming that the probability of each of the events occurring can be determined simply from the knowledge of x and y. This is again the Markov postulate which we mentioned in Sect. 1.2.1. In the case of Brownian motion, very convincing arguments can be made in favour of this Markov assumption. Here it is by no means clear. The concept of heredity, i.e., that the behaviour of progeny is related to that of parents, clearly contradicts this assumption. How to *include* heredity is another matter; by no means does a unique prescription exist.

The assumption of the Markov postulate in this context is valid to the extent that different individuals of the same species are similar; it is invalid to the extent that, nevertheless, perceptible inheritable differences do exist.

This type of model has a wide application—in fact to any system to which a population of individuals may be attributed, for example systems of molecules of various chemical compounds, of electrons, of photons and similar physical particles as well as biological systems. The particular choice of transition probabilities is made on various grounds determined by the degree to which details of the births and deaths involved are known. The simple multiplicative laws, as illustrated in (1.4.3a–1.4.3d), are the most elementary choice, ignoring, as they do, almost all details of the processes involved. In some of the physical processes we can derive the transition probabilities in much greater detail and with greater precision.

Equation (1.4.4) has no simple solution, but one major property differentiates equations like it from an equation of Langevin's type, in which the fluctuation term is simply added to the differential equation. Solutions of (1.4.4) determine both the gross deterministic motion and the fluctuations; the fluctuations are typically of the

same order of magnitude as the square roots of the *numbers* of individuals involved. It is not difficult to simulate a sample time development of the process as in Fig. 1.3c. The figure does show the correct general features, but the model is so obviously simplified that exact agreement can never be expected. Thus, in contrast to the situation in Brownian motion, we are not dealing here so much with a theory of a phenomenon, as with a class of mathematical models, which are simple enough to have a very wide range of approximate validity. We will see in Chap. 11 that a theory can be developed which can deal with a wide range of models in this category, and that there is indeed a close connection between this kind of theory and that of stochastic differential equations.

1.5 Noise in Electronic Systems

The early days of radio with low transmission powers and primitive receivers, made it evident to every ear that there were a great number of highly irregular electrical signals which occurred either in the atmosphere, the receiver, or the radio transmitter, and which were given the collective name of "noise", since this is certainly what they sounded like on a radio. Two principal sources of noise are shot noise and Johnson noise.

1.5.1 Shot Noise

In a vacuum tube (and in solid-state devices) we get a nonsteady electrical current, since it is generated by individual electrons, which are accelerated across a distance and deposit their charge one at a time on the anode. The electric current arising from such a process can be written

$$I(t) = \sum_{t_k} F(t - t_k)\,, \tag{1.5.1}$$

where $F(t-t_k)$ represents the contribution to the current of an electron which arrives at time t_k. Each electron is therefore assumed to give rise to the same shaped pulse, but with an appropriate delay, as in Fig. 1.4.

A statistical aspect arises immediately we consider what kind of choice must be made for t_k. The simplest choice is that each electron arrives independently of the previous one—that is, the times t_k are randomly distributed with a certain average number per unit time in the range $(-\infty, \infty)$, or whatever time is under consideration.

The analysis of such noise was developed during the 1920's and 1930's and was summarised and largely completed by *Rice* [1.17]. It was first considered as early as 1918 by *Schottky* [1.18].

We shall find that there is a close connection between shot noise and processes described by birth-death master equations. For, if we consider n, the number of electrons which have arrived up to a time t, to be a statistical quantity described by a probability $P(n,t)$, then the assumption that the electrons arrive independently is clearly the Markov assumption. Then, assuming the probability that an electron will

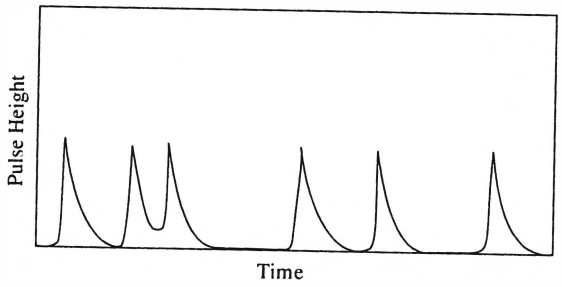


Fig. 1.4. Illustration of shot noise: identical electric pulses arrive at random times

arrive in the time interval between t and $t + \Delta t$ is completely independent of t and n, its only dependence can be on Δt. By choosing an appropriate constant λ, we may write

$$\text{Prob}\,(n \to n+1, \text{in time } \Delta t) = \lambda\,\Delta t\,, \tag{1.5.2}$$

so that

$$P(n, t+\Delta t) = P(n,t)(1 - \lambda\,\Delta t) + P(n-1,t)\lambda\,\Delta t\,, \tag{1.5.3}$$

and taking the limit $\Delta t \to 0$

$$\frac{\partial P(n,t)}{\partial t} = \lambda[P(n-1,t) - P(n,t)]\,, \tag{1.5.4}$$

which is a pure birth process. By writing

$$G(s,t) = \sum s^n P(n,t)\,, \tag{1.5.5}$$

[here, $G(s,t)$ is known as the generating function for $P(n,t)$, and the particular technique of solving (1.5.4) is very widely used], we find

$$\frac{\partial G(s,t)}{\partial t} = \lambda(s-1)G(s,t), \tag{1.5.6}$$

so that

$$G(s,t) = \exp[\lambda(s-1)t]G(s,0)\,. \tag{1.5.7}$$

By requiring at time $t = 0$ that no electrons had arrived, it is clear that $P(0,0)$ is and $P(n,0)$ is zero for all $n \geqslant 1$, so that $G(s,0) = 1$. Expanding the solution (1.5.7 in powers of s, we find

$$P(n,t) = \exp(-\lambda t)(\lambda t)^n/n!\,, \tag{1.5.8}$$

which is known as a *Poisson distribution* (Sect. 2.8.3). Let us introduce the variabl $N(t)$, which is to be considered as the number of electrons which have arrived up t time t, and is a random quantity. Then,

$$P(n,t) = \text{Prob}\,\{N(t) = n\}\,, \tag{1.5.9}$$

and $N(t)$ can be called a *Poisson process variable.* Then clearly, the quantity $\mu(t)$ formally defined by

$$\mu(t) = dN(t)/dt\,, \tag{1.5.10}$$

is zero, except when $N(t)$ increases by 1; at that stage it is a Dirac delta function, i.e.,

$$\mu(t) = \sum_k \delta(t - t_k)\,, \tag{1.5.11}$$

where the t_k are the times of arrival of the individual electrons. We may write

$$I(t) = \int_{-\infty}^{\infty} dt' F(t - t')\mu(t')\,. \tag{1.5.12}$$

A very reasonable restriction on $F(t - t')$ is that it vanishes if $t < t'$, and that for $t \to \infty$, it also vanishes. This simply means that no current arises from an electron before it arrives, and that the effect of its arrival eventually dies out. We assume then, for simplicity, the very commonly encountered form

$$F(t) = \begin{cases} q\,\mathrm{e}^{-\alpha t}\,, & (t > 0)\,, \\ 0\,, & (t < 0)\,, \end{cases} \tag{1.5.13}$$

so that (1.5.12) can be rewritten as

$$I(t) = \int_{-\infty}^{t} dt'\, q\,\mathrm{e}^{-\alpha(t-t')}\frac{dN(t')}{dt'}\,. \tag{1.5.14}$$

We can derive a simple differential equation. We differentiate $I(t)$ to obtain

$$\frac{dI(t)}{dt} = -\int_{-\infty}^{t} dt'\,\alpha q\,\mathrm{e}^{-\alpha(t-t')}\frac{dN(t')}{dt'} + \left[q\,\mathrm{e}^{-\alpha(t-t')}\frac{dN(t')}{dt}\right]_{t'=t}\,, \tag{1.5.15}$$

so that

$$\frac{dI(t)}{dt} = -\alpha I(t) + q\mu(t)\,. \tag{1.5.16}$$

This is a kind of stochastic differential equation, similar to Langevin's equation, in which, however, the fluctuating force is given by $q\mu(t)$, where $\mu(t)$ is the derivative of the Poisson process, as given by (1.5.11). However, the mean of $\mu(t)$ is nonzero, in fact, from (1.5.10)

$$\langle \mu(t)dt\rangle = \langle dN(t)\rangle = \lambda\, dt\,, \tag{1.5.17}$$

$$\langle [dN(t) - \lambda dt]^2\rangle = \lambda\, dt\,, \tag{1.5.18}$$

from the properties of the Poisson distribution, for which the variance equals the mean. Defining, then, the fluctuation as the difference between the mean value and $dN(t)$, we write

$$d\eta(t) = dN(t) - \lambda dt\,, \tag{1.5.19}$$

so that the stochastic differential equation (1.5.16) takes the form

$$dI(t) = [\lambda q - \alpha I(t)]dt + q\,d\eta(t)\,. \tag{1.5.20}$$

Now how does one solve such an equation? In this case, we have an academic problem anyway since the solution is known, but one would like to have a technique. Suppose we try to follow the method used by Langevin—what will we get as an answer? The short reply to this question is: nonsense. For example, using ordinary calculus and assuming $\langle I(t)d\eta(t)\rangle = 0$, we can derive

$$\frac{d\langle I(t)\rangle}{dt} = \lambda q - \alpha\langle I(t)\rangle\,, \tag{1.5.21}$$

$$\frac{1}{2}\frac{d\langle I^2(t)\rangle}{dt} = \lambda q\langle I(t)\rangle - \alpha\langle I^2(t)\rangle\,. \tag{1.5.22}$$

Solving in the limit $t \to \infty$, where the mean values would reasonably be expected to be constant one finds

$$\langle I(\infty)\rangle = \lambda q/\alpha, \tag{1.5.23}$$

$$\langle I^2(\infty)\rangle = (\lambda q/\alpha)^2. \tag{1.5.24}$$

The first answer is reasonable—it merely gives the average current through the system in a reasonable equation, but the second implies that the mean square current is the same as the square of the mean, i.e., the current at $t \to \infty$ does not fluctuate! This is rather unreasonable, and the solution to the problem will show that stochastic differential equations are rather more subtle than we have so far presented.

Firstly, the notation in terms of differentials used in (1.5.17–1.5.20) has been chosen deliberately. In deriving (1.5.22), one uses ordinary calculus, i.e., one writes

$$d(I^2) \equiv (I + dI)^2 - I^2 = 2I\,dI + (dI)^2\,, \tag{1.5.25}$$

and then one drops the $(dI)^2$ as being of second order in dI. But now look at (1.5.18): this is equivalent to

$$\langle d\eta(t)^2\rangle = \lambda\, dt\,, \tag{1.5.26}$$

so that a *quantity of second order in $d\eta$ is actually of first order in dt.* The reason is not difficult to find. Clearly,

$$d\eta(t) = dN(t) - \lambda\, dt\,, \tag{1.5.27}$$

but the curve of $N(t)$ is a step function, discontinuous, and certainly not differentiable, at the times of arrival of the individual electrons. In the ordinary sense, none of these calculus manipulations is permissible. But we can make sense out of them as follows. Let us simply calculate $\langle d(I^2)\rangle$ using (1.5.20, 1.5.25, 1.5.26):

$$\langle d(I)^2\rangle = 2\langle I\{[\lambda q - \alpha I]dt + q\,d\eta(t)\}\rangle + \langle\{[\lambda q - \alpha I]dt + q\,d\eta(t)\}^2\rangle\,. \tag{1.5.28}$$

We now assume again that $\langle I(t)d\eta(t)\rangle = 0$ and expand, after taking averages using the fact that $\langle d\eta(t)^2\rangle = \lambda\, dt$, to 1*st order in* dt. We obtain

$$\tfrac{1}{2}d\langle I^2\rangle = \left[\lambda q\langle I\rangle - \alpha\langle I^2\rangle + \tfrac{1}{2}q^2\lambda\right]dt\,, \tag{1.5.29}$$

and this gives

$$\langle I^2(\infty)\rangle - \langle I(\infty)\rangle^2 = \frac{q^2\lambda}{2\alpha}\,. \tag{1.5.30}$$

Thus, there are fluctuations from this point of view, as $t \to \infty$. The extra term in (1.5.29) as compared to (1.5.22) arises directly out of the statistical considerations implicit in $N(t)$ being a discontinuous random function.

Thus we have discovered a somewhat deeper way of looking at Langevin's kind of equation—the treatment of which, from this point of view, now seems extremely naive. In Langevin's method the fluctuating force X is not specified, but it will become clear in this book that problems such as we have just considered are very widespread in this subject. The moral is that random functions cannot normally be differentiated according to the usual laws of calculus; special rules have to be developed, and a precise specification of what one means by differentiation becomes important. We will specify these problems and their solutions in Chap. 4 which will concern itself with situations in which the fluctuations are Gaussian.

1.5.2 Autocorrelation Functions and Spectra

The measurements which one can carry out on fluctuating systems such as electric circuits are, in practice, not of unlimited variety. So far, we have considered the distribution functions, which tell us, at any time, what the probability distribution of the values of a stochastic quantity are. If we are considering a measurable quantity $x(t)$ which fluctuates with time, in practice we can sometimes determine the distribution of the values of x, though more usually, what is available at one time are the mean $\bar{x}(t)$ and the variance var$[x(t)]$.

The mean and the variance do not tell a great deal about the underlying dynamics of what is happening. What would be of interest is some quantity which is a measure of the influence of a value of x at time t on the value at time $t + \tau$. Such a quantity is the *autocorrelation function*, which was apparently first introduced by *Taylor* [1.19] as

$$G(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, x(t)x(t+\tau)\,. \tag{1.5.31}$$

This is the time average of a two-time product over an arbitrary large time T, which is then allowed to become infinite. Using modern computerized data collection technology it is straightforward to construct an autocorrelation from any stream of data, either in real time or from recorded data.

A closely connected approach is to compute the *spectrum* of the quantity $x(t)$. This is defined in two stages. First, define

$$y(\omega) = \int_0^T dt\, e^{-i\omega t} x(t)\,, \tag{1.5.32}$$

then the spectrum is defined by

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} |y(\omega)|^2. \tag{1.5.33}$$

The autocorrelation function and the spectrum are closely connected. By a little manipulation one finds

$$S(\omega) = \lim_{T\to\infty} \left[\frac{1}{\pi} \int_0^T \cos(\omega\tau) d\tau \frac{1}{T} \int_0^{T-\tau} x(t)x(t+\tau) dt \right], \tag{1.5.34}$$

and taking the limit $T \to \infty$ (under suitable assumptions to ensure the validity of certain interchanges of order), one finds

$$S(\omega) = \frac{1}{\pi} \int_0^\infty \cos(\omega\tau) G(\tau) d\tau\,. \tag{1.5.35}$$

This is a fundamental result which relates the Fourier transform of the autocorrelation function to the spectrum. The result may be put in a slightly different form when one notices that

$$G(-\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\tau}^{T-\tau} dt\, x(t+\tau)x(t) = G(\tau)\,, \tag{1.5.36}$$

so we obtain

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} G(\tau)\, d\tau\,, \tag{1.5.37}$$

with the corresponding inverse

$$G(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega)\, d\omega\,. \tag{1.5.38}$$

This result is known as the *Wiener-Khinchin theorem* [1.20, 1.21] and has widespread application.

It means that one may either directly measure the autocorrelation function of a signal, or the spectrum, and convert back and forth, which by means of the fast Fourier transform and computer is relatively straightforward.

1.5.3 Fourier Analysis of Fluctuating Functions: Stationary Systems

The autocorrelation function has been defined so far as a time average of a signal, but we may also consider the *ensemble average*, in which we repeat the same measurement many times, and compute averages, denoted by $\langle\ \rangle$. It will be shown that for very many systems, the time average is equal to the ensemble average; such systems are termed *ergodic*—see Sect. 3.7.1.

If we have such a fluctuating quantity $x(t)$, then we can consider the average of the product of two time-values of x

$$\langle x(t)x(t+\tau)\rangle = G(\tau)\,. \tag{1.5.39}$$

The fact that the result is independent of the absolute time t is a consequence of our ergodic assumption.

Now it is very natural to write a Fourier transform for the stochastic quantity $x(t)$

$$x(t) = \int d\omega\, c(\omega) e^{i\omega t}\,, \tag{1.5.40}$$

and consequently,

$$c(\omega) = \frac{1}{2\pi}\int dt\, x(t) e^{-i\omega t}\,. \tag{1.5.41}$$

Note that $x(t)$ real implies

$$c(\omega) = c^*(-\omega)\,. \tag{1.5.42}$$

If the system is ergodic, we must have a constant $\langle x(t)\rangle$, since the time average is clearly constant. The process is then *stationary* by which we mean that all time-dependent averages are functions only of time differences, i.e., averages of functions $x(t_1)$, $x(t_2)$, $\ldots x(t_n)$ are equal to those of $x(t_1+\Delta)$, $x(t_2+\Delta)$, $\ldots x(t_n+\Delta)$.

For convenience, in what follows we assume $\langle x\rangle = 0$. Hence,

$$\begin{aligned}
\langle c(\omega)\rangle &= \frac{1}{2\pi}\int dt\,\langle x\rangle e^{-i\omega t} = 0\,, \qquad (1.5.43)\\
\langle c(\omega)c^*(\omega')\rangle &= \frac{1}{(2\pi)^2}\int\int dt\,dt' e^{-i\omega t + i\omega' t'}\langle x(t)x(t')\rangle\,,\\
&= \frac{1}{(2\pi)}\delta(\omega-\omega')\int d\tau\, e^{i\omega\tau}G(\tau)\,,\\
&= \delta(\omega-\omega')S(\omega)\,. \qquad (1.5.44)
\end{aligned}$$

Here we find not only a relationship between the mean square $\langle|c(\omega)|^2\rangle$ and the spectrum, but also the result that stationarity alone implies that $c(\omega)$ and $c^*(\omega')$ are uncorrelated, since the term $\delta(\omega-\omega')$ arises because $\langle x(t)x(t')\rangle$ is a function only of $t-t'$.

1.5.4 Johnson Noise and Nyquist's Theorem

Two brief and elegant papers appeared in 1928 in which *Johnson* [1.22] demonstrated experimentally that an electric resistor automatically generated fluctuations of electric voltage, and *Nyquist* [1.23] demonstrated its theoretical derivation, in complete accordance with Johnson's experiment. The principle involved was already known by *Schottky* [1.18] and is the same as that used by Einstein and Langevin. This principle is that of thermal equilibrium. If a resistor R produces electric fluctuations, these will produce a current which will generate heat. The heat produced in the resistor must exactly balance the energy taken out of the fluctuations. The detailed working out of this principle is not the subject of this section, but we will find that such results are common throughout the physics and chemistry of stochastic processes, where the principles of statistical mechanics, whose basis is not essentially stochastic, are brought in to complement those of stochastic processes—such results are known as *fluctuation-dissipation theorems*.

Nyquist's experimental result was the following. We have an electric resistor of resistance R at absolute temperature T. Suppose by means of a suitable filter we measure $E(\omega)d\omega$, the voltage across the resistor with angular frequency in the range $(\omega, \omega+d\omega)$. Then, if k is Boltzmann's constant,

$$\langle E^2(\omega)\rangle = RkT/\pi\,. \tag{1.5.45}$$

This result is known nowadays as *Nyquist's theorem*. Johnson remarked. "The effect is one of the causes of what is called 'tube noise' in vacuum tube amplifiers. Indeed, it is often by far the larger part of the 'noise' of a good amplifier."

Johnson noise is easily described by the formalism of the previous subsection. The mean noise voltage is zero across a resistor, and the system is arranged so that it is in a steady state and is expected to be well represented by a stationary process. Johnson's quantity is, in practice, a limit of the kind (1.5.33) and may be summarised by saying that the voltage spectrum $S(\omega)$ is given by

$$S(\omega) = RkT/\pi, \tag{1.5.46}$$

that is, the spectrum is flat, i.e., a constant function of ω. In the case of light, the frequencies correspond to different colours of light. If we perceive light to be white it is found that in practice all colours are present in equal proportions—the optical spectrum of white light is thus flat—at least within the visible range. In analogy the term *white noise* is applied to a noise voltage (or any other fluctuating quantity) whose spectrum is flat.

White noise cannot actually exist. The simplest demonstration is to note that the mean power dissipated in the resistor in the frequency range (ω_1, ω_2) is given by

$$\int_{\omega_1}^{\omega_2} d\omega\, S(\omega)/R = kT(\omega_1-\omega_2)/\pi\,, \tag{1.5.47}$$

so that the total power dissipated in all frequencies is infinite! Nyquist realised this and noted that, in practice, there would be quantum corrections which would, at room temperature, make the spectrum flat only up to 7×10^{13} Hz, which is not detectable in practice, in a radio situation. The actual power dissipated in the resistor would be somewhat less than infinite—10^{-10} W in fact! And in practice there are other limiting factors such as the inductance of the system, which would limit the spectrum to even lower frequencies.

From the definition of the spectrum in terms of the autocorrelation function given in Sect. 1.5, we have

$$\begin{aligned}
\langle E(t+\tau)E(t)\rangle &= G(\tau)\,, \qquad (1.5.48)\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}2RkT\,, \qquad (1.5.49)\\
&= 2RkT\delta(\tau)\,, \qquad (1.5.50)
\end{aligned}$$

which implies that no matter how small the time difference τ, $E(t+\tau)$ and $E(\tau)$ are not correlated. This is, of course, a direct result of the flatness of the spectrum. A typical model of $S(\omega)$ that is almost flat is

$$S(\omega) = RkT/[\pi(\omega^2\tau_C^2+1)]\,. \tag{1.5.51}$$

This is flat provided $\omega \ll \tau_C^{-1}$. The Fourier transform can be explicitly evaluated in this case to give

$$\langle E(t+\tau)E(t)\rangle = (R\,kT/\tau_C)\exp(-\tau/\tau_C)\,, \tag{1.5.52}$$

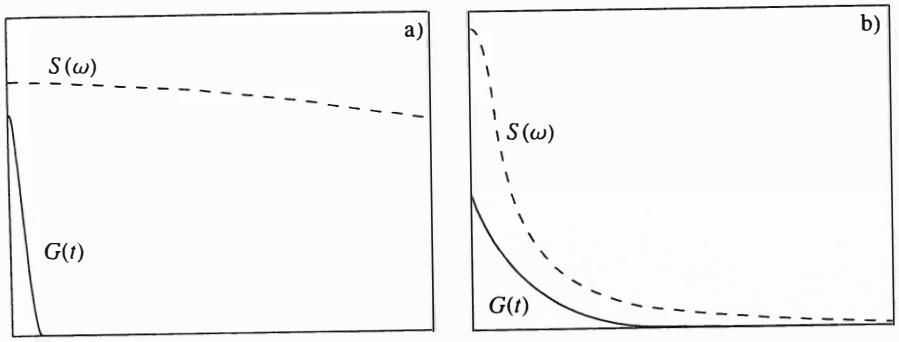


Fig. 1.5. Correlation Functions (———) and corresponding spectra (- - - - -) for (**a**) short correlation time corresponding to an almost flat spectrum; (**b**) long correlation time, giving a quite rapidly decreasing spectrum

so that the autocorrelation function vanishes only for $\tau \gg \tau_C$, which is called the *correlation time* of the fluctuating voltage. Thus, the delta function correlation function appears as an idealisation, only valid on a sufficiently long time scale.

This is very reminiscent of Einstein's assumption regarding Brownian motion and of the behaviour of Langevin's fluctuating force. The idealised *white noise* will play a highly important role in this book but, in just the same way as the fluctuation term that arises in a stochastic differential equation is not the same as an ordinary differential, we will find that differential equations which include white noise as a driving term have to be handled with great care. Such equations arise very naturally in any fluctuating system and it is possible to arrange by means of *Stratonovich's rules* for ordinary calculus rules to apply, but at the cost of imprecise mathematical definition and some difficulties in stochastic manipulation. It turns out to be far better to abandon ordinary calculus and use the *Ito calculus*, which is not very different (it is, in fact, very similar to the calculus presented for shot noise) and to preserve tractable statistical properties. All these matters will be discussed thoroughly in Chap. 4.

White noise, as we have noted above, does not exist as a physically realisable process and the rather singular behaviour it exhibits does not arise in any realisable context. It is, however, fundamental in a mathematical, and indeed in a physical sense, in that it is an idealisation of very many processes that do occur. The slightly strange rules which we will develop for the calculus of white noise are not really very difficult and are very much easier to handle than any method which always deals with a real noise. Furthermore, situations in which white noise is not a good approximation can very often be indirectly expressed quite simply in terms of white noise. In this sense, white noise is the starting point from which a wide range of stochastic descriptions can be derived, and is therefore fundamental to the subject of this book.

2. Probability Concepts

In the preceding chapter, we introduced probability notions without any definitions. In order to formulate essential concepts more precisely, it is necessary to have some more precise expression of these concepts. The intention of this chapter is to provide some background, and to present a number of essential results. It is not a thorough outline of mathematical probability, for which the reader is referred to standard mathematical texts such as those by *Feller* [2.1] and *Papoulis* [2.2].

2.1 Events, and Sets of Events

It is convenient to use a notation which is as general as possible in order to describe those occurrences to which we might wish to assign probabilities. For example, we may wish to talk about a situation in which there are 6.4×10^{14} molecules in a certain region of space; or a situation in which a Brownian particle is at a certain point $\boldsymbol{x}$ in space; or possibly there are 10 mice and 3 owls in a certain region of a forest.

These occurrences are all examples of practical realisations of *events*. More abstractly, an event is simply a member of a certain space, which in the cases most practically occurring can be characterised by a vector of integers

$$\boldsymbol{n} = (n_1, n_2, n_3 \ldots), \tag{2.1.1}$$

or a vector of real numbers

$$\boldsymbol{x} = (x_1, x_2, x_3 \ldots). \tag{2.1.2}$$

The dimension of the vector is arbitrary.

It is convenient to use the language of set theory, introduce the concept of a *set of events*, and use the notation

$$\omega \in A, \tag{2.1.3}$$

to indicate that the event ω is one of events contained in A. For example, one may consider the set $A(25)$ of events in the ecological population in which there are no more than 25 animals present; clearly the event $\bar{\omega}$ that there are 3 mice, a tiger, and no other animals present satisfies

$$\bar{\omega} \in A(25). \tag{2.1.4}$$

More significantly, suppose we define the set of events $A(\boldsymbol{r}, \Delta V)$ that a molecule is within a volume element ΔV centred on a point $\boldsymbol{r}$. In this case, the practical significance of working in terms of sets of events becomes clear, because we should