Stochastic Methods

John R. Boccio

November 20, 2012

Contents

Chapter 1

A Historical Introduction

1.1 Motivation

Theoretical science up to the end of the 19^{th} century can be viewed as the study of solutions of differential equations and the modeling of natural phenomena by deterministic solutions of these differential equations. It was commonly thought at that time that if all initial data was known, one would be able to predict the future with certainty.

We now know this is not so, in at least two ways. First, the advent of quantum mechanics in the early party of the 20^{th} gave rise to a new physics and thus a new theoretical basis for all science, which had at it core a purely statistical element. Second, more recently, the concept of chaos has been studied. in which even quite simple differential equation systems have the rather alarming property of giving rise to essentially unpredictable behavior. 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 (exponentially) magnified that no practical predictability remains.

We will not be discussing chaos and quantum mechanics in these notes. We will be giving a semi-historical outline of how a phenomenological theory of fluctuating phenomena arose and what its essential point are. The clear usefulness of predictable models indicates that life is not entirely chaos. But there is a limit to predictability and what we will be most concerned with in these notes are models of limited predictability. The experience of careful measurements in science normally gives us data like that of Figure 1, representing the growth of the number of molecules of a substance X formed by a chemical reaction of the form $X \rightleftharpoons A$.

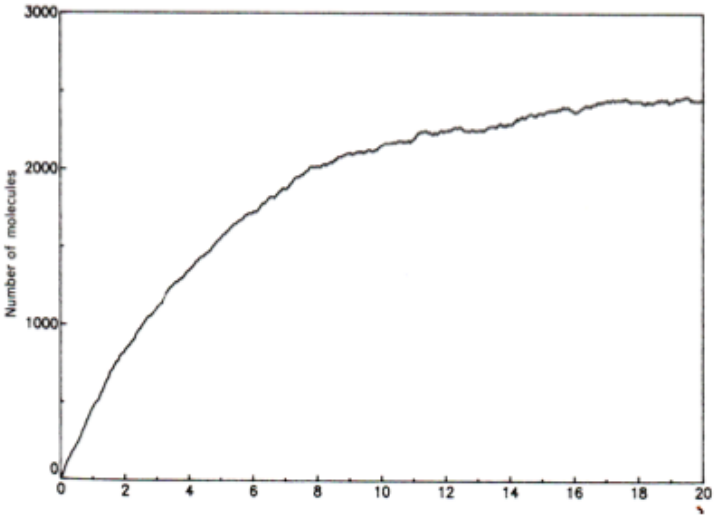

Figure 1.1: Stochastic simulation of an isomerization reaction $X \rightleftharpoons A$

A well-defined deterministic motion is quite evident and it is reproducible, unlike the fluctuations 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, and the observed phenomena became known as *Brownian motion* because of his fundamental pioneering work. Brown was a botanist and thus tested whether this motion was in some way a manifestation of life. By showing that the motion was present in any suspension of fine particles - glass, minerals and even a fragment of the sphinx - he ruled out any specifically organics origin of this motion. The motion is illustrated in Figure 2.

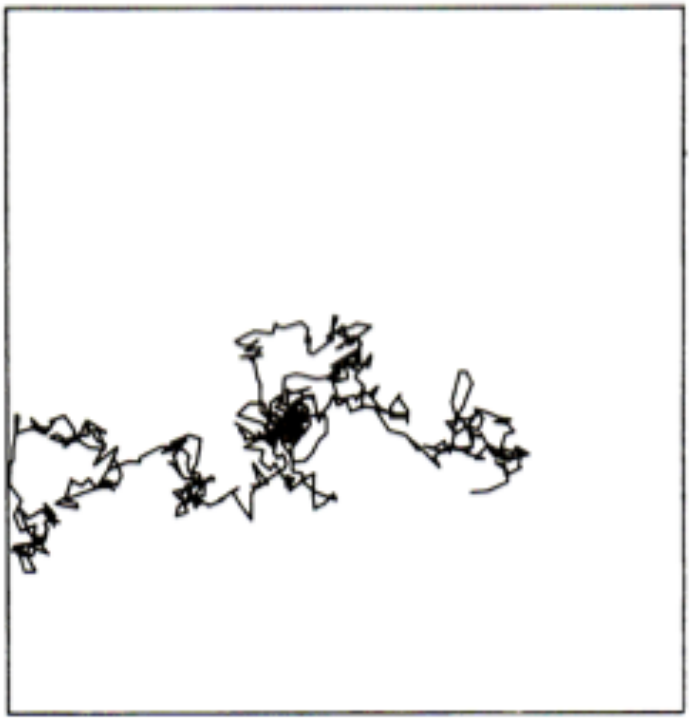

Figure 1.2: Motion of a point undergoing Brownian motion

The riddle of Brownian motion was not quickly solved, and a satisfactory explanation did not come about until 1905, when Einstein published an explanation under the rather modest title "über die von der molekular-kinetischen Theorie de 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). The same explanation was independently developed by Smoluchowski, who was responsible for much of the later systematic development and for much of the experimental verification of brownian motion.

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

(i) The motion is caused by the exceedingly frequent impacts on the pollen grain of the incessantly moving molecules of the liquid in which it is suspended.

(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 those described by Einstein calls out for a statistical explanation if this type of phenomena. Statistics has 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 arising and not as an intrinsic part of the time evolution of the system. Rayleigh 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 modeling of natural phenomena. *Stochastic* is an adjective that refers to systems whose behavior is intrinsically non-deterministic, sporadic, and categorically not intermittent (i.e. random). A stochastic process is one whose behavior is non-deterministic, in that a system's subsequent state is determined both by the process's predictable actions and by a random element.

Einstein's reasoning is very clear and elegant. It contains all of the basic concepts that will make up the subject matter of these notes. Rather than paraphrase a classic piece of work, I will simply give an extended excerpt from Einstein's original paper(translated).

It must be clearly 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-coordinate 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*

$$dn = n\phi(\Delta)d\Delta \tag{1.1}$$

where

$$\int_{-\infty}^{\infty} \phi(\Delta)d\Delta = 1 \tag{1.2}$$

and ϕ is only different from zero for very small values of Δ, and satisfies the condition

$$\phi(\Delta) = \phi(-\Delta) \tag{1.3}$$

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 $t+\tau$ are found between two planes perpendicular to the x-axis and passing through the points x and $x+dx$. One obtains

$$f(x,t+\tau)dx = dx\int_{=\infty}^{\infty} f(x+\Delta),t)\Phi(\Delta)d\Delta \tag{1.4}$$

But since τ is very small

$$f(x,t+\tau) = f(x,t) + \tau\frac{\partial f}{\partial t} \tag{1.5}$$

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

$$f(x+\Delta,t) = f(x,t_+\Delta\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!}\frac{\partial^2 f(x,t)}{\partial x^2} + \cdots \tag{1.6}$$

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

$$f + \frac{\partial f}{\partial t}\tau = f\int_{=\infty}^{\infty}\Phi(\Delta)d\Delta + \frac{\partial f}{\partial x}\int_{=\infty}^{\infty}\Delta\Phi(\Delta)d\Delta + \frac{\partial^2 f}{\partial x^2}\int_{=\infty}^{\infty}\frac{\Delta^2}{2!}\Phi(\Delta)d\Delta \tag{1.7}$$

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

$$\int_{=\infty}^{\infty}\Phi(\Delta)d\Delta = 1 \tag{1.8}$$

and setting

$$\frac{1}{\tau}\int_{=\infty}^{\infty}\frac{\Delta^2}{2!}\Phi(\Delta)d\Delta = D \tag{1.9}$$

and keeping only the first and third terms on the right-hand side,

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}\cdots \tag{1.10}$$

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

$$f(x,t) = \frac{n}{\sqrt{4\pi D}} \frac{e^{-x^2/4Dt}}{\sqrt{t}} \tag{1.11}$$

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 on the direction of the X-axis; it is

$$\lambda_x = \sqrt{\bar{x}^2} = \sqrt{2Dt} \tag{1.12}$$

Einstein's derivation is really based on a discrete time assumption, that impacts happen only at times $0, \tau, 2\tau, 3\tau, ...$, and his resulting equation (1.10) for the distribution function $f(x,t)$ and its solution (1.11) 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 most of the major concepts which have been developed more generally and more rigorously since then, and which are central to these notes. For example:

(i) *The Chapman-Kolmogorov Equation* occurs as Einstein's equation (1.4). 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 position $x+\delta$, 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 (1.4) is a special form, is the central dynamical equation to all Markov processes. We will discuss Markov processes in Chapter 3 of these notes.

(ii) *The Fokker-Planck Equation*: Eq. (1.10) is the diffusion equation, a special case of the Fokker-Planck 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 the pollen grain's position, if thought of as obeying a probabilistic law given by solving the diffusion equation (1.10), 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 his famous equation, which we will discuss in Chapter 4 of these notes.

(iii) *The Kramers-Moyal* and similar expansions are essentially the same as that used by Einstein to go from (1.4) (the Chapman-Kolmogorov equation) to the diffusion equation (1.10). The use of this type of approximation, which effectively replaces a process whose sample paths need not be

continuous with one where paths are continuous is a much discussed topic in the literature.

1.2.2 Langevin's Equation

Sometime after Einstein's original derivation, Langevin 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

$$\langle \tfrac{1}{2} m v^2 \rangle = \tfrac{1}{2} kT \tag{1.13}$$

(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 but he same formula as in microscopic hydrodynamics, this is $-6\pi\eta a dx/dt$, η being 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 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

$$m\frac{d^2x}{dt^2} = -6\pi\eta a\frac{dx}{dt} + X \tag{1.14}$$

and multiplying by x, this can be written

$$\frac{m}{2}\frac{d^2}{dt^2}(x^2) - mv^2 = -3\pi\eta a\frac{d(x^2)}{dt} + Xx \tag{1.15}$$

where $v = dx/dt$. We now average over a large number of different particles and use (1.13) to obtain an equation for $\langle x^2 \rangle$:

$$\frac{m}{2}\frac{d^2}{dt^2}(x^2) + 3\pi\eta a\frac{d(x^2)}{dt} = kT \tag{1.16}$$

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

$$\frac{d(x^2)}{dt} = \frac{kT}{3\pi\eta a} + Ce^{-6\pi\eta a t/m} \tag{1.17}$$

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

$$\langle x^2 \rangle - \langle x_0^2 \rangle = \left[\frac{kT}{3\pi\eta a} \right] t \tag{1.18}$$

This corresponds to (1.12) as deduced by Einstein, provided we identify

$$D = \frac{kT}{6\pi\eta a} \tag{1.19}$$

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 conspire to always be in the same direction as those of X, so that, possibly, the product could not be set equal to zero; these are really equivalent to Einstein's independence assumption. The method Langevin gives is clearly more direct, at least at first glance, and gives a very natural way of generalizing 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. In this formulation, a precise statement of the independence of X and x will lead to the calculus of stochastic differentials, which we will discuss in Chapter 4.

1.3 Birth-Death Processes

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

and A the food of the prey, the process under consideration might be

$$X+A \to 2X \tag{1.20a}$$
$$X+Y \to 2Y \tag{1.20b}$$
$$Y \to B \tag{1.20c}$$

which have the following naive, but charming interpretation. The first equation symbolizes the prey eating one unit of food, and reproducing immediately. The second equation symbolizes a predator consuming a prey (who thereby dies - this is the only death mechanism considered for the prey) and immediately reproducing. The final equation symbolizes 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 symbolizes a rate of production of X proportional to the product of x and the amount of food; the second equation a production rate of Y (and an equal rate of consumption of X) proportional to xy, and the last equation a death rate for Y, in which the rate of death of Y is simply proportional to y; thus, we may write

$$\frac{dx}{dt} = k_1 ax - k_2 xy \tag{1.21a}$$
$$\frac{dy}{dt} = k_2 xy - k_3 y \tag{1.21b}$$

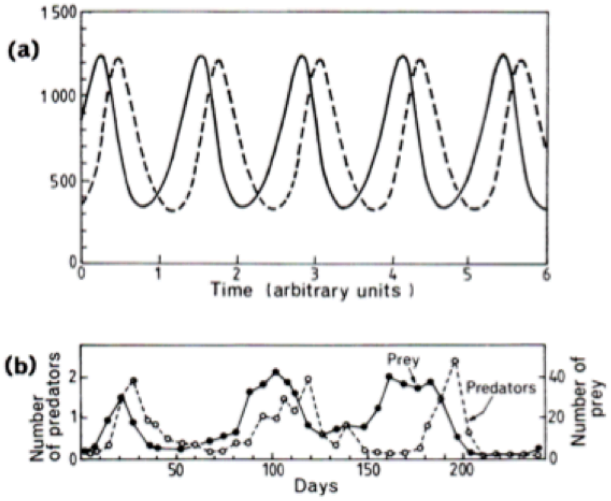

Figure 1.3: Time development in predator-prey systems. (a) Solutions of deterministic equations (1.21) (x = solid, y = dashed). (b) Data for real predator-prey system. Predator and prey are mites.

The solutions of these equations, which were independently developed by Lotka and Volterra, have very interesting oscillating solutions, as shown in Figure 1.3(a) above. These oscillations are qualitatively easily explicable. In the absence of significant number of predators, the prey population grows rapidly until the presence of so much prey for the predators to eat stimulates their rapid reproduction, at 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 predator-prey systems and example is given in Figure 1.3(b) above.

Of course, the realistic systems do not follow the solutions of differential equations exactly - they flute 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 (1.21). This is done by assuming that in an infinitesimal time Δt, the following *transition probability* laws hold.

$$Prob(x \to x+1; y \to y) = k_1 a x \Delta t \tag{1.22a}$$

$$Prob(x \to x-1; y \to y+1) = k_2 x y \Delta t \tag{1.22b}$$

$$Prob(x \to x; y \to y-1) = k_3 y \Delta t \tag{1.22c}$$

$$Prob(x \to x; y \to y) = 1 - (k_1 a x + k_2 x y + k_3 y)\Delta t \tag{1.22d}$$

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

$$\begin{aligned} &\frac{P(x, y, t+\delta t) - P(x, y, t)}{\Delta t} \\ &\quad = k_1 a (x-1) P(x-1, y, t) + k_2 (x+1)(y-1) P(x+1, y-1, t) \\ &\qquad + k_3 (y+1) P(x, y+1, t) - (k_1 a x + k_2 x y + k_3 y) P(x, y, t) \end{aligned} \tag{1.23}$$

and letting $\Delta t \to 0$, the left-hand side $= \partial P(x, y, t)/\partial t$. In writing the assumed probability laws (1.22), 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 was mentioned in Section 1.2.1. In the case of brownian motion, very convincing arguments can be made in favor of this Markov assumption. Here it is by no means clear. The concept of heredity, i.e., that the behavior 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 model has 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 (1.23), 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 greeter precision.

Equation (1.23) has no simple solution., but one major property differentiates equations like it from an equation of Langevin's type, in which the fluctuation terms is simply added to the differential equation. Solutions of (1.23) determined 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 shown in Figure 1.4 below.

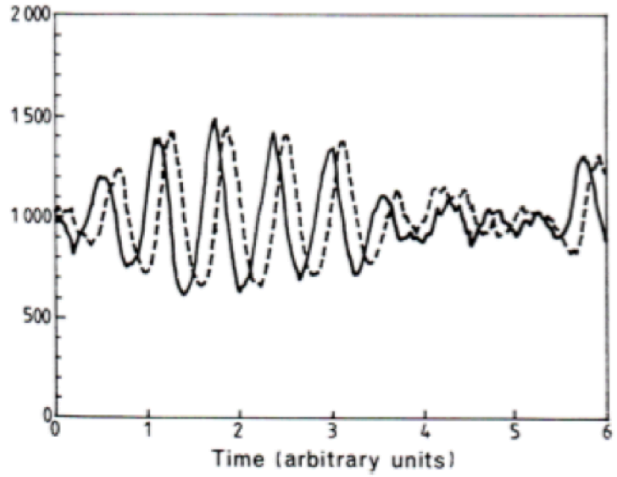

Figure 1.4: Simulation of stochastic equations (1.22).

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. In fact, a theory can be developed which can deal with a wide range of models in this category and

there is a close connection between this kind of theory and that of stochastic differential equations, which we are discussing in these notes.

1.4 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 the radio. Two principal sources of noise are shot noise and Johnson noise.

1.4.1 Shot Noise

In a vacuum tube (and in solid-state devices) we get a non steady 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

$$I9t0 = \sum_{t_k} F(t - t_k) \tag{1.24}$$

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 Figure 1.5.

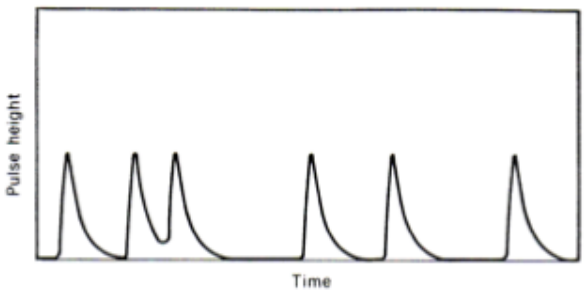

Figure 1.5: Illustration of shot noise: identical electric pulses arrive at random times.

A statistical aspect arises immediately if 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 by Schottky and Rice.

We will 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 that arrived up to 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 that the probability that an electron will 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

$$Prob(n \to n+1, \text{in time } \Delta t) = \lambda \Delta t \tag{1.25}$$

so that

$$P(n, t+\Delta t) = P(n,t)(1-\lambda\Delta t) + P(n-1,t)\lambda\Delta t \tag{1.26}$$

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

$$\frac{\partial P(n,t)}{\partial t} = \lambda[P(n-1,t) - P(n,t)] \tag{1.27}$$

which is a pure birth process. By writing

$$G(s,t) = \sum_n s^n P(n,t) \tag{1.28}$$

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

$$\frac{\partial G(s,t)}{\partial t} = \lambda(s-1)G(s,t) \tag{1.29}$$

so that

$$G(s,t) = e^{\lambda(s-1)t}G(s,0) \tag{1.30}$$

By requiring at time $t=0$ that no electrons has arrived, it is clear that $P(0,0)$ is 1 and $P(n,0$ is zero for all $n \gg 1$, so that $G(s,0)=1$. Expanding the solution (1.30) in powers of s, we find

$$P(n,t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \tag{1.31}$$

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

$$P(n,t) = Prob\{N(t) = n\} \tag{1.32}$$

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

$$\mu(t) = \frac{dN(t)}{dt} \tag{1.33}$$

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

$$\mu(t) = \sum_k \delta(t - t_k) \tag{1.34}$$

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

$$I(t) = \int_{-\infty}^{\infty} dt'\, F(t - t')\mu(t') \tag{1.35}$$

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 imply 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 standard (Green function) form

$$F(t) = \begin{cases} e^{-\alpha t} & (t > 0) \\ 0 & (t < 0) \end{cases} \tag{1.36}$$

so that (1.35) can be rewritten as

$$I(t) = \int_{-\infty}^{t} dt'\, qe^{-\alpha(t-t')} \frac{dN(t')}{dt'} \tag{1.37}$$

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

$$\frac{dI(t)}{dt} = \left[qe^{-\alpha(t-t')} \frac{dN(t')}{dt'} \right]_{t'=t} + \int_{-\infty}^{t} dt'\, (-\alpha q) e^{-\alpha(t-t')} \frac{dN(t')}{dt'} \tag{1.38}$$

so that

$$\frac{dI(t)}{dt} = -\alpha I(t) + q\mu(t) \tag{1.39}$$

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 (1.34). However, the mean of $\mu(t)$ is nonzero, in fact, from (1.33)

$$\langle \mu(t)dt \rangle = \langle dN(t) \rangle = \lambda dt \tag{1.40}$$

$$\langle [dN(t) - \lambda dt]^2 \rangle = \lambda dt \tag{1.41}$$

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

$$d\eta(t) = dN(t) - \lambda dt \tag{1.42}$$

so that the stochastic differential equation (1.39) takes the form

$$dI(t) = [\lambda q - \alpha I(t)]dt + q d\eta(t) \tag{1.43}$$

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

$$\frac{d\langle I(t)\rangle}{dt} = \lambda q - \alpha\langle I(t)\rangle \tag{1.44}$$

$$\frac{1}{2}\frac{d\langle I^2(t)\rangle}{dt} = \lambda q\langle I(t)\rangle - \alpha\langle I^2(t)\rangle \tag{1.45}$$

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

$$\langle I(\infty)\rangle = \frac{\lambda q}{\alpha} \quad \text{and} \tag{1.46}$$

$$\langle I^2(\infty)\rangle = \left(\frac{\lambda q}{\alpha}\right)^2 \tag{1.47}$$

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 (1.40-1.43) has been chosen deliberately. In deriving (1.45), one uses ordinary calculus, i.e., one writes

$$d(I^2) \equiv (I + dI)^4 - I^2 = 2IdI + (dI)^2 \tag{1.48}$$

and then one drops the $(dI)^2$ term as being of second order in dI. But now look at (1.41): this is equivalent to

$$\langle d\eta(t)^2\rangle = \lambda dt \tag{1.49}$$

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,

$$d\eta(t) = dN(t) - \lambda dt \tag{1.50}$$

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 (1.43,1.48, 1.49):

$$\langle d(I^2)\rangle = 2\langle I\{[\lambda q - \alpha I]dt + qd\eta(t)\}\rangle + \langle\{[\lambda q - \alpha I]dt + qd\eta(t)]\}^2\rangle \tag{1.51}$$

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

$$\frac{1}{2}d\langle d(I^2)\rangle = \left[\lambda q\langle I\rangle - \alpha\langle I^2\rangle + \frac{q^2\lambda}{2}\right]dt \tag{1.52}$$

and this gives

$$\langle I^2(\infty)\rangle - \langle I(\infty)\rangle^2 = \frac{q^2\lambda}{2\alpha} \tag{1.53}$$

Thus, there are fluctuations from this point of view, as $t \to \infty$. The extra term in (1.52) as compared to (1.45) 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 these notes that problems such as we have just considered are very widespread. 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 Chapter 4 which will discuss situations in which fluctuations are Gaussian.

1.4.2 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 first introduced by Taylor as

$$G(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\, x(t)x(t+\tau) \tag{1.54}$$

This is the time average of a two-time product over an arbitrary large time T, which is then allowed to become infinite.

At this time, single purpose auotcorrelators exist, which sample data and directly construct the autocorrelation function of a desired process, from laser light scattering signals to bacterial counts.

A more traditional approach to autocorrelation is to compute the spectrum of the quantity $x(t)$. This is defined in two stages. First, define

$$y(\omega) = \int_0^T dt\, e^{-i\omega t} x(t) \tag{1.55}$$

then the spectrum is defined by

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} |y(\omega)|^2 \tag{1.56}$$

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

$$S(\omega) = \lim_{T\to\infty} \left[\frac{1}{\pi} \int_0^T \cos(\omega\tau) d\tau \frac{1}{T} \int_0^{T-\tau} x(t)x(t+\tau)dt \right] \tag{1.57}$$

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

$$S(\omega) = \frac{1}{\pi} \int_0^\infty \cos(\omega\tau) G(\tau) d\tau \tag{1.58}$$

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

$$G(-\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\tau}^{T-\tau} dt\, x(t+\tau)x(t) = G(\tau) \tag{1.59}$$

so we obtain

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty e^{-i\omega\tau} G(\tau) d\tau \tag{1.60}$$

with the corresponding inverse

$$G(\tau) = \int_{-\infty}^\infty e^{i\omega\tau} S(\omega) d\omega \tag{1.61}$$

This result is known as the *Wiener-Khinchin theorem* 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 (FFT) and high-speed computers is relatively straightforward.

1.4.3 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* (Section 3.7.1).

If we have a fluctuating quantity $x(t)$, then we can consider the average

$$\langle x(t)x(t+\tau)\rangle = G(\tau) \tag{1.62}$$

this result is a consequence of our ergodic assumptions.

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

$$x(t) == \int d\omega\, c(\omega)e^{i\omega t} \tag{1.63}$$

and consequently,

$$c(\omega) = \frac{1}{2\pi}\int dt\, x(t)e^{-i\omega t} \tag{1.64}$$

Note that $x(t)$ era implies

$$c(\omega) = c^*(-\omega) \tag{1.65}$$

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 $x(t_1), x(t_2), ..., x(t_n)$ are equal to those of $x(t_1+\Delta), x(t_2+\Delta), ..., x(t_n+\Delta)$.

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

$$\langle c(\omega)\rangle = \frac{1}{2\pi}\int dt\, \langle x\rangle e^{-i\omega t} = 0 \tag{1.66}$$

$$\begin{aligned}\langle c(\omega)c^*(\omega')\rangle &= \frac{1}{(2\pi)^2}\int\int dt\, dt'\, e^{-i\omega t+i\omega' t'}\langle x(t)x(t')\rangle \\ &= \frac{1}{2\pi}\delta(\omega-\omega')\int d\tau\, e^{i\omega\tau}G(\tau) \\ &= \delta(\omega-\omega')S(\omega)\end{aligned} \tag{1.67}$$

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 $t-t'$.

1.4.4 Johnson Noise and Nyquist's Theorem

Two brief and elegant papers appeared in 1928 in which Johnson demonstrated experimentally that an electric resistor automatically generated fluctuations of

electric voltage, and Nyquist demonstrated its theoretical derivation, in complete agreement with Johnson's experiment. The principle involved was already known by Schottky 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 will not be carried out here. We note, however, 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. The experimental result found was the following. We have and electric resistor or 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,

$$\langle E^2(\omega)\rangle = 2RkT \tag{1.68}$$

This result is known 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 largest 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 the 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 (1.56) and may be summarized by saying that the voltage spectrum $S(\omega)$ is given by

$$S(\omega) = 2RkT \tag{1.69}$$

that is, the spectrum is flat, i.e., a constant function of ω. In the case of light, the frequencies correspond to different colors of light. If we perceive light to be white, it is found that in practice all colors 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

$$\int_{\omega_1}^{\omega_2} d\omega\, S(\omega)/R = 2kT(\omega_2 - \omega_1) \tag{1.70}$$

so that the total power dissipated at all frequencies is infinite. Nyquist realized this, and noted that, in practice, there would be quantum corrections which would, at room temperature, make the spectrum flat only up to $7 \times 10^{13}\, Hz$, which is not detectable in practice, in a radio situation. The actual power dissipated 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 in Section 1.4, we have

$$\begin{aligned}\langle E(t+\tau)E(t)\rangle &= G(\tau) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau} 2RkT \\ &= 2RkT\delta(\tau) \end{aligned} \tag{1.71}$$

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

$$S(\omega) = \frac{2RkT}{\omega^2\tau_C^2 + 1} \tag{1.72}$$

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

$$\langle E(t+\tau)E(t)\rangle = \frac{RkT}{\tau_c} e^{-\tau/\tau_c} \tag{1.73}$$

so that the autocorrelation function vanshes only for $\tau \gg \tau_c$, which is called the *correlation time* of the fluctuating voltage. Thus, the delay function correlation function appears as an idealization, only valid on a sufficiently long time scale. Some examples are shown below.

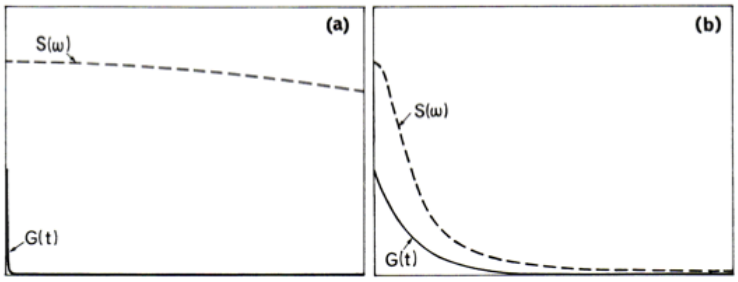

Figure 1.6: Correlation Functions(solid) and corresponding spectraI(dashed) for (a) short correlation time corresponding to an almost flat spectrum; (b) long correlation time, giving a quite rapidly decreasing spectrum.

This discussion is very reminiscent of Einstein's assumption regarding Brownian motion and of the behavior of Langevin's fluctuating force. The idealized *white noise* will play a highly important role in these notes 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 that 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 of these ideas will be discussed in Chapter 4.

White noise, as we have noted above, does not exist as a physically realizable process and the rather singular behavior it exhibits does not arise in any realizable context. It is, however, fundamental in a mathematical, and indeed in a physical sense, in that it is an idealization of very many processes that do occur. The slightly strange rule 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 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 discussions in these notes.

1.5 The Stock Market

The equations of Brownian motion were in fact6 first derived by Bachelier in a doctoral thesis 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* $x = C - X_0$ 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

$$p_{cx,t_1+t_2} = \int p_{x,t_1} p_{z-x,t_2} dz \tag{1.74}$$

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

$$p = Ae^{-B^2x^2} \tag{1.75}$$

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

> "The definitive expression for the probability is thus
>
> $$p = \frac{1}{2\pi k\sqrt{t}} e^{-\frac{x^2}{4\pi k^2 t}} \tag{1.76}$$

The mathematical expectation

$$\int_0^\infty pxdx = k\sqrt{t} \tag{1.77}$$

is proportional to the square root of the time"

Bachelier gave another derivation 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 1900's. The value of the woek 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.

1.5.1 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 fianc circles until the 1960s, when Samuelson decided to develop the approach further. Damuelson rather unfairly criticized bachelor 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 is what has become known as the *return* on the share price, given by

$$r = \frac{x}{X} \tag{1.78}$$

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

$$p = \log X \tag{1.79}$$

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 today. Samuelson termed the process *geometric Brownian motion* or alternatively *economic Brownian motion.*

1.5.2 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 future delivery of wheat. Rather than deal with an individual grower, the baker can buy the ingrown 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 but it with the view to selling it at a future date, without ever having 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 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.

1.5.3 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 mice of cash, stocks and options could be devised in which the short term fluctuations in 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:

> "A great economist of an earlier generation said that, usefull 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 sector 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 by Black and Scholes and a different derivation was given by Merton. The formula depends critically on the 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-dramatized view of modern financial theory.

1.5.4 Heavy Tailed Distributions

There is however, no doubt that the geometric Brownian motion model of financial markets is not exact, and even misses out on 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* is an impressive theoretical edifice built on a flawed foundation of brownian motion, but so far it appears to be the most practical method for modeling 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.

Chapter 2

Probability Concepts

In the last 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 intneion of this chapter is to provide some background, and to present a number of essential results. We will not give a through outline of mathematical probability, but only those parts that we need.

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 $\mathbf{x}$ 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 realizations of *events*. More abstractly, an event is simply a member of a certain space, which in the cases most practically occurring can be characterized by a vector of integers

$$\mathbf{n} = (n_1, n_2, n_3, ...) \tag{2.1}$$

or a vector of real numbers

$$\mathbf{x} = (x_1, x_2, x_3, ...) \tag{2.2}$$

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

$$\omega \in A \tag{2.3}$$

to indicate that the event ω is one of the 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

$$\bar{\omega} \in A(25) \tag{2.4}$$

More significantly, suppose we define the set of events $A(\mathbf{r}, \Delta V)$ that a molecule is within a volume element ΔV centered on a point $\mathbf{r}$. In this case, the practical significance of working in terms of sets of events becomes clear, because we should normally be able to determine whether or not a molecule is within a neighborhood ΔV of $\mathbf{r}$, but to determine whether the particle is exactly at $\mathbf{r}$ is impossible. Thus, if we define the event $\omega(\mathbf{y})$ that the molecule is at point $\mathbf{y}$, it makes sense to ask whether

$$\mathbf{y} \in A(\mathbf{r}, \Delta V) \tag{2.5}$$

and to assign a certain probability to the *set* $A(\mathbf{r}, \Delta V)$, which is to be interpreted as the probability of the occurrence of (2.5).

2.2 Probabilities

Most people have an intuitive conception of a probability, based on their own experience. However, a precise formulation of intuitive concepts is fraught with difficulties, and it has been found most convenient to axiomatize probability theory as an essentially abstract science, in which a probability measure $P(A)$ is *assigned* to every set A, in the space of events, including

$$\text{the set of all events: } \Omega \tag{2.6}$$

$$\text{the set of no events: } \emptyset \tag{2.7}$$

in order to define probability, we need our sets of events to form a closed system (known by mathematicians as a σ-algebra) under the set theoretic operations of union and intersection.

2.2.1 Probability Axioms

We introduce the probability of A, $P(A)$, as a function of A satisfying the following *probability axioms*:

(i)

$$P(A) \geq 0 \text{ for all } A \tag{2.8}$$

(ii)

$$P(\Omega) = 1 \tag{2.9}$$

(iii) if $A_i (i = 1, 2, 3, ...)$ is a countable (but possibly infinite) collection of *non overlapping sets*, i.e., such that

$$A_i \cap A_j = \emptyset \tag{2.10}$$

for all $i \neq j$, then

$$P\left(\underset{i}{\cup} A\right) = \sum_i P(A_i) \tag{2.11}$$

(iv) if λ is the complement of A, i.e., the set of all events not contained in A, then

$$P(\tilde{A}) = 1 - P(A) \tag{2.12}$$

(v)

$$P(\emptyset) = 0 \tag{2.13}$$

2.2.2 The Meaning of $P(A)$

There is no way of making probability theory correspond to reality without requiring a certain degree of intuition. The probability $P(A)$, as axiomatized above, is the intuitive probability that an "*arbitrary*" event ω, i.e., an event ω "*chosen at random*", will satisfy $\omega \in A$. Or more explicitly, if we choose an event "*at random*" from Ω N times, the relative frequency that the particular event chosen will satisfy $\omega \in A$ approaches $P(A)$ as the number of times, N, we choose the event , approaches infinity. The number of choices N can be visualized as being done one after the other("*independent*" tosses of one die) or at the same time (N dice are thrown at the same time "*independently*"). All definitions of this kind must be intuitive, as we can see by the way the undefined terms ("*arbitrary*", "*at random*", "*independent*") keep turning up. By eliminating what we now think of as intuitive ideas and axiomatizing probability, Kolmogorov cleared the road for a rigorous development of mathematical probability. But the circular definition problems posed by wanting an intuitive understanding remain. The simplest way of looking at axiomatic probability is as a formal method of manipulating probabilities using the axioms. In order to apply the theory, the probability space must be defined *and* and the probability measure P assigned. These are *a priori probabilities*, which are simply assumed. Examples of such a priori probabilities abound in applied disciplines. For example, in equilibrium statistical mechanics one assigns equal probabilities to equal volumes of phase space. Einstein's reasoning in Brownian motion assigned a probability $\phi(\Delta)$ to the probability of a "push" Δ from a position x at time t.

The task of applying probability is (i) to assume some set of *a priori* probabilities which seem reasonable and deduce results from this and from the structure of the probability space, (ii) to measure experimental results with some apparatus which is constructed to measure quantities in accordance with these a priori probabilities.

The structure of the probability space is very important, especially when the space of events is compounded by the additional concept of time. This extension makes the effective probability space infinite-dimensional, since we can construct events such as "the particle was at points $\mathbf{x_n}$ at times t_n, $n = 1, 2, ..., \infty$".

2.2.3 The Meaning of the Axioms

Any intuitive concept of probability gives rise to nonnegative probabilities, and the probability that an arbitrary event is contained in the set of all events must be 1 no matter what our definition of the word arbitrary. Hence, axioms (i) and (ii) are understandable. The heart of the matter lies in axiom (iii). Suppose we are dealing with only 2 sets A and B, and $A \cap B = \emptyset$. This means there are *no* events contained in both A and B. Therefore, the probability that $\omega \in A \cup B$ is the probability that *either* $\omega \in A$ or $\omega \in B$. Intuitive considerations tell us this probability is the sum of the individual probabilities, i.e.,

$$P(A \cup B) \equiv P(\omega \in A) \text{ or } P(\omega \in B) = P(A) + P(B) \tag{2.14}$$

(notice that this not a proof - merely an explanation).

The extension now to any finite number of non overlapping sets is obvious, but the extension to any *countable* number of non overlapping sets requires some comment.

The extension must be made restrictive because of the existence of sets labelled by a continuous index, for example $\mathbf{x}$, the position in space. The probability of a molecule being in the set whose only element is $\mathbf{x}$ is zero; but the probability of being in a region R of finite volume is nonzero. The region R is a union of sets of the form $\{\mathbf{x}\}$ - but not a *countable* union. Thus, axiom (iii) is not applicable and the probability of being in R is *not* equal to the sum of the probabilities of being in $\{\mathbf{x}\}$.

2.2.4 Random Variables

The concept of a random variable is a notational convenience which is central to these notes. Suppose we have an abstract probability space whose events can be written $\mathbf{x}$. Then we can introduce the random variable $F(\mathbf{x})$ which is a function of $\mathbf{x}$, which takes on certain values for each $\mathbf{x}$. In particular, the identity function of $\mathbf{x}$, written $\mathbf{X}(\mathbf{x})$ is of interest; it is given by

$$\mathbf{X}(\mathbf{x}) = \mathbf{x} \tag{2.15}$$

We will normally use capitals in these notes to denote random variables and small letters $\mathbf{x}$ to denote their values whenever it is necessary to make a distinction.

Very often, we have some quite different underlying probability space Ω with

values ω, and talk about $\mathbf{X}(\omega)$ which is some function of ω, and then omit explicit mention of ω. This can be for either of two reasons:

(i) we specify the events by the values of $\mathbf{x}$ anyway, i.e., we identify $\mathbf{x}$ and ω

(ii) the underlying events ω are too complicated to describe, or sometimes, even to know

For example, in the case of the position of a molecule in a liquid, we really should interpret each ω as being capable of specifying all the positions, momenta and orientations of each molecule in that volume of the liquid; but this is simply too difficult to write down, and often unnecessary.

One great advantage of introducing the concepts of a random variable is the simplicity with which one may handle functions of random variables, e.g., $\mathbf{X^2}$, $\sin(\mathbf{a}\cdot\mathbf{X})$, etc, and compute means and distributions of these. Further, by defining stochastic differential equations, one can also quite simply talk about the time development of random variables in a way which is quite analogous to the classical description by means of differential equations of non probabilistic systems.

2.3 Joint and Conditional Probabilities: Independence

2.3.1 Joint Probabilities

We explained in Section 2.2.3 how the occurrence of mutually exclusive events is related to the concept of nonintersecting sets. We must now consider the concept $P(A\cap B)$, where $A\cap B$ is nonempty. An event ω which satisfies $\omega\in A$ will only satisfy $\omega\in A\cap B$ if $\omega\in B$ as well.

$$\text{Thus, } P(A\cap B)=P\{(\omega\in A)\text{ and }(\omega\in B)\} \tag{2.16}$$

and $P(A\cap B)$ is called the *joint probability* that the event ω is contained in both classes, or, alternatively, that both the events $\omega\in A$ and $\omega\in B$ occur. Joint probabilities occur naturally in the context of these notes in two ways:

(i) When the event is specified by a vector, e.g., m mice and n tigers. The probability of this event is the joint probability of [m mice (and any number of tigers)] and [n tigers (and any number of mice)]. All vector specifications are implicitly joint probabilities in this sense.

(ii) *When more than one time is considered*: what is the probability that (at time t_1 there are m_1 tigers and n_1 mice) and (at time t_2 there are m_2 tigers and n_2 mice). To consider such a probability, we have effectively created out of the events at time t_1 and the vents at time t_2, *joint events* involving one event at each time. In essence, there is no difference between these two cases except for the fundamental dynamical role of time.

2.3.2 Conditional Probabilities

We may specify conditions on the events we are interested in and consider only these, e.g., the probability of 21 buffaloes given that we know there are 100 lions. What does this mean? Clearly, we will be interested in those events contained in the set $B = \{\text{all events where exactly 100 lions occur}\}$. This means that we need to define conditional probabilities, which are defined only on the collection of all sets contained in B. We define the conditional probability as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.17}$$

and this satisfies our intuitive conception that the conditional probability that $\omega \in A$ (given that we know $\omega \in B$), is given by dividing the probability of joint occurrence by the probability ($\omega \in b$).

We can define in both directions, i.e., we have

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \tag{2.18}$$

There is no particular conceptual difference between, say, the probability of {(21 buffaloes) given (100 lions)} and the reversed concept. However, when two times are involved, we do see a difference. For example, the probability that a particle is at position $\mathbf{x_1}$ at time t_1, given that it was at $\mathbf{x_2}$ at the *previous* time t_2, is a very natural thing to consider; indeed, it will turn out to be a central concept in these notes. The converse sounds strange, i.e., the probability that a particle is at position $\mathbf{x_1}$ at time t_1, given that it will be at position $\mathbf{x_2}$ at a *later* time t_2. It smacks of clairvoyance - we cannot conceive of any natural way in which we would wish to consider it, although it is, in principle, a quantity very similar to the "natural" conditional probability, in which the condition precedes the event under consideration.

The natural definition has already occurred in these notes, for example, the $\phi(\Delta)d\Delta$ of Einstein (Section 1.2.1) is the probability that a particle at x at time t will be in the range $[x + \Delta, x + \delta + d\Delta]$ at time $t + \tau$. Our intuition tells us as it told Einstein (as can be seen by reading the extract from his paper) that this kind of conditional probability is directly related to the time development of a probabilistic system.

2.3.3 Relationship Between Joint Probabilities of Different Orders

Suppose we have a collection of sets B_i such that

$$B_i \cap B_j = \emptyset \quad i \neq j \tag{2.19}$$

$$\underset{i}{\cup} B_i = \Omega \tag{2.20}$$

so that the sets divide up the space Ω into non overlapping subsets.

Then

$$\bigcup_i (A \cap B_i) = A \cap \bigcup_i B_i = A \cap \Omega = A \tag{2.21}$$

Now, using the probability axiom (iii), we see that the $A \cap B_i$ satisfy the conditions on the A_i used there, so that

$$\sum_i P(A \cap B_i) = P\left[\bigcup_i (A \cap B_i)\right] \tag{2.22}$$

$$= P(A) \tag{2.23}$$

and thus

$$\sum_i P(A|B_i)P(B_i) = P(A) \tag{2.24}$$

Thus, summing over all mutually exclusive possibilities of B in the joint probability eliminates that variable.

Hence, in general,

$$\sum_i P(A_i \cap B_j \cap C_k \cdots) = P(B_j \cap C_k \cdots) \tag{2.25}$$

The result (2.24) has very significant consequences in the development of the theory of stochastic processes, which depend heavily on joint probabilities.

2.3.4 Independence

We need a probabilistic way of specifying what we mean by independent events. Two sets of events A and B should represent independent sets of events if the specification that a particular event is contained in B has no influence on the probability of that event belonging to A. Thus, the conditional probability $P(A|B)$ should be independent of B, and hence

$$P(A \cap B) = P(A)P(B) \tag{2.26}$$

In the case of several events, we need a somewhat stronger specification. The events $(\omega \in A_i)\,(i = 1, 2, ..., n)$ will be considered to be independent if for any susbset $(i_1, i_2, ..., i_k)$ of the set $(1, 2, ...n)$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}).....P(A_{i_k}) \tag{2.27}$$

It is important to require factorization for all possible combination, as in (2.27). For example, for three sets A_i, it is quite conceivable that

$$P(A_i \cap A_j) = P(A_i)P(A_j) \tag{2.28}$$

for all different i and j, but also that

$$A_1 \cap A_2 = A_2 \cap A_3 = A_3 \cap A_1 \tag{2.29}$$

as shown in the figure below.

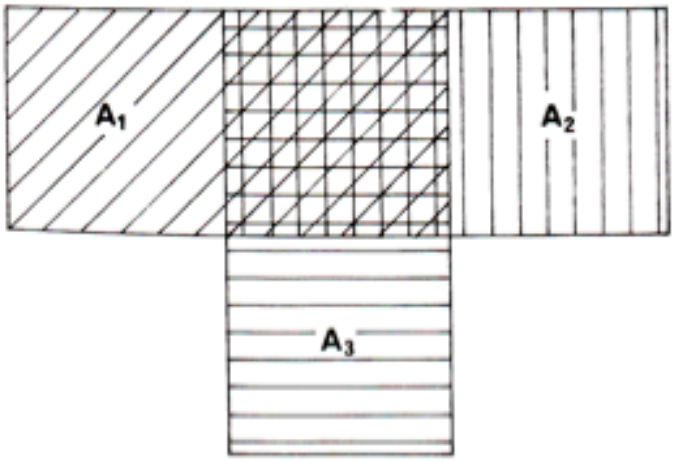

Figure 2.1: Illustration of statistical independence in pairs, but not in threes. In the three sets $A_j \cap A_i$ is, in all cases, the central region. By appropriate choice of probabilities, we can arrange for $P(A_i \cap A_j) = P(A_i)P(A_j)$.

This requires

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_2 \cap A_3 \cap A_3) = P(A_2 \cap A_3) \\ &= P(A_2)P(A_3) \neq P(A_1)P(A_2)P(A_3) \end{aligned} \tag{2.30}$$

We can see that the occurrence of $\omega \in A_2$ and $\omega \in A_3$, necessarily implies the occurrence of $\omega \in A_1$. In this sense, the events are obviously not independent.

Random variable $X_1, X_2, X_3, ...$, will be said to be independent random variables, if for all sets of the form $A_i = (x \text{ such that } a_i \leq x \leq b_i)$ the events $X_1 \in A_1, X_2 \in A_2, X_3 \in A_3, ...$ are independent events. This will mean that all values of the X_i are assumed independently of those of the remaining X_i.

2.4 Mean Values and Probability Density

The mean value of a random variable $R(\omega)$ in which the basic events ω countably specifiable is given by

$$\langle R \rangle = \sum_{\omega} P(\omega)R(\omega) \tag{2.31}$$

where $P(\omega)$ means the probability of the set containing only the single event ω. In the case of a continuous variable, the probability axioms enable us to define a probability density $p(\omega)$ such that if $A(\omega_0, d\omega_0)$ is the set

$$(\omega_0 \leq \omega < \omega_0 + d\omega_0) \tag{2.32}$$

then

$$p(\omega_)d\omega_0 = P[A(\omega_0, d\omega_0)] \tag{2.33}$$
$$\equiv p(\omega_0, d\omega_0) \tag{2.34}$$

The last is a notation often used by mathematicians. Details of how this is done are nicely explained by Feller. In this case,

$$\langle R\rangle = \int_{\omega\in|OMega} d\omega\, R(\omega)p(\omega) \tag{2.35}$$

One can often (as mentioned in Section 2.2.4) use R itself to specify the event, so we will often write

$$\langle R\rangle = \int dR\, Rp(R) \tag{2.36}$$

Obviously, $p(R)$ is not the same function of R as $p(\omega)$ is of ω - more precisely

$$p(R_0)dR_0 = P[R_0 < R < R_0 + dR_0] \tag{2.37}$$

2.4.1 Determination of Probability Density by Means of Arbitrary Functions

Suppose for every function $f(R)$ we know

$$\langle f(R)\rangle = \int dR\, f(R)p(R) \tag{2.38}$$

then we know $p(R)$, which is known as a *probability density*. The proof follows by choosing

$$f(R) = \begin{cases} 1 & R_0 \le R < R_0 + dR_0 \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

Because the expectation of an arbitrary function is sometimes a little easier to work with than a density, this relation will be used occasionally in these notes.

Notation: The notation $\langle A\rangle$ for the expectation used in these notes is a physicist's notation. The most common mathematical notation is $E(A)$, which is in my opinion a little less intuitive.

2.4.2 Sets of Probability Zero

If a density $p(R)$ exists, the probability that R is in the interval $(R_0, R_0 + dR)$ goes to zero with dR. Hence, the probability that R has *exactly* the value R_0 is zero; and similarly for any other value.

Thus, in such a case, there are sets $S(R_i)$, each containing only one point R_i, which have zero probability. From probability axiom (iii), any countable union

of such sets, i.e., any set containing only a countable number of points (e.g., all rational numbers) has probability zero. In general, all equalities in probability theory are at best only "*almost certainly true*", i.e., they may be untrue on sets of probability zero. Alternatively, one says, for example,

$$X = Y \text{ with probability 1} \tag{2.40}$$

which is by no means the same as saying that

$$X(R) = Y(R) \text{ for all } R \tag{2.41}$$

Of course, if the theory is to have any connection with reality, events with probability zero do not occur.

In particular, notice that our previous result if inspected carefully, only implies that we know $p(R)$ only with probability 1, given that we know $\langle f(R) \rangle$ for all $f(R)$.

2.5 The Interpretation of Mean Values

The question of what to measure in a probabilistic system is nontrivial. In practice, one measures either a set of individual values of a random variable (the number of animals of a certain kind in a certain region at certain points in time; the electric current passing through a given circuit element in each of a large number of replicas of that circuit, etc.) or alternatively, the measuring procedure may implicitly construct an average of some kind. For example, to measure an electric current, we may measure the electric charge transferred and divide by the time taken - this gives a measure of the average number of electrons transferred per unit time. It is important to note the essential difference in this case, that it will not normally be possible to measure anything other than a few selected averages and thus, higher moments (for example) will be unavailable.

In contrast, when we measure individual events (as in counting animals), we can then construct averages of the observables by the obvious methods

$$\bar{X}_N = \frac{1}{N} \sum_{n=1}^{N} X(n) \tag{2.42}$$

The quantities $X(n)$ are the individual observed values of the quantity X. We expect that as the number of samples N becomes very large, the quantity $\bar{X}_N$ approaches the mean $\langle X \rangle$ and that, in fact,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f[X(n)] = \lim_{N \to \infty} \overline{f(X)}_N = \langle f(X) \rangle \tag{2.43}$$

and such a procedure will determine the probability density function $p(x)$ of X if we carry out this procedure for all functions f. The validity of this procedure

depends on the degree of independence of the successive measurements and is dealt with in Section 2.5.2.

In the case where only averages themselves are directly determined by the measuring method, it will not normally be possible to measure $X(n)$ and therefore, it will not, in general, be possible to determine $\overline{f(X)}_N$. All that will be available will be $f(\bar{X}_N)$ - quite a different thing unless f is linear. We can often find situation in which measurable quantities are related (by means of some theory) to mean values of certain functions, but to hope to measure, for example, the mean value of an arbitrary function of the number of electrons in a conductor is quite hopeless. The mean number - yes, and indeed even the mean square number, but the measuring methods available are not direct. We do *not* enumerate the individual numbers of electrons at different times and hence arbitrary functions are not attainable.

2.5.1 Moments, Correlations, and Covariances

Quantities of interest are given by the *moments* $\langle X^n \rangle$ since these are often easily calculated. However, probability densities must always vanish as $x \to \pm\infty$, so we see that higher moments tell us only about the properties of unlikely large values of X. In practice we find that the most important quantities are related to the first and second moments. In particular, for a single variable X, *the variance* is defined by

$$var[X] \equiv [\sigma[X]]^2 \equiv \langle [X - \langle X \rangle]^2 \rangle \tag{2.44}$$

and as is well known, the *variance* $var[X]$ or its square root the *standard deviation* $\sigma[X]$, is a measure of the degree to which the values of X deviate from the mean value $\langle X \rangle$.

In the case of several variables, we define the *covariance matrix* as

$$\langle X_i, X_j \rangle \equiv \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle \equiv \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \tag{2.45}$$

Obviously,

$$\langle X_i, X_j \rangle = var[X_i] \tag{2.46}$$

If the variables are independent *in pairs*, the covariance matrix is diagonal.

2.5.2 The Law of Large Numbers

As an application of the previous concepts, let us investigate the following model of measurement. We assume that we measure the same quantity N times, obtaining sample values of the random variable $X(n) :', (n = 1, 2, ..., N)$. Since these are all measurements of the same quantity at successive times, we assume that for every n, $X(n)$ has the same probability distribution but we do not

assume the $X(n)$ to be independent. However, provided that the covariance matrix $\langle X(n), X(m)\rangle$ vanishes sufficiently rapidly as $|n-m| \to \infty$, then defining

$$\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X(n) \tag{2.47}$$

we will show

$$\lim_{N\to\infty} \bar{X}_N = \langle X\rangle \tag{2.48}$$

It is clear that

$$\langle \bar{X}_N\rangle = \langle X\rangle \tag{2.49}$$

We now calculate the variance of $\bar{X}_N$ and show that as $N \to \infty$ it vanishes under certain conditions:

$$\langle \bar{X}_N \bar{X}_N\rangle^2 = \frac{1}{N^2}\sum_{n,m=1}^{N} \langle X_n, X_m\rangle \tag{2.50}$$

Provided $\langle X_n, X_m\rangle$ falls off sufficiently rapidly as $|n-m| \to \infty$, we find

$$\lim_{N\to\infty} (var[\bar{X}_N]) = 0 \tag{2.51}$$

so that $\lim_{N\to\infty} \bar{X}_N$ is a deterministic variable equal to $\langle X\rangle$.

Two models of $\langle X_n, X_m\rangle$ can be chosen.

(a)

$$\langle X_n, X_m\rangle \sim K\lambda^{|n-m|}, \quad (\lambda < 1) \tag{2.52}$$

for which one finds

$$var[\bar{X}_N] \sim \frac{2K}{N^2}\left(\frac{\lambda^{N+2} - N(\lambda-1) - \lambda}{(\lambda-1)^2}\right) - \frac{K}{N} \to 0 \tag{2.53}$$

(b)

$$\langle X_n, X_m\rangle \sim |n-m|^{-1}, \quad (n \neq m) \tag{2.54}$$

and one approximately finds

$$var[\bar{X}_N] \sim \frac{2}{N}\log N - \frac{1}{N} \to 0 \tag{2.55}$$

In both these cases, $var[X_N] \to 0$, but the rate of convergence is very different. Interpreting n, m as the times at which the measurement is carried out, one sees that even very slowly decaying correlations are permissible. The law of large numbers comes in many forms. The central limit theorem is an even more precise result in which the limiting distribution function of $\bar{X}_N - \langle X\rangle$ is determined (see Section 2.8.2).

2.6 Characteristic Function

One would like a condition where the variables are independent, not just in pairs. To this end (and others) we define the characteristic function.

If $\mathbf{s}$ is the vector $(s_1, s_2, ..., s_n)$, and $\mathbf{X} = (X_1, X_2, ..., X_n)$ is a vector of random variables, then the characteristic function (or moment generating function) is defined by

$$\phi(\mathbf{s}) = \langle \exp(i\mathbf{s}\cdot\mathbf{X})\rangle = \int d\mathbf{x}\, p(\mathbf{x}) \exp(i\mathbf{s}\cdot\mathbf{x}) \tag{2.56}$$

the characteristic function has the following properties:

(i) $\phi(\mathbf{0}) = 1$

(ii) $|\phi(\mathbf{s})| \leq 1$

(iii) $\phi(\mathbf{s})$ is a uniformly continuous function of its arguments for all finite real $\mathbf{s}$

(iv) If the *moments* $\langle \prod_i X_i^{m_i} \rangle$ exist, then

$$\left\langle \prod_i X_i^{m_i} \right\rangle = \left[\prod_i \left(-i \frac{\partial}{\partial s_i} \right)^{m_i} \phi(\mathbf{s}) \right]_{\mathbf{s}=\mathbf{0}} \tag{2.57}$$

(v) A sequence of probability densities converges to limiting probability density if and only if the corresponding characteristic functions converge to the corresponding characteristic function of the limiting probability density.

(vi) Fourier inversion formula

$$p(\mathbf{x}) = \frac{1}{(2\pi)^n} \int d\mathbf{s}\, \phi(\mathbf{s}) \exp(-i\mathbf{x}\cdot\mathbf{s}) \tag{2.58}$$

Because of this inversion formula $\phi(\mathbf{s})$ determines $p(\mathbf{x})$ with probability 1. Hence, the characteristic function does truly *characterize* the probability density.

(vii) Independent random variables: from the definition of independent random variables in Section 2.3.4, it follows that the variables $X_1, X_2, ...$ are independent if and only if

$$p(x_1, x_2, ..., x_n) = p(x_1)p(x_2).....p(x_n) \tag{2.59}$$

in which case,

$$\phi(s_1, s_2, ..., s_n) = \phi(s_1)\phi(s_2).....\phi(s_n) \tag{2.60}$$

(viii) Sum of independent random variables: if $X_1, X_2, ...$ are independent random variables and if

$$Y = \sum_{i=1}^{N} X_i \tag{2.61}$$

and the characteristic function of Y is

$$\phi_y(s) = \langle \exp(isY) \rangle \tag{2.62}$$

then

$$\phi_y(s) = \prod_{i=1}^{n} \phi_i(s) \tag{2.63}$$

The characteristic function plays an important role in these notes which arises from the convergence property (v), which allows us to perform limiting processes on the characteristic function rather than the probability distribution itself, and often makes proofs easier. Further, the fact that the characteristic function is truly characteristic, i.e., the inversion formula (vi), shows that different characteristic functions arise from different distributions. As well as this, the straightforward derivation of the moments by (2.57) makes any determination of the characteristic function directly relevant to measurable quantities.

2.7 Cumulant Generating Function: Correlation Functions and Cumulants

A further important property of the characteristic function arises by considering its logarithm

$$\Phi(\mathbf{s}) = \log \phi(\mathbf{s}) \tag{2.64}$$

which is called the *cumulant generating function*. Let us assume that all moments exist so that $\phi(\mathbf{s})$ and hence, $\Phi(\mathbf{s})$, is expandable in a power series which can be written as

$$\Phi(\mathbf{s}) = \sum_{r=1}^{\infty} i^r \sum_{(m)} \langle\langle X_1^{m_1} X_2^{m_2} ..... X_n^{m_n} \rangle\rangle \frac{s_1^{m_1} s_2^{m_2} ..... s_n^{m_n}}{m_1! m_2! ..... m_n!} \delta\left(r, \sum_{i=1}^{n} m_i\right) \tag{2.65}$$

where the quantities $\langle\langle X_1^{m_1} X_2^{m_2} ..... X_n^{m_n} \rangle\rangle$ are called the *cumulants* of the variable $\mathbf{X}$. The notation chosen should not be taken to mean that the cumulants are functions of the particular product of powers of the $\mathbf{X}$; it rather indicates the moment of highest order which occurs in their expression in terms of moments. Stratonovich also uses the term *correlation functions*, a term which we will reserve for cumulants which involve more than one X_i. For, if the $\mathbf{X}$ are all independent, the factorization property (2.61) implies that $\Phi(\mathbf{s})$ (the cumulant generating function) is a sum of n terms, each of which is a function of only one s_i and hence the coefficient of mixed terms, i.e., the *corraltion functions* (in our terminology) are all zero and the converse is true. Thus, the magnitude of the

correlation functions is a measure of the degree of correlation.

The cumulants and correlation functions can be evaluated in terms of moments by expanding the characteristic function as a power series:

$$\phi(\mathbf{s}) = \sum_{r=1}^{\infty} \frac{i^r}{r!} \sum_{(m)} \langle X_1^{m_1} X_2^{m_2} ..... X_n^{m_n} \rangle \frac{r!}{m_1! m_2! ..... m_n!} \delta \left(r, \sum_{i=1}^{n} m_i \right) s_1^{m_1} s_2^{m_2} ..... s_n^{m_n} \tag{2.66}$$

Expanding the logarithm in a power series, and comparing it with (2.65) for $\Phi(\mathbf{s})$, the relationship between the cumulants and the moments can be deduced. No *simple* formula can be given, but the first few cumulants can be exhibited: we find

$$\langle\langle X_i \rangle\rangle = \langle X_i \rangle \tag{2.67}$$

$$\langle\langle X_i X_j \rangle\rangle = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \tag{2.68}$$

$$\begin{aligned} \langle\langle X_i X_j X_k \rangle\rangle = \langle X_i X_j X_k \rangle - \langle X_i X_j \rangle \langle X_k \rangle - \langle X_i \rangle \langle X_J X_k \rangle \\ - \langle X_i X_k \rangle \langle X_j \rangle + 2 \langle X_i \rangle \langle X_j \rangle \langle X_k \rangle \end{aligned} \tag{2.69}$$

Here, all formulae are also valid for any number of equal i, j, k, l. An expoicit general formula can be given as follows. Suppose we wish to calculate the cumulant $\langle\langle X_1 X_2 X_3 ..... X_n \rangle\rangle$. The procedure is the following:

(i) Write a sequence of n dots .....

(ii) Divide into $p+1$ subsets by inserting angle brackets

$$\langle ... \rangle \langle .. \rangle \langle ..... \rangle .. \langle .. \rangle \tag{2.70}$$

(iii) Distribute the symbols $X_1 .... X_n$ in places of the dots in such a way that all *different* expressions of this kind occur, e.g.,

$$\langle X_1 \rangle \langle X_2 X_3 \rangle = \langle X_1 \rangle \langle X_3 X_2 \rangle \neq \langle X_3 \rangle \langle X_1 X_2 \rangle \tag{2.71}$$

(iv) Take the sum of all such terms for a given p. Call this $C_p(X_1, X_2, ..., X_n)$

(v)

$$\langle\langle X_1 X_2 X_3 ..... X_n \rangle\rangle = \sum_{p=0}^{n-1} (-1)^p p! C_p(X_1, X_2, ..., X_n) \tag{2.72}$$

A derivation of this formula was given by Merton. The particular procedure is due to van Kampen.

(vi) *Cumulants in which there is one or more repeated element*: For example $\langle\langle X_1^2 X_2 X_3 \rangle\rangle$ - simply evaluate $\langle\langle X_1 X_2 X_3 X_4 \rangle\rangle$ and set $X_1 = X_4$ in the resulting expression.

2.7.1 Example: Cumulant of Order 4: $\langle\langle X_1X_2X_3X_4\rangle\rangle$

(a) $p = 0$

Only term is $\langle X_1X_2X_3X_4\rangle = C_0(X_1, X_2, X_3, X_4)$

(b) $p = 1$

Partition $\langle.\rangle\langle...\rangle$
Term $\{\langle X_1\rangle\langle X_2X_3X_4\rangle + \langle X_2\rangle\langle X_3X_4X_1\rangle + \langle X_3\rangle\langle X_4X_1X_2\rangle$
$+ \langle X_4\rangle\langle X_1X_2X_3\rangle \equiv D_1$

Partition $\langle..\rangle\langle..\rangle$
Term $\langle X_1X_2\rangle\langle X_3X_4\rangle + \langle X_1X_3\rangle\langle X_2X_4\rangle + \langle X_1X_4\rangle\langle X_2X_3\rangle \equiv D_2$

Hence

$$D_1 + D_2 = C_1(X_1, X_2, X_3, X_4) \tag{2.73}$$

(c) $p = 2$

Partition $\langle.\rangle\langle.\rangle\langle..\rangle$
Term $\langle X_1\rangle\langle X_2\rangle\langle X_3X_4\rangle + \langle X_1\rangle\langle X_3\rangle\langle X_2X_4\rangle + \langle X_1\rangle\langle X_4\rangle\langle X_2X_3\rangle$
$+ \langle X_2\rangle\langle X_3\rangle\langle X_1X_4\rangle + \langle X_2\rangle\langle X_4\rangle\langle X_1X_3\rangle + \langle X_3\rangle\langle X_4\rangle\langle X_1X_2\rangle$
$= C_2(X_1, X_2, X_3, X_4)$

(d) $p = 3$

Partition $\langle.\rangle\langle.\rangle\langle.\rangle\langle.\rangle$
Term $\langle X_1\rangle\langle X_2\rangle\langle X_3\rangle\langle X_4\rangle = C_3(X_1, X_2, X_3, X_4)$

Hence,

$$\langle\langle X_1X_2X_3X_4\rangle\rangle = C_0 - C_1 + 2C_2 - 6C_3 \tag{2.74}$$

2.7.2 Significance of Cumulants

From (2.67), 2.68) we see that the first two cumulants are the means $\langle X_i\rangle$ and covariances $\langle X_i, X_j\rangle$. Higher-order cumulants contain information of decreasing significance, unlike higher-order moments. We cannot set all *moments* higher than a certain order equal to zero since $\langle X^{2n}\rangle \geq \langle X^n\rangle^2$ and thus, all moments contain information about lower moments.

For cumulants, however, we can consistently set

$$\begin{aligned}
&\langle\langle X\rangle\rangle = a \\
&\langle\langle X^2\rangle\rangle = \sigma^2 \\
&\langle\langle X^n\rangle\rangle = 0 \quad (n > 2)
\end{aligned}$$

and we can easily deduce by using the inversion formula for the characteristic function that

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) \tag{2.75}$$

that is, a Gaussian probability distribution. It does not, however, seem possible to give more than this intuitive justification. Indeed, the theorem of Marcinkiewicz shows that the cumulant generating function cannot be a polynomial of degree greater than 2, that is, either all but the first 2 cumulants vanish or there are an infinite number of nonvanishing cumulants. The greatest significance of cumulants lies in the definition of the correlation functions of different variables in terms of them; this leads further to important approximation methods.

2.8 Gaussian and Poissonian Probability Distributions

2.8.1 The Gaussian Distribution

By far the most important probability distribution is the Gaussian, or normal distribution. Here we collect together the most important facts about it.

If $\mathbf{X}$ is a vector of n Gaussian random variables, the corresponding multivariate probability density function can be written

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\sigma)}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T \sigma^{-1} (\mathbf{x}-\bar{\mathbf{x}})\right] \tag{2.76}$$

so that

$$\langle \mathbf{X} \rangle = \int d\mathbf{x}\, \mathbf{x} p(\mathbf{x}) = \bar{\mathbf{x}} \tag{2.77}$$

$$\langle \mathbf{X}\mathbf{X}^T \rangle = \int d\mathbf{x}\, \mathbf{x}\mathbf{x}^T p(\mathbf{x}) = \bar{\mathbf{x}}\bar{\mathbf{x}}^T + \sigma \tag{2.78}$$

and the characteristic function is given by

$$\phi(s) = \langle \exp\left(i\mathbf{s}^T\mathbf{X}\right) \rangle = \exp\left(i\mathbf{s}^T\bar{\mathbf{x}} - \tfrac{1}{2}\mathbf{s}^T\sigma\mathbf{s}\right) \tag{2.79}$$

This particularly simple characteristic function implies that all cumulants of higher order than 2 vanish, and hence means that all moments of order higher than 2 are expressible in terms of those of order 1 and 2. The relationship (2.78) means that σ is the covariance matrix (as defined in Section 2.5.1), i.e., the matrix whose elements are the second-order correlation functions. Of course, σ is symmetric.

The precise relationship between the higher moments and the covariance matrix σ can be written down straightforwardly by using the relationship between the

moments and the characteristic function [Section 2.6 (iv)]. The formula is only simple if $\bar{\mathbf{x}} = 0$, in which case the odd moments vanish and the even moments satisfy

$$\langle X_i X_j X_k ..... \rangle = \frac{(2N)!}{N! 2^N} \{\sigma_{ij}\sigma_{kl}\sigma_{mn}....\}_{sym} \tag{2.80}$$

where the subscript "sum" means the symmetrized form of the product of σ's, and $2N$ is the order of the moment. For example,

$$\begin{aligned} \langle X_1 X_2 X_3 X_4 \rangle &= \frac{4!}{4 \cdot 2!} \left\{ \frac{1}{3} [\sigma_{12}\sigma_{34} + \sigma_{41}\sigma_{23} + \sigma_{13}\sigma_{24} \right\} \\ &= \sigma_{12}\sigma_{34} + \sigma_{41}\sigma_{23} + \sigma_{13}\sigma_{24} \end{aligned} \tag{2.81}$$

$$\langle X_1^4 \rangle = \frac{4!}{4 \cdot 2!} \left\{ \sigma_{11}^2 \right\} = 3\sigma_{11}^2 \tag{2.82}$$

2.8.2 Central Limit Theorem

The Gaussian distribution is important for a variety of reasons. Many variables are, in practice, empirically well approximated by Gaussians and the reason for this arises from the *central limit theorem*, which, roughly speaking, asserts that a random variable composed of the sum of many parts, each independent but arbitrarily distributed, is Gaussian. More precisely, let $X_1, X_2, X_3, ..., X_n$ be independent random variables such that

$$\langle X_i \rangle = 0 \quad , \quad var[X_i] = b_i^2 \tag{2.83}$$

and let the distribution function of X_i be $p_i(x_i)$.

Define

$$S_n = \sum_{i=1}^{n} X_i \tag{2.84}$$

and

$$\sigma_n^2 = var[S_n] = \sum_{i=1}^{n} b_i^2 \tag{2.85}$$

We require further the fulfilment of the *Lindeberg condition*:

$$\lim_{n\to\infty} \left[\frac{1}{\sigma_n^2} \sum_{i=1}^{n} \int\limits_{|x|>t\sigma_n} dx\, x^2 p_i(x) \right] = 0 \tag{2.86}$$

for any fixed $t > 0$. Then, under these conditions, the distribution of the normalized sums S_n/σ_n tends to the Gaussian with zero mean and unit variance.

The proof of the theorem can be found in many texts. It is worthwhile commenting on the hypotheses, however. We first note that the summands X_i

are required to be independent. This condition is not absolutely necessary; for example, choose

$$X_i = \sum_{r=i}^{i+j} Y_r \tag{2.87}$$

where the Y_j are independent. Since the sum of the X's can be rewritten as a sum of Y's (with certain finite coefficients), the theorem is still true.

Roughly speaking, as long as the correlation between X_i and X_j goes to zero sufficiently rapidly as $|i-j| \to \infty$, a central limit theorem will be expected. The Lindeberg condition (2.86) is not an obviously understandable condition but it is the weakest condition which expresses the requirement that the probability for $|X_i|$ to be large is very small. For example, if all the b_i are infinite or greater than some constant C, it is clear that σ_n^2 diverges as $n \to \infty$. The sum of integrals in (2.86(is the sum of contributions to variances for all $|X_i| > t\sigma_n$, and it is clear that as $n \to \infty$, each contribution goes to zero. The Lindeberg condition requires the sum of all the contributions not diverge as fast as σ_n^2. In practice, it is a rather weak requirement; satisfied if $|X_i| < C$ for all X_i, or if $p_i(x)$ goes to zero sufficiently rapidly as $x \to \pm\infty$. An exception is

$$p_i(x) = \frac{a_i}{\pi(x^2 + a_i^2)} \tag{2.88}$$

the *Cauchy*, or *Lorentzian* distribution. The variance of this distribution is infinite and, in fact, the sum of all the X_i has a distribution of the same form as (2.88) with a_i replaced by $\sum_{i=1}^n a_i$. Obviously, the Lindeberg condition is not satisfied.

2.8.3 The Poisson Distribution

A distribution which plays a central role in the study of random variables which take on positive integer values is the Poisson distribution. If X is the relevant variable the Poisson distribution is defined by

$$P(X = x) \equiv P(x) = \frac{e^{-\alpha}\alpha^x}{x!} \tag{2.89}$$

and clearly, the *factorial moments*, defined by

$$\langle X^r \rangle_f = \langle x(x-1).....(x-r+1) \rangle \tag{2.90}$$

are given by

$$\langle X^r \rangle_f = \alpha^r \tag{2.91}$$

For variables whose range bis nonnegative integral, we can very naturally define the *generating function*

$$G(s) = \sum_{x=0}^{\infty} s^x P(x) = \langle s^x \rangle \tag{2.92}$$

which is related to the characteristic function by

$$G(s) = \phi(-i \log s) \tag{2.93}$$

The generating function has the useful property that

$$\langle X^r \rangle_f = \left[\left(\frac{\partial}{\partial s} \right)^r G(s) \right]_{s=1} \tag{2.94}$$

For the Poisson distribution we have

$$G(s) = \sum_{x=0}^{\infty} \frac{e^{-\alpha}(s\alpha)^x}{x!} = e^{\alpha(s-1)} \tag{2.95}$$

We may also define the factorial cumulant generating function $g(s)$ by

$$g(s) = \log G(s) \tag{2.96}$$

and the *factorial cumulants* $\langle\langle X^r \rangle\rangle_f$ by

$$g(s) = \sum_{x=1}^{\infty} \langle\langle X^r \rangle\rangle_f \frac{(s-1)^r}{r!} \tag{2.97}$$

We see that the Poisson distribution has all but the first factorial cumulant zero.

The Poisson distribution arises naturally in very many contexts, for example, we have already met it is Section 1.5.1 as the solution of a simple master equation. It plays a similar central role in the study of random variables which take on integer values to that occupied by the Gaussian distribution in the study of variables with a continuous range. However, the only simple multivariate generalization of the Poisson is simply a product of Poissons, i.e., of the form

$$P(x_1, x_2, x_3, ...) = \prod_{i=1}^{n} \frac{e^{-\alpha_i}(\alpha_i)_i^x}{x_i!} \tag{2.98}$$

There is no logical concept of a correlated multipoisson distribution, similar to that of a correlated multivariate Gaussian distribution.

2.9 Limits of Sequences of random Variables

Much of computational work consists of determining *approximations* to random variables, in which the concept of a *limit of a sequence of random variables* naturally arises. However, there is no unique way of defining such a limit.

For suppose we have a probability space Ω, and a sequence of random variables X_n defined on Ω. Then by the limit of the sequence as $n \to \infty$

$$X = \lim_{n \to \infty} X_n \tag{2.99}$$

we mean a random variable X which, in some sense, is approached by the sequence of random variables X_n. The various possibilities arise when one considers that the probability space Ω has elements ω which have a probability density $p(\omega)$. Then we can choose the following definitions.

2.9.1 Almost Certain Limit

X_n converges *almost certainly* to X if, for all ω except a set of probability zero

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \tag{2.100}$$

Thus each realization of X_n converges on X and we write

$$ac - \lim_{n\to\infty} X_n = X \tag{2.101}$$

2.9.2 Mean Square Limit (Limit in the Mean)

Another possibility is to regard the $X_n(\omega)$ as functions of ω, and look for the mean square deviation of $X_n(\omega)$ from $X(\omega)$. Thus, we say that X_n converges to X in the *mean square* if

$$\lim_{n\to\infty} \int d\omega\, p(\omega)[X_n(\omega) - X(\omega)]^2 \equiv \lim_{n\to\infty} \langle (X_n - X)^2 \rangle = 0 \tag{2.102}$$

This is the kind of limit which is well known in Hilbert space theory. We write

$$ms - \lim_{n\to\infty} X_n = X \tag{2.103}$$

2.9.3 Stochastic Limit, or Limit in Probability

We can consider the possibility that $X_n(\omega)$ approaches X because the probability of deviation from X approaches zero precisely, this means that if for any $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \tag{2.104}$$

then the *stochastic limit* go X_n is X.

In this case we write

$$st - \lim_{n\to\infty} X_n = X \tag{2.105}$$

2.9.4 Limit in Distribution

An even weaker form of convergence occurs if, for any continuous bounded function $f(x)$

$$\lim_{n\to infty} \langle f(X_n) \rangle = \langle f(X) \rangle \tag{2.106}$$

In this case the convergence of the limit is said to be *in distribution*. In particular, using $\exp ixs$) for $f(x)$, we find that the characteristic functions approach each other, and hence the probability density of X_n approach that of X.

2.9.5 Relationships Between Limits

The following relations can be show.

$$\text{Almost certain convergence} \Rightarrow \text{stochastic convergence}$$
$$\text{Convergence in mean square} \Rightarrow \text{stochastic convergence}$$
$$\text{Stochastic convergence} \Rightarrow \text{convergence in distribution}$$

All of these limits have uses in applications.

Chapter 3

Markov Processes

3.1 Stochastic Processes

All the examples given in Chapter 1 can be mathematically described as *stochastic processes* by which we mean, in a loose sense, systems which evolve probabilistically in time or more precisely, systems in which a certain time-dependent random variable $\mathbf{X}(t)$ exists. We can measure values $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ...$ etc., of $\mathbf{X}(t)$ at times $t_1, t_2, t_3, ...$ and we assume that a set of joint probability densities

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; ....) \tag{3.1}$$

exists, which completely describe the system.

In terms of these joint probability density functions, one can also define conditional probability densities:

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ....|\mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; ....) \equiv \frac{p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ....; \mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; ....)}{p(\mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; ....)} \tag{3.2}$$

These definitions are valid independently of the ordering of the times, although it is usual to consider only times which increase from right to left, i.e.,

$$t_1 \geq t_2 \geq t_3 \geq ..... \geq \tau_1 \geq \tau_2 \geq ..... \tag{3.3}$$

The concept of an evolution equation leads us to consider the conditional probabilities as predictions of the future values of $\mathbf{X}(t)$ (i.e., $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ...$ at times $t_1, t_2, t_3, ...$) given the knowledge of the past (values $\mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; ....$ at times $\tau_1, \tau_2, \tau_3, ...$).

3.1.1 Kinds of Stochastic Process

The concept of a general stochastic process is very loose. To define the process we need to know at least all possible joint probabilities of the kind (3.1). If such

knowledge does define the process, it is known as a *separable stochastic process*. All the processes considered in these notes will be assumed to be separable.

(a) **Complete Independence**: This is the most simple kind of stochastic process; it satisfies the property

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; ....) = \prod_i p(\mathbf{x}_i, t_i) \tag{3.4}$$

which means that the value of $\mathbf{X}$ at time t is completely independent of its values in the past (or future).

(b) **Bernoulli Trials**: An even more special case occurs when the $p(\mathbf{x}_i, t_i)$ are independent of t_i, so that the same probability law governs the process at all times. We then have the *Bernoulli trials*, in which a probabilistic process is repeated at successive times.

(c) **Martingales**: The conditional mean value of $X(t)$ given that $X(t_0) = x_0$ is defined as

$$\langle X(t)|[x_0, t_0]\rangle \equiv \int dx\, xp(x, t|x_0, t_0) \tag{3.5}$$

In a Martingale this has the simple property

$$\langle X(t)|[x_0, t_0]\rangle = x_0 \tag{3.6}$$

The martingale property is a rather strong property, and is associated with many similar and related processes, such as local martingales, sub-martingales, super-martingales, etc., which have come to be extensively studied and used in the past 25 years.

(d) **Markov Processes**: The next most simple idea is that of the *Markpv process* in which knowledge of only the present determines the future, and most of these notes are built around this concept.

3.2 Markov Process

The *Markov assumption* is formulated in terms of conditional probabilities. We require that if the times satisfy the ordering (3.3), the conditional probability is determined entirely by the knowledge of the most recent condition, i.e.,

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ....|\mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; ....) = p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ....|\mathbf{y}_1, \tau_1) \tag{3.7}$$

This is simply a more precise statement of the assumptions made by Einstein, Smoluchowski and others. It is, even by itself, extremely powerful. For it means that we can define everything in terms of simple conditional probabilities $p(\mathbf{x}_1, t_1|\mathbf{y}_1, \tau_1)$. For example, , by definition of the conditional probability density $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2|\mathbf{y}_1, \tau_1) = p(\mathbf{x}_1, t_1|\mathbf{x}_2, t_2; \mathbf{y}_1, \tau_1)p(\mathbf{x}_2, t_2|\mathbf{y}_1, \tau_1)$ and using the Markov assumption (3.7, we find)

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2|\mathbf{y}_1, \tau_1) = p(\mathbf{x}_1, t_1|\mathbf{x}_2, t_2)p(\mathbf{x}_2, t_2|\mathbf{y}_1, \tau_1) \tag{3.8}$$

and it is not difficult to see that an arbitrary joint probability can be expressed simply as

$$\begin{aligned} &p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; ....; \mathbf{x}_n, t_n) \\ &\quad = p(\mathbf{x}_1, t_1|\mathbf{x}_2, t_2)p(\mathbf{x}_2, t_2|\mathbf{x}_3, t_3)p(\mathbf{x}_3, t_3|\mathbf{x}_4, t_4)... \\ &\quad ...p(\mathbf{x}_{n-1}, t_{n-1}|\mathbf{x}_n, t_n) \end{aligned} \tag{3.9}$$

provided

$$t_1 \geq t_2 \geq t_3 \geq ..... \geq t_{n-1} \geq t_n \tag{3.10}$$

3.2.1 Consistency - the Chapman-Kolmogorov Equation

From Section 2.3.3 we require that summing over all mutually exclusive events of one kind in a joint probability eliminates that variable, i.e.,

$$\sum_B P(A \cap B \cap C...) = P(A \cap C...) \tag{3.11}$$

and when this is applied to stochastic processes, we get two deceptively similar equations:

$$p(\mathbf{x}_1, t_1) = \int d\mathbf{x}_2 \, p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2) = \int d\mathbf{x}_2 \, p(\mathbf{x}_1, t_1|\mathbf{x}_2, t_2)p(\mathbf{x}_2, t_2) \tag{3.12}$$

This equation is an identity valid for all stochastic processes and is the first in a hierarchy of equations, the second of which is

$$\begin{aligned} p(\mathbf{x}_1, t_1|\mathbf{x}_3, t_3) &= \int d\mathbf{x}_2 \, p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2|\mathbf{x}_3, t_3) \\ &= \int d\mathbf{x}_2 \, p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2|\mathbf{x}_3, t_3)p(\mathbf{x}_2, t_2|\mathbf{x}_3, t_3) \end{aligned} \tag{3.13}$$

This equation is always valid. We now introduce the Markov assumption. If $t_1 \geq t_2 \geq t_3$, we can drop the t_3 dependence in the doubly conditional probability and write

$$p(\mathbf{x}_1, t_1|\mathbf{x}_3, t_3) = \int d\mathbf{x}_2 \, p(\mathbf{x}_1, t_1|\mathbf{x}_2, t_2)p(\mathbf{x}_2, t_2|\mathbf{x}_3, t_3) \tag{3.14}$$

which is the *Chapman-Kolmogorov equation*.

What is the essential difference between (3.14) and (3.12)? The obvious answer is that (3.12) is for unconditioned probabilities, whereas (3.14) is for conditioned probabilities. Equation (3.14) is a rather complex nonlinear functional equation relating all conditional probabilities $p(\mathbf{x}_i, t_i|\mathbf{x}_j, t_j)$ to each other, whereas (3.12) simply constructs the one time probabilities in the future t_1 of t_2, given the conditional probability $p(\mathbf{x}_1, t_1|\mathbf{x}_2, t_2)$.

The Chapman-Kolmogorov equation has many solutions. These are best understood by deriving the differential form which is done later in Section 3.4.1 under certain rather mild conditions.

3.2.2 Discrete State Spaces

In the case where we have a discrete variable, we will use the symbol $N = (N_1, N_2 < N_3, ...)$, where The N_i are random variables which take on integral values. Clearly, now we can replace

$$\int d\mathbf{x} \longleftrightarrow \sum_n \tag{3.15}$$

and we can now write the Chapman-Kolmogorov equation for such a process as

$$P(\mathbf{n}_1, t_1|\mathbf{n}_3, t_3) = \sum_{n_2} P(\mathbf{n}_1, t_1|\mathbf{n}_2, t_2)P(\mathbf{n}_2, t_2|\mathbf{n}_3, t_3) \tag{3.16}$$

This is now matrix multiplication, with possibly infinite matrices.

3.2.3 More general Measures

A more general formulation would assume a measure $d\mu(\mathbf{x})$ instead of $d\mathbf{x}$ where a variety of choices can be made. For example, if $\mu(\mathbf{x})$ is a step function with steps at integral values of $\mathbf{x}$, we recover the discrete space form. Most mathematical works attempt to be as general as possible. For applications, such generality can lead to lack of clarity so, where possible, we will favor a more specific notation.

3.3 Continuity in Stochastic Processes

Whether or not the random variable $\mathbf{X}(t)$ has a continuous range of possible values is a completely different question from whether the sample path of $\mathbf{X}(t)$ is a continuous function of t. For example, in a gas composed of molecules with velocities $\mathbf{V}(t)$, it is clear that all possible values of $\mathbf{V}(t)$ are in principle realizable, so that the *range* of $\mathbf{V}(t)$ is continuous. However, a model of collisions in a gas of hard spheres as occurring instantaneously is often considered, and in such a model the velocity before the collision $\mathbf{v}_i$, will change instantaneously at the time of impact to another value $\mathbf{v}_f$, so the sample path of $\mathbf{V}(t)$ is not continuous. Nevertheless, in such a model, the *position* of a gas molecule $\mathbf{X}(t)$ would be expected to change continuously.

A major question now arises. Do *Markov* processes with *continuous sample paths* actually exist in reality? Notice the combination of *Markov* and *continuous*. It is almost certainly the case that in a classical picture (i.e., not quantum mechanical), all variables with a continuous range have continuous sample paths.Even the hard sphere gas mentioned above is an idealization and more realistically, one should allow some potential to act which would continuously deflect the molecules during a collision. But it would also be the case that, if we observe on such a fine time scale, the process will probably not be Markovian. The immediate history of the whole system will almost certainly be required to predict even the probabilistic future. This is certainly born out out in all

attempts to derive Markovian probabilistic equations from mechanics. Equations which are derived are rarely truly Markovian - rather there is a curtain characteristic memory time during which the previous history is important.

This means that in the real world there is really no such thing as a Markov process; rather, there may be systems whose memory time is so small that, on the time scale on which we carry out observations, it is fair to regard them a being well approximated by a Markov process. But in this case, the question of whether the sample paths are continuous is not relevant. The sample paths of the approximating Markov process certainly need not be continuous. Even if collisions of molecules are not accurately modeled by hard spheres, during the time taken for a collision, a finite change of velocity takes place and this will appear in the approximating Markov process as a discrete step. On this time scale, even the position may change discontinuously, thus giving the picture of Brownian motion as modeled by Einstein.

In chemical reactions, for example, the time taken for an individual reaction to proceed to completion - roughly of the same order of magnitude as the collision time for molecules - provides yet another minimum time, since during this time, states which cannot be described in terms of individual molecules exist. Here, therefore, the very description of the state in terms of individual molecules requires a certain minimum time scale to be considered.

However, Markov processes with continuous sample paths do exist *mathematically* and are useful in describing reality. The model of the gas mentioned above provides a useful example. The position of the molecule is indeed probably best modeled as changing discontinuously by discrete jumps. Compared to the distance travelled, however, these jumps are infinitesimal and a continuous curve provides a good approximation to the sample path. On the other hand, the velocities can change by amounts which are the same order of magnitude as typical values attained in practice. The average velocity of a molecule in a gas is about $1000\,m/s$ and during a collision can easily reverse sign. The velocities simply cannot reach (with any significant probability) values for which the change of velocity can be regarded as very small. Hence, there is no sense in a continuous path description of velocities of a gas.

3.3.1 Mathematical Definition of a Continuous Markov Process

For a *Markov process*, it can be shown that with probability one, the sample paths are continuous functions of t, if for any $\epsilon > 0$ we have

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|x-z|>\epsilon} d\mathbf{x}\, p(\mathbf{x}, t+\Delta t|\mathbf{z}, t) = 0 \tag{3.17}$$

uniformly in $\mathbf{z}, t$ and δt.

This means that the probability for the final position $\mathbf{x}$ to be finitely different from $\mathbf{z}$ goes to zero *faster* than Δt, as δt goes to zero. Equation (3.17) is sometime called the *Lindeberg condition*. Note this different than the earlier one.

Examples

(i) *Einstein's solution* for his $f(x,t)$ (Section 1.2.1) is really the conditional probability $p(x,t|0,0)$. Following his method we would find

$$p(x,t+\Delta t|z,t) = \frac{1}{\sqrt{4\pi D\Delta t}} e^{-\frac{(x-z)^2}{4D\Delta t}} \tag{3.18}$$

and it is easy to check that (3.17) is satisfied in this case. Thus, Brownian motion in Einstein's formulation has continuous sample paths.

(ii) *Cauchy process*: Suppose

$$p(x,t+\Delta t|z,t) = \frac{\Delta t}{\pi[(x-z)^2+\Delta t^2]} \tag{3.19}$$

Then this does satisfy (3.17) so the sample paths are discontinuous.

However, in both cases, we have as required for consistency

$$\lim_{\Delta t\to 0} p(x,t+\Delta t|z,t) = \delta(x-z) \tag{3.20}$$

and it is easy to show that in both cases, the Chapman-Kolmogorov equation is satisfied.

The difference between the two processes just described is illustrated in the figure below

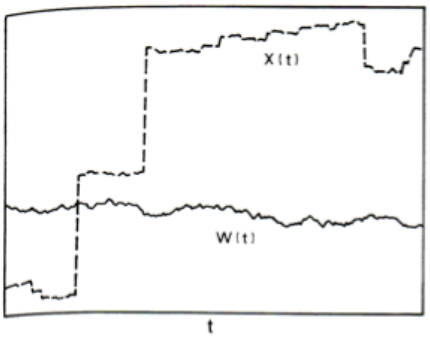

Figure 3.1: Illustration of sample paths of the Cauchy process $X(t)$(t) (dashed) and Brownian motion $W(t)$ (solid).

in which simulations of both processes are given. The difference between the two is striking. Notice, however, that even the Brownian motion curve is extremely irregular, even though continuous - in fact it is nowhere differentiable. The Cauchy process curve, however, is only piecewise continuous.

3.4 Differential Chapman-Kolmogorov Equation

Under appropriate assumptions, the Chapman-Kolmogorov equation can be reduced to a differential equation. The assumptions made are closely connected with the continuity properties of the process under consideration. Because of the form of the continuity condition (3.17), one is led to consider a method of dividing the differentiability conditions into two parts, one corresponding to continuous motion of a representative point and the other to discontinuous motion.

We require the following conditions *for all* $\epsilon > 0$:

(i)

$$\lim_{\Delta t \to 0} \frac{p(x, t + \Delta t | \mathbf{z}, t)}{\Delta t} = W(\mathbf{x} | \mathbf{z}, t) \tag{3.21}$$

uniformly in $\mathbf{x}, \mathbf{z}$, and t for $|\mathbf{x} - \mathbf{z}| \geq \epsilon$.

(ii)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|x-z|>\epsilon} d\mathbf{x}\,(x_i - z_i) p(x, t + \Delta t | \mathbf{z}, t) = A_i(\mathbf{z}, t) + O(\epsilon) \tag{3.22}$$

(iii)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|x-z|>\epsilon} d\mathbf{x}\,(x_i - z_i)(x_j - z_j) p(x, t + \Delta t | \mathbf{z}, t) = B_{ij}(\mathbf{z}, t) + O(\epsilon) \tag{3.23}$$

the last two being uniform in $\mathbf{x}, \mathbf{z}$, and t.

Notice that all higher-order coefficients of the form (3.22,3.23) must vanish. For example, consider the third-order quantity defined by

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|x-z|>\epsilon} d\mathbf{x}\,(x_i - z_i)(x_j - z_j)(x_k - z_k) p(x, t + \Delta t | \mathbf{z}, t) = C_{ijk}(\mathbf{z}, t) + O(\epsilon) \tag{3.24}$$

Since C_{ijk} is symmetric in i, j, k consider

$$\sum_{i,j,k} \alpha_i \alpha_j \alpha_k C_{ijk}(\mathbf{z}, t) \equiv \bar{C}(\alpha, \mathbf{z}, t) \tag{3.25}$$

so that

$$C_{ijk}(\mathbf{z},t) = \frac{1}{3!}\frac{\partial^3}{\partial\alpha_i\partial\alpha_j\partial\alpha_k}\bar{C}(\alpha,\mathbf{z},t) \tag{3.26}$$

Then,

$$\begin{aligned}
|\bar{C}(\alpha,\mathbf{z},t)| &\leq \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int\limits_{|x-z|>\epsilon} |\alpha\cdot((x)-\mathbf{z})|[\alpha\cdot((x)-\mathbf{z})]^2 p(\mathbf{x},t+\Delta t|\mathbf{z},t)d\mathbf{x} + O(\epsilon) \\
&\leq |\alpha|\epsilon\lim_{\Delta t\to 0}\int[\alpha\cdot((x)-\mathbf{z})]^2 p(\mathbf{x},t+\Delta t|\mathbf{z},t)d\mathbf{x} + O(\epsilon) \\
&= \epsilon|\alpha|[\alpha_i\alpha_j B_{ij}(z,t) + O(\epsilon)] + O(\epsilon) \\
&= O(\epsilon)
\end{aligned} \tag{3.27}$$

so that C is zero. Similarly, we can show that all corresponding higher-order qualities also vanish.

According to the condition for continuity (3.17), the process can only have continuous paths of $W(\mathbf{x}|\mathbf{z},t)$ vanishes for all $\mathbf{x}\neq\mathbf{z}$. Thus, this function must in some way describe discontinuous motion, while the quantities A_i and B_{ij} must be connected with continuous motion.

3.4.1 Derivation of the Differential Chapman-Kolmogorov Equation

We consider the time evolution of the expectation of a function $f(\mathbf{z})$ which is twice continuously differentiable.

Thus,

$$\begin{aligned}
&\partial_t\int d\mathbf{x}\, f(\mathbf{x})p(\mathbf{x},t|\mathbf{y},t') \\
&= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\int d\mathbf{x}\, f(\mathbf{x})[p(\mathbf{x},t+\Delta t|\mathbf{y},t') - p(\mathbf{x},t|\mathbf{y},t')]\right\} &(3.28)\\
&= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\int d\mathbf{x}\int d\mathbf{z}\, f(\mathbf{x})p(\mathbf{x},t+\Delta t|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') - \int d\mathbf{z}\, f(\mathbf{z})p(\mathbf{z},t|\mathbf{y},t')\right\} &(3.29)
\end{aligned}$$

where we have used the Chapman-Kolmogorov equation in the positive term of (3.28) to produce the corresponding term in (3.29)

We now divide the integral over $\mathbf{x}$ into two regions $|\mathbf{x}-\mathbf{z}|\geq\epsilon$ and $|\mathbf{x}-\mathbf{z}|<\epsilon$. When $|\mathbf{x}-\mathbf{z}|<\epsilon$, since $f(\mathbf{z})$ is, by assumption, twice continuously differentiable, we may write

$$f(\mathbf{x}) = f(\mathbf{z}) + \sum_i\frac{\partial f(\mathbf{z})}{\partial z_i}(x_i-z_i) + \tfrac{1}{2}\sum_{i,j}\frac{\partial^2 f(\mathbf{z})}{\partial z_i\partial z_j}(x_i-z_i)(x_j-z_j) + |\mathbf{x}-\mathbf{z}|^2 R(\mathbf{x},\mathbf{z}) \tag{3.30}$$

where we have(again by the twice continuously differentiability)

$$|R(\mathbf{x},\mathbf{z})| \to 0 \quad \text{as} \quad |\mathbf{x}-\mathbf{z}| \to 0 \tag{3.31}$$

Now substitute in (3.29)

$$\begin{aligned}(3.29) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\Bigg\{ &\int\int_{|x-z|<\epsilon} d\mathbf{x}d\mathbf{z}\left[\sum_i (x_i-z_i)\frac{\partial f}{\partial z_i} + \tfrac{1}{2}\sum_{i,j}(x_i-z_i)(x_j-z_j)\frac{\partial^2 f}{\partial z_i \partial z_j}\right] \\ &\qquad \times p(\mathbf{x},t+\Delta t|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') \\ &+ \int\int_{|x-z|<\epsilon} d\mathbf{x}d\mathbf{z}|\mathbf{x}-\mathbf{z}|^2 p(\mathbf{x},t+\Delta t|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') \\ &+ \int\int_{|x-z|<\epsilon} d\mathbf{x}d\mathbf{z} f(\mathbf{x})p(\mathbf{x},t+\Delta t|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') \\ &+ \int\int_{|x-z|<\epsilon} d\mathbf{x}d\mathbf{z} f(\mathbf{z})p(\mathbf{x},t+\Delta t|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') \\ &- \int\int d\mathbf{x}d\mathbf{z} f(\mathbf{z})p(\mathbf{x},t+\Delta t|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')\Bigg\}\end{aligned} \tag{3.32}$$

Note that since $p(\mathbf{x},t+\Delta t|\mathbf{z},t)$ is a probability, the integral over $\mathbf{x}$ in the last term gives 1 - this is simply the last term in (3.29).

We now consider this expression line by line.

Lines 1 and 2: By the assumed uniform convergence, we take the limit inside the integral to obtain [using conditions (ii) and (iii) of Section 3.4]

$$\int d\mathbf{z}\left[\sum_i A_i(\mathbf{z})\frac{\partial f}{\partial z_i} + \tfrac{1}{2}\sum_{i,j} B_{ij}\frac{\partial^2 f}{\partial z_i \partial z_j}\right] p(\mathbf{z},t|\mathbf{y},t') + O(\epsilon) \tag{3.33}$$

Line 3: This is a remainder term and vanishes as $\epsilon \to 0$. For

$$\begin{aligned}&\left|\frac{1}{\Delta t}\int\limits_{|x-z|<\epsilon} |\mathbf{x}-\mathbf{z}|^2 R(\mathbf{x},\mathbf{z})p(\mathbf{x},t+\Delta t|\mathbf{z},t)\right| \\ &\quad \le \left[\frac{1}{\Delta t}\int\limits_{|x-z|<\epsilon} |\mathbf{x}-\mathbf{z}|^2 p(\mathbf{x},t+\Delta t|\mathbf{z},t)\right] \max_{|x-z|<\epsilon}|R(\mathbf{x},\mathbf{z})| \\ &\quad \to \left[\sum_{i,j} B_{ij}(\mathbf{z},t) + O(\epsilon)\right]\left\{\max_{|x-z|<\epsilon}|R(\mathbf{x},\mathbf{z})|\right\}\end{aligned} \tag{3.34}$$

From (3.31) we can see that as $\epsilon \to 0$, the factor in curly brackets vanishes.

Lines 4-6: We can put these all together to obtain

$$\int\int_{|x-z|<\epsilon} d\mathbf{x}d\mathbf{z}\, f(\mathbf{z})[W(\mathbf{z}|\mathbf{x},t)p(\mathbf{x},t|\mathbf{y},t') - W(\mathbf{x}|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')] \tag{3.35}$$

The whole right-hand side of (3.32) is independent of ϵ. Hence taking the limit $\epsilon \to 0$, we find

$$\partial_t \int d\mathbf{z}\, f(\mathbf{z})p(\mathbf{z},t|\mathbf{y},t') = \int d\mathbf{z} \left[\sum_i A_i(\mathbf{z})\frac{\partial f(\mathbf{z})}{\partial z_i} + \tfrac{1}{2}\sum_{i,j} B_{ij}(\mathbf{z})\frac{\partial^2 f(\mathbf{z})}{\partial z_i \partial z_j}\right] p(\mathbf{z},t|\mathbf{y},t')$$
$$+ \int d\mathbf{z}\, f(\mathbf{z}) \left\{ -\!\!\!\!\!\!\int d\mathbf{x}[W(\mathbf{z}|\mathbf{x},t)p(\mathbf{x},t|\mathbf{y},t') - W(\mathbf{x}|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')]\right\} \tag{3.36}$$

Notice that we use the definition

$$\lim_{\epsilon\to 0} \int_{|x-z|<\epsilon} d\mathbf{x}\, F(\mathbf{x},\mathbf{z}) \equiv -\!\!\!\!\!\!\int d\mathbf{x}\, F(\mathbf{x},\mathbf{z}) \tag{3.37}$$

for the principal value integral of a function $F(\mathbf{x},\mathbf{z})$. For (3.36) to have any meaning, this integral should exist. Equation (3.21) defines $W(\mathbf{x}|\mathbf{z},t)$ only for $\mathbf{x} \neq \mathbf{z}$ and hence leaves open the possibility that it is infinite at $\mathbf{x} = \mathbf{z}$, as is indeed the case for the Cauchy process, discussed in Section 3.3.1, for which

$$W(x|z,t) = \frac{1}{\pi(x-z)^2} \tag{3.38}$$

However, if $p(\mathbf{z},t|\mathbf{y},t')$ is continuous and once differentiable, then the principal value integral exists. In the remainder of these notes we will not write this integral explicitly as a principal value integral since one rarely considers the singular cases for which it is necessary.

The final step is now to integrate by parts. We find

$$\int d\mathbf{z}\, f(\mathbf{z})\partial_t p(\mathbf{z},t|\mathbf{y},t')$$
$$= \int d\mathbf{z}\, f(\mathbf{z}) \left\{ -\sum_i \frac{\partial}{\partial z_i} A_i(\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') + \tfrac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j} B_{ij}(\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t') \right.$$
$$+ \int d\mathbf{x}\ [W(\mathbf{z}|\mathbf{x},t)p(\mathbf{x},t|\mathbf{y},t') - W(\mathbf{x}|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')]\}$$
$$+ \text{ surface terms} \tag{3.39}$$

We have not specified the range of the integrals. Suppose the process is confined to a region R with surface S. Then clearly,

$$p(\mathbf{x},t|\mathbf{z},t') = 0 \quad \text{unless both } \mathbf{x} \text{ and } \mathbf{z} \ \in R \tag{3.40}$$

It is clear that by definition we have

$$W(\mathbf{x}|\mathbf{z},t) = 0 \quad \text{unless both } \mathbf{x} \text{ and } \mathbf{z} \ \in R \tag{3.41}$$

But the condition on $A_i(\mathbf{z},t)$ and $B_{ij}(\mathbf{z},t)$ can result in discontinuities in these functions as defined by (3.22) and (3.23) since $p(\mathbf{x}, t+\Delta t|\mathbf{z},t')$, the conditional probability, can very reasonably change discontinuous ly a $\mathbf{z}$ crosses the boundary of R, reflecting the fact that no transitions are allowed from outside R to inside R.

In integrating by parts, we are forced to differentiate both A_i and B_{ij} and by our reasoning above, one cannot assume that this is possible on the boundary of the region. Hence, let us choose $f(\mathbf{z})$ to be arbitrary but nonvanishing only in an arbitrary region R' entirely contained in R. We can then deduce that for all $\mathbf{z}$ in *the interior* or R,

$$\boxed{\begin{aligned}\frac{\partial p(\mathbf{z},t|\mathbf{y},t')}{\partial t} &= -\sum_i \frac{\partial}{\partial z_i} A_i(\mathbf{z},t) p(\mathbf{z},t|\mathbf{y},t') \\ &+ \tfrac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j} B_{ij}(\mathbf{z},t) p(\mathbf{z},t|\mathbf{y},t') \\ &+ \int d\mathbf{x}\ [W(\mathbf{z}|\mathbf{x},t)p(\mathbf{x},t|\mathbf{y},t') - W(\mathbf{x}|\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')]\end{aligned}} \tag{3.42}$$

Surface terms do not arise, since they necessarily vanish.

This equation is the *differential Chapman-Kolmogorov* equation.

3.4.2 Status of the Differential Chapman-Kolmogorov Equation

From our serivation it is not clear to what extent solutions of the differential Chapman-Kolmogorov equation are solutions of the Chapman-Kolmogorov equation itself or indeed, to what extent solutions exist. It is certainly true, however, that a set of conditional probabilities which obey the Chapman-Kolmogorov equation does generate a Markov process, in the sense that the joint probabilities so generated satisfy all probability axioms.

It can be shown that, under certain conditions, if we specify $\mathbf{A}(\mathbf{x},t), \mathbf{B}(\mathbf{x},t)$ (which must be positive semi-definite), and $W(\mathbf{x}|\mathbf{y},t)$ (which must be non-negative), that a non-negative solution to the differential Chapman-Kolmogorov equation exists, and this solution also satisfies the Chapman-Kolmogorov equation. The conditions to be satisfied are the *initial condition*,

$$p(\mathbf{z},t|\mathbf{y},t) = \delta(\mathbf{y}-\mathbf{z}) \tag{3.43}$$

which follows from the definition of the conditional probability density, and any appropriate boundary conditions.

3.5 Interpretation of Conditions and Results

Each of the conditions (i), (ii), (iii) of Section 3.4 can now be seen to give rise to a distinctive part of the equation, whose interpretation is rather straightforward. We can identify three processes taking place, which are known as jumps, drift and diffusion.

3.5.1 Jump Processes: The Master Equation

We consider a case in which

$$A_i(\mathbf{z}, t) = B_{ij}(\mathbf{z}, t) = 0 \tag{3.44}$$

so that we now have the *master equation*:

$$\frac{\partial p(\mathbf{z}, t|\mathbf{y}, t')}{\partial t} = \int d\mathbf{x}\ [W(\mathbf{z}|\mathbf{x}, t)p(\mathbf{x}, t|\mathbf{y}, t') - W(\mathbf{x}|\mathbf{z}, t)p(\mathbf{z}, t|\mathbf{y}, t')] \tag{3.45}$$

To first order in Δt we solve approximately, as follows. Notice that

$$p(\mathbf{z}, t|\mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z}) \tag{3.46}$$

Hence,

$$\frac{p(\mathbf{z}, t + \Delta t|\mathbf{y}, t) - p(\mathbf{z}, t|\mathbf{y}, t)}{\Delta t} = \int d\mathbf{x}\ [W(\mathbf{z}|\mathbf{x}, t)p(\mathbf{x}, t|\mathbf{y}, t) - W(\mathbf{x}|\mathbf{z}, t)p(\mathbf{z}, t|\mathbf{y}, t)]$$

$$p(\mathbf{z}, t+\Delta t|\mathbf{y}, t) = p(\mathbf{z}, t|\mathbf{y}, t)+\Delta t \int d\mathbf{x}\ [W(\mathbf{z}|\mathbf{x}, t)p(\mathbf{x}, t|\mathbf{y}, t) - W(\mathbf{x}|\mathbf{z}, t)p(\mathbf{z}, t|\mathbf{y}, t)]$$

$$p(\mathbf{z}, t + \Delta t|\mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z}) + \Delta t \int d\mathbf{x}\ [W(\mathbf{z}|\mathbf{x}, t)\delta(\mathbf{y} - \mathbf{x}) - W(\mathbf{x}|\mathbf{z}, t)\delta(\mathbf{y} - \mathbf{z})]$$

so that

$$p(\mathbf{z}, t + \Delta t|\mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z}) \left[1 - \int d\mathbf{x}\, W(\mathbf{x}|\mathbf{y}, t)\Delta t\right] + W(\mathbf{z}|\mathbf{y}, t)\Delta t \tag{3.47}$$

We see that for any Δt there is a finite probability, given by the coefficient of the $\delta(\mathbf{y} - \mathbf{z})$ in (3.47), for the particle to stay at the original position $\mathbf{y}$. The distribution of those particles which do not remain at $\mathbf{y}$ is given by $W(\mathbf{z}|\mathbf{y}, t)$ after appropriate normalization. Thus, a typical path $\mathbf{X}(t)$ will consist of sections of straight lines $\mathbf{X}(t) =$ constant, interspersed with discontinuous jumps whose distribution is given by $W(\mathbf{z}|\mathbf{y}, t)$. For this reason, the process is known as a jump process. The paths are discontinuous at discrete points.

(a) Time Between Jumps: The probability $Q(\mathbf{y}, t, t_0)$ that, given that we start from point $\mathbf{y}$ at time t_0, we are still at point $\mathbf{y}$ at time t is given for infinitesimal Δt by

$$Q(\mathbf{y}, t_0 + \Delta t, t_0) = 1 - \int d\mathbf{x}\, W(\mathbf{x}|\mathbf{y}, t_0)\Delta t \tag{3.48}$$

Clearly this means that

$$\frac{\partial Q(\mathbf{y},t,t_0)}{\partial t} = -\int d\mathbf{x}\, W(\mathbf{x}|\mathbf{y},t) Q(\mathbf{y},t,t_0) \tag{3.49}$$

If the jump probability is independent of t, then this has the simple solution

$$Q(\mathbf{y},t,t_0) = \exp\left(-\lambda t\right) \tag{3.50}$$

$$\text{with} \quad \lambda = \int d\mathbf{x} W(\mathbf{x}|\mathbf{y}) \tag{3.51}$$

Thus, the jump times are exponentially distributed, and can be simulated very simply. To simulate, one first chooses a jump time according to the probability law (3.50), and then chooses the value of $\mathbf{x}$ to which the jump was made according to a probability law

$$w(\mathbf{x}|\mathbf{y}) = \frac{W(\mathbf{x}|\mathbf{y})}{\lambda} \tag{3.52}$$

If the jump probability is not independent of t, the same procedure is in principle possible, but instead of using the exponential form (3.50), one must solve the differential equation (3.49).

(b) Integer State Space: In the case where the state space consists of integers only, the master equation takes the form

$$\partial_t P(\mathbf{n},t|\mathbf{n}',t') = \sum_m [W(\mathbf{n}|\mathbf{m},t)P(\mathbf{m},t|\mathbf{n}',t') - W(\mathbf{m}|\mathbf{n},t)P(\mathbf{n},t|\mathbf{n}',t')] \tag{3.53}$$

There is no longer any question that only jumps can occur, since only discrete values of the state variable $\mathbf{N}(t)$ are allowed. It is most important, however, to be aware that a pure jump process can occur even though the variable $\mathbf{X}(t)$ can take on a continuous range of variables.

3.5.2 Diffusion Processes: The Fokker-Planck Equation

When the quantities $W(\mathbf{z}|\mathbf{x},t)$ are zero, the differential Chapman-Kolmogorov equation reduces to the *Fokker-Planck equation:*

$$\frac{\partial p(\mathbf{z},t|\mathbf{y},t')}{\partial t} = -\sum_i \frac{\partial}{\partial z_i}[A_i(\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')] + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial z_i \partial z_j}[B_{ij}(\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')] \tag{3.54}$$

and the corresponding process is known mathematically as a *diffusion process.* The vector $\mathbf{A}(\mathbf{z},t)$ is known as the *drift vector* and the matrix $\mathbf{B}(\mathbf{z},t)$ as the *diffusion matrix.* The diffusion matrix is positive semidefinite and symmetric as a result of its definition (3.23). It is easy to see from (3.21), the definition of $W(\mathbf{x}|\mathbf{z},t)$, that the requirement (3.17) for continuity of the sample paths is satisfied if $W(\mathbf{x}|\mathbf{z},t)$ is zero. Hence, the Fokker-Planck equation describes a

process in which $\mathbf{X}(t)$ has continuous sample paths.

In fact, we can heuristically give a much more definite description of the process. Let us consider computing $p(\mathbf{z}, t + \Delta t|\mathbf{y}, t)$, given that

$$p(\mathbf{z}, t|\mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z}) \tag{3.55}$$

For small Δt, the solution of the Fokker-Planck equation will still be on the whole sharply peaked, and hence derivative of $A_i(\mathbf{z}, t)$ and $B_{ij}(\mathbf{z}, t)$ will be negligible compared to those of p. We are thus reduced to solving, approximately

$$\frac{\partial p(\mathbf{z}, t|\mathbf{y}, t')}{\partial t} = -\sum_i A_i(\mathbf{z}, t)\frac{\partial p(\mathbf{z}, t|\mathbf{y}, t')}{\partial z_i} + \frac{1}{2}\sum_{i,j} B_{ij}(\mathbf{z}, t)\frac{\partial^2 p(\mathbf{z}, t|\mathbf{y}, t')}{\partial z_i \partial z_j} \tag{3.56}$$

where we have neglected the time dependence of A_i and B_{ij} for small $t - t'$. Equation (3.56) can now be solved, subject to the initial condition (3.55), and we get

$$\begin{aligned} p(\mathbf{z}, t + \Delta t|\mathbf{y}, t) &= \{(2\pi)^N \det[\mathbf{B}(\mathbf{y}, t)\Delta t]\}^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}\frac{[\mathbf{z} - \mathbf{y} - \mathbf{A}(\mathbf{y}, t)\Delta t]^T[\mathbf{B}(\mathbf{y}, t)\Delta t]^{-1}[\mathbf{z} - \mathbf{y} - \mathbf{A}(\mathbf{y}, t)\Delta t]}{\Delta t}\right\} \end{aligned} \tag{3.57}$$

that is, a Gaussian distribution with variance matrix $\mathbf{B}(\mathbf{y}, t)$ and mean $\mathbf{y} + \mathbf{A}(\mathbf{y}, t)\Delta t$. We get the picture of a system moving with a systematic drift, whose velocity is $\mathbf{A}(\mathbf{y}, t)$, on which is superimposed a Gaussian fluctuation with covariance matrix $\mathbf{B}(\mathbf{y}, t)\Delta t$, that is we can write

$$\mathbf{y}(t + \Delta t) = \mathbf{y}(t) + \mathbf{A}(\mathbf{y}, t)\Delta t + \eta(t)\Delta t^{1/2} \tag{3.58}$$

where

$$\langle \eta(t) \rangle = 0 \tag{3.59}$$

$$\langle \eta(t)\eta(t)^T \rangle = \mathbf{B}(\mathbf{y}, t) \tag{3.60}$$

It is easy to see that this picture gives

(i) Sample paths which are always continuous, since it is clear that $\mathbf{y}(t + \Delta t) \to \mathbf{y}(t)$ as $\Delta t \to 0$.

(ii) Sample paths which are nowhere differentiable, because of the $\Delta t^{1/2}$ occurring in (3.58).

We will see later, in Chapter 4 that the heuristic picture of (3.58) can be made much more precise and leads to the concept of the *stochastic differential equation.*

3.5.3 Deterministic Processes: Liouville's Equation

It is possible that in the differential Chapman-Kolmogorov equation (3.42) only the first term is nonzero, so we are led to the special case of a *Liouville equation*:

$$\frac{\partial p(\mathbf{z},t|\mathbf{y},t')}{\partial t} = -\sum_i \frac{\partial}{\partial z_i}[A_i(\mathbf{z},t)p(\mathbf{z},t|\mathbf{y},t')] \tag{3.61}$$

which occurs in classical mechanics. This equation describes a completely deterministic motion, i.e., if $\mathbf{x})\mathbf{y},t)$ is the solution of the ordinary differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}[\mathbf{x}(t),t] \tag{3.62}$$

with

$$\mathbf{x}(\mathbf{y},t') = \mathbf{y} \tag{3.63}$$

then the solution to (3.61) with initial condition

$$p(\mathbf{z},t'|\mathbf{y},t') = \delta(\mathbf{z}-\mathbf{y}) \tag{3.64}$$

is

$$p(\mathbf{z},t|\mathbf{y},t') = \delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)] \tag{3.65}$$

The proof of this assertion is best obtained by direct substitution. For

$$\sum_i \frac{\partial}{\partial z_i}\{A_i(\mathbf{z},t)\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)]\} = \sum_i \frac{\partial}{\partial z_i}\{A_i(\mathbf{x}(\mathbf{y},t),t)\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)]\} \tag{3.66}$$

$$= \sum_i \left\{ A_i(\mathbf{x}(\mathbf{y},t),t)\frac{\partial}{\partial z_i}\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)] \right\} \tag{3.67}$$

and

$$\frac{\partial}{\partial t}\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)] = -\sum_i \frac{\partial}{\partial z_i}\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)]\frac{dx_i(\mathbf{y},t)}{dt} \tag{3.68}$$

and by use of (3.62), we see that (3.67) and (3.68) are equal. Thus, if the particle is in a well-defined initial position $\mathbf{y}$ at time t', it stays on the trajectory obtained by solving the ordinary differential equation (3.62).

Hence, deterministic motion, as defined by a first-order differential equation of the form (3.62), is an elementary form of Markov process. The solution (3.65) is, of course, merely a special case of the kind of process approximated by equations like (3.57) in which the Gaussian part is zero.

3.5.4 General Processes

In general, none of the quantities in $\mathbf{A}(\mathbf{z},t)$, $\mathbf{B}(\mathbf{z},t)$ and $W(\mathbf{x}|\mathbf{z},t)$ need vanish, and in this case we obtain a process whose sample paths are as illustrated in the figure below

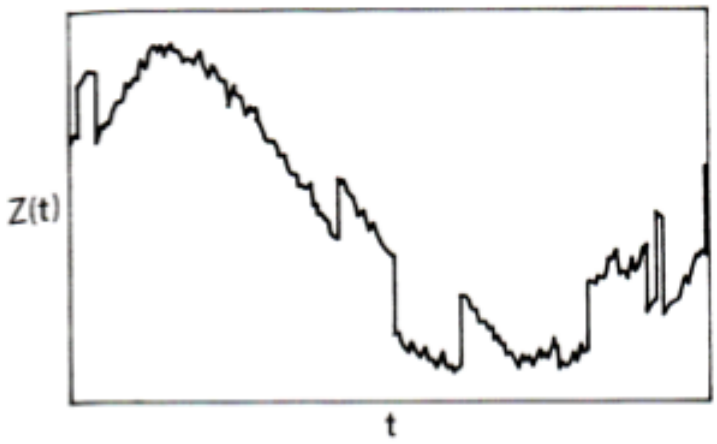

Figure 3.2: Illustration of a sample path of a general Markov process, in which drift, diffusion and jumps exist.

i.e., a piecewise continuous path made up of pieces which correspond to a diffusion process with a nonzero drift, onto which is superimposed a fluctuating part.

It is also possible that $\mathbf{A}(\mathbf{z}, t)$ is nonzero, but $\mathbf{B}(\mathbf{z}, t)$ is zero and here the sample paths are, as in the figure below

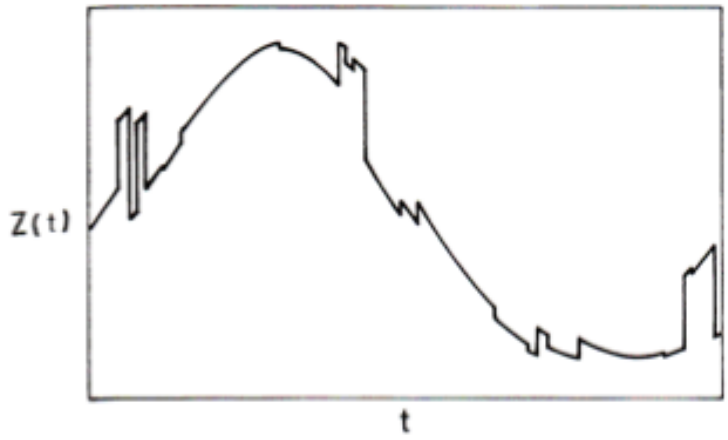

Figure 3.3: Sample path of a Markov process with only drift and jumps.

composed of pieces of smooth curve[solutions of (3.62)] with discontinuities

superimposed. This is very like the picture one would expect in a dilute gas where the particles move freely between collisions which cause an instantaneous change in momentum, though not position.

3.6 Equations for Time Development in Initial Time - Backward Equations

We can derive much more simply than in Section 3.4, some equations which give the time development with respect to the initial variables $\mathbf{y}, t'$ of $p(\mathbf{x}, t|\mathbf{y}, t')$.

We consider

$$\lim_{\Delta t' \to 0} \frac{1}{\Delta t'} \{p(\mathbf{x}, t|\mathbf{y}, t' + \Delta t') - p(\mathbf{x}, t|\mathbf{y}, t')\} \tag{3.69}$$

$$= \lim_{\Delta t' \to 0} \frac{1}{\Delta t'} \int d\mathbf{z}\, p(\mathbf{z}, t' + \Delta t'|\mathbf{y}, t') \{p(\mathbf{x}, t|\mathbf{y}, t' + \Delta t') - p(\mathbf{x}, t|\mathbf{z}, t' + \Delta t')\} \tag{3.70}$$

The second line follows by use of the Chapman-Kolmogorov equation in the second term and by noting that the first term gives $1 \times p(\mathbf{x}, t|\mathbf{y}, t' + \Delta t')$.

The assumptions that are necessary are now the existence of all relevant derivatives, and that $p(\mathbf{x}, t|\mathbf{y}, t')$ is continuous and bounded in $\mathbf{x}, t, t'$ for some range $t - t' > \delta > 0$. We may then write

$$(3.70) = \lim_{\Delta t' \to 0} \frac{1}{\Delta t'} \int d\mathbf{z}\, p(\mathbf{z}, t' + \Delta t'|\mathbf{y}, t') \{p(\mathbf{x}, t|\mathbf{y}, t') - p(\mathbf{x}, t|\mathbf{z}, t')\} \tag{3.71}$$

We now proceed using similar techniques to those used in Section 3.4.1 and finally derive

$$\frac{\partial p(\mathbf{x}, t|\mathbf{y}, t')}{\partial t} = -\sum_i A_i(\mathbf{y}, t') \frac{\partial p(\mathbf{x}, t|\mathbf{y}, t')}{\partial y_i} + \tfrac{1}{2} \sum_{i,j} B_{ij}(\mathbf{y}, t') \frac{\partial^2 p(\mathbf{x}, t|\mathbf{y}, t')}{\partial y_i \partial y_j}$$
$$+ \int d\mathbf{z}\, W(\mathbf{z}|\mathbf{y}, t') \{p(\mathbf{x}, t|\mathbf{y}, t') - p(\mathbf{x}, t|\mathbf{z}, t')\} \tag{3.72}$$

This will be called the *backward differential Chapman-Kolmogorov equation.* In a mathematical sense, it is better defined than the corresponding forward equation (3.42). The appropriate initial condition for both equations is

$$p(\mathbf{x}, t|\mathbf{y}, t) = \delta(\mathbf{x} - \mathbf{y}) \quad \text{for all } t \tag{3.73}$$

representing the obvious fact that if the particle is at $\mathbf{y}$ at time t, the probability deistic for finding it a $\mathbf{x}$ at the same time is $\delta(\mathbf{x} - \mathbf{y})$.

The forward and backward equations are equivalent to each other. For, solutions of the forward equation, subject to the initial condition (3.73) and any

appropriate boundary conditions, yield solutions of the Chapman-Kolmogorov equation as noted in Section 3.4.2. But these have just been shown to yield the backward equation. The basic difference is which set of variables is held fixed. In the case of the forward equation, we hold $\mathbf{y}$ and t' fixed, and solutions exist for $t \geq t'$, so that (3.73) is an *initial condition* for the forward equation. For the backward equation, solutions exist for $t' \leq t$, so that since the backwards equation expresses development in t', (3.73) is better termed a *finsl condition* in this case.

Since they are equivalent, the forward and backward equations are both useful. The forward equation gives more directly the values of measurable quantities as a function of the observed time t, and tends to be used more commonly in applications. The backward equation finds most application in the study of *first passage time or exit problems*, in which we find the probability that a particle leaves a region in a given time.

3.7 Stationary and Homogeneous Markov Processes

On Section 1.5.3 we met the concept of a stationary process, which represents the stochastic motion of a system which has settled down to a steady state, and whose stochastic properties are independent of when they are measured. Stationarity can be defined in various degrees, but we will reserve the term "*stationary process*" for a strict definition, namely, a stochastic process $\mathbf{X}(t)$ is stationary if $\mathbf{X}(t)$ and the process $\mathbf{X}(t+\epsilon)$ have the same statistics for any ϵ. This is equivalent to saying that all joint probability densities satisfy time translation invariance, i.e.,

$$\begin{aligned} &p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; .....; \mathbf{x}_n, t_n) \\ &\quad = p(\mathbf{x}_1 + \epsilon, t_1; \mathbf{x}_2, t_2 + \epsilon; \mathbf{x}_3, t_3 + \epsilon; .....; \mathbf{x}_n, t_n + \epsilon) \end{aligned} \tag{3.74}$$

and hence such probabilities are only functions of the time differences, $t_i - t_j$. In particular, the one-time probability is independent of time and can simply be written as

$$p_s(\mathbf{x}) \tag{3.75}$$

and the two-time joint probability as

$$p_s(\mathbf{x}_1, t_1 - t_2; \mathbf{x}_2, 0) \tag{3.76}$$

Finally, the conditional probability can also be written as

$$p_s((\mathbf{x}_1, t_1 - t_2 | \mathbf{x}_2, 0) \tag{3.77}$$

For a *Markov process*, since all joint probabilities can be written as products of the two-time conditional probability and the one-time probability, a necessary and sufficient condition for stationarity is the ability to write the one- and tow-time probabilities in the forms given in (3.74-3.77).

3.7.1 Ergodic Properties

If we have a stationary process, it is reasonable to expect that average measurements could be constructed by taking values of the variable x at successive times, and averaging various functions of these. This is effectively a belief that the law of large numbers (Section 2.5.2) applies to the variables defined by successive measurements in a stochastic process.

(a) Ergodic Property of the Mean: Let us define the variable $\bar{X}(T)$ by

$$\bar{X}(T) = \frac{1}{2T}\int_{-T}^{T} dt\, x(t) \tag{3.78}$$

where $x(t)$ is a stationary process, and consider the limit $T \to \infty$. This represents a possible model of measurement of the mean by averaging over all times

$$\langle \bar{X}(T) \rangle = \langle x \rangle_s \tag{3.79}$$

We now calculate the variance of $\bar{X}(T)$. Thus

$$\langle \bar{X}(T)^2 \rangle = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} dt_1 dt_2\, \langle x(t_1)x(t_2) \rangle \tag{3.80}$$

and if the process is stationary,

$$\langle x(t_1)x(t_2) \rangle \equiv R(t_1 - t_2) + \langle x \rangle^2 \tag{3.81}$$

where R is the two-time correlation function. Hence,

$$\langle \bar{X}(T)^2 \rangle - \langle x \rangle^2 = \frac{1}{4T^2}\int_{-2T}^{2T} d\tau\, R(\tau)(2T - |\tau|) \tag{3.82}$$

where the last factor follows by changing variables to

$$\tau = t_1 - t_2 \quad , \quad t = t_1 \tag{3.83}$$

and integrating t.

The left-hand side is now the variance of $\bar{X}(T)$ and we will show that under certain conditions, this vanishes as $T \to \infty$. Most straightforwardly, all we require is that

$$\lim_{T\to\infty} \frac{1}{T}\int_{-2T}^{2T} d\tau\, \left(1 - \frac{|\tau|}{2T}\right) R(\tau) = 0 \tag{3.84}$$

which is a little obscure. However, it is clear that a sufficient condition for this limit to be zero is for

$$\int_0^\infty d\tau\, |R(\tau)| < \infty \tag{3.85}$$

in which case, we simply require that the correlation function $\langle x(t_1)x(t_2)\rangle$ should tend to zero sufficient rapidly as $|t_1 - t_2| \to \infty$. In cases of interest it is frequently found that the asymptotic behavior of $R(\tau)$ is

$$R(\tau) \sim \Re[A \exp(-\tau/\tau_c)] \tag{3.86}$$

where τ_c is a (possibly complex) parameter known as the *correlation time*. Clearly the criterion (3.85) is satisfied, and we find in this case that the variance in $\bar{X}(T)$ approaches zero, so that using (3.79) and (2.102), we may write

$$\underset{T\to\infty}{ms-lim}\, \bar{X}(T) = \langle x\rangle_s \tag{3.87}$$

This means that the averaging procedure (3.78) is indeed valid. It is not difficult to extend the result to an average of an infinite set of measurements at discrete times $t_n = t_0 + n\Delta t$.

Other ergodic hypotheses can easily be stated, and the two quantities that are of the most interest are the autocorrelation function and the distribution function.

(b) Ergodic Property of the Autocorrelation Function: As already mentioned in Section 1.5.2, the most natural way of measuring an autocorrelation function is through the definition

$$G(\tau, T) = \frac{1}{T}\int_0^T dt\, x(t)x(t+\tau) \tag{3.88}$$

and we can rather easily carry through similar reasoning to show that

$$\underset{T\to\infty}{ms-lim}\, G(\tau, T) = \langle x(t)x(t+\tau)\rangle_s \tag{3.89}$$

provided the following condition is satisfied. Namely, define $\rho(\tau, \lambda)$ by

$$\langle x(t+\lambda+\tau)x(t+\lambda)x(t+\tau)x(t)\rangle_s = \rho(\tau,\lambda) + \langle x(t+\tau)x(t)\rangle_s^2 \tag{3.90}$$

Then we require

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|\tau|}{2T}\right)\rho(\tau,\lambda)d\lambda = 0 \tag{3.91}$$

We can see that this means that for sufficiently large λ, the four-time average (3.90) factorizes into a product of two-time averages, and that the "error term" $\rho(\tau, \lambda)$ must vanish sufficiently rapidly for $\lambda \to \infty$. Exponential behavior, such as that given in (3.86) is sufficient, and usually found.

(c) Ergodic Property of the Spectrum: We similarly find that the spectrum, given by the Fourier transform

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}G(\tau)d\tau \tag{3.92}$$

as seen in Section 1.5.2, is also given by the procedure

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} \int_0^T dt\, e^{-i\omega t} x(t) \tag{3.93}$$

(d) Ergodic Property of the Distribution Function: Finally, the practical method of measuring the distribution function is to consider an interval (x_1, x_2) and measure $x(t)$ repeatedly to determine whether it is in the range or not. This gives a measure of $\int_{x_1}^{x_2} dx\, p_s(x)$. Essentially, we are then measuring the time average value of the function $\chi(x)$ defined by

$$\chi(x) = \begin{cases} 1 & x_1 < x < x_2 \\ 0 & \text{otherwise} \end{cases} \tag{3.94}$$

and we adapt the method of proving the ergodicity of $\langle x \rangle$ to find that the distribution is ergodic provided

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T}\right) \int_{x_1}^{x_2} dx'\, p_s(x') \left\{ \int_{x_1}^{x_2} dx\, [p(x, \tau|x', 0) - p_s(x)] \right\} = 0 \tag{3.95}$$

The most obvious sufficient condition here is that

$$\lim_{\tau\to\infty} p(x, \tau|x', 0) = p_s(x) \tag{3.96}$$

and that this limit is approached sufficiently rapidly. In practice, an exponential approach isa frequently found and this is, as in the case of the mean, quite sufficiently rapid.

This condition is, in fact, sufficient for ergodicity of the mean and autocorrelation function for a Markov process, since all means can be expressed in terms of conditional probabilities and the sufficiently rapid achievement of the limit (3.96) can be readily seen to be sufficient to guarantee both (3.91) and (3.84). We will call a Markov process simply *ergodic* if this rather strong condition is satisfied.

3.7.2 Homogeneous Properties

If the condition (3.96) is satisfied for a stationary Markov process, then we clearly have a way of constructing from the stationary Markov process a non stationary process whose limit as time becomes large is the stationary process. We simply define the process for $t, t' > t_0$ by

$$p(\mathbf{x}, t) = p_s(\mathbf{x}, t|\mathbf{x}_0, t_0) \tag{3.97}$$

$$p(\mathbf{x}, t|\mathbf{x}', t') = p_s(\mathbf{x}, t|\mathbf{x}', t') \tag{3.98}$$

and all other joint probabilities are obtained from these in the usual manner for a Markov process. Clearly, if (3.96) is satisfied, we find that as $t \to \infty$ or as $t_0 \to -\infty$,

$$p(\mathbf{x}, t) \to p_s(\mathbf{x}) \tag{3.99}$$

and all other probabilities become stationary because the conditional probability is stationary. Such a process is known as a *homogeneous* process.

The physical interpretation is rather obvious. We have a stochastic system whose variable $\mathbf{x}$ is by some external agency fixed to have a value $\mathbf{x}_0$ at time $t_)$. It then evolves back to a stationary system with the passage of time. This is how many stationary systems are created in practice.

From the point of view of the differential Chapman-Kolmogorov equation, we will find that the stationary distribution function $p_s(x)$ is a solution of the stationary differential Chapman-Kolmogorov equation, which takes the form

$$\begin{aligned} 0 = &- \sum_i \frac{\partial}{\partial z_i} A_i(\mathbf{z}) p(\mathbf{z}, t|\mathbf{y}, t') + \tfrac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j} B_{ij}(\mathbf{z}) p(\mathbf{z}, t|\mathbf{y}, t') \\ &+ \int d\mathbf{x} \, [W(\mathbf{z}|\mathbf{x}) p(\mathbf{x}, t|\mathbf{y}, t') - W(\mathbf{x}|\mathbf{z}) p(\mathbf{z}, t|\mathbf{y}, t')] \end{aligned} \tag{3.100}$$

where we have used the fact that the process is homogeneous to note that A, B, and W, as defined by (3.21-3.23), are independent of t. This is an alternative definition of a homogeneous process.

3.7.3 Approach to a Stationary Process

A converse problem also exists. Suppose A, B, and W are independent of time and $p_s(\mathbf{z})$ satisfies (3.100). Under what conditions does a solution of the differential Chapman-Kolmogorov equation approach a stationary solution of $p_s(\mathbf{z})$?

There does not appear to be a complete answer to this problem. However, we can give a reasonably good picture as follows. We define a *Lyapunov functional* K of any two solutions p_1 and p_2 of the differential Chapman-Kolmogorov equation by

$$K = \int d\mathbf{x} \, p_1(\mathbf{x}, t) \left[\log \frac{p_1(\mathbf{x}, t)}{p_2(\mathbf{x}, t)} \right] \tag{3.101}$$

and assume for the moment that neither p_1 nor p_2 are zero anywhere. We will now show that K is always positive and dK/dt is always negative.

Firstly, noting that both $p_2(\mathbf{x}, t)$ and $p_1(\mathbf{x}, t)$ are normalized to one, we write

$$K(p_1, p_2, t) = \int d\mathbf{x} \, p_1(\mathbf{x}, t) \left\{ \log \frac{p_1(\mathbf{x}, t)}{p_2(\mathbf{x}, t)} + \frac{p_2(\mathbf{x}, t)}{p_1(\mathbf{x}, t)} - 1 \right\} \tag{3.102}$$

and use the inequality valid for all, $z > 0$,

$$-\log z + z - 1 \geq 0 \tag{3.103}$$

to show that $K \geq 0$.

Now let us show that $dK/dt \leq 0$, We can write (using an abbreviated notation)

$$\frac{dK}{dt} = \int d\mathbf{x} \left\{ \frac{\partial p_1}{\partial t} [\log p_1 + 1 - \log p_2] - \frac{\partial p_2}{\partial t} \left[\frac{p_1}{p_2} \right] \right\} \tag{3.104}$$

We now calculate one by one the contributions to dK/dt from drift, diffusion, and jump terms in the differential Chapman-Kolmogorov equation:

$$\left(\frac{dK}{dt} \right)_{drift} = \sum_i \int \mathbf{x} \left\{ - \left[\log \frac{p_1}{p_2} + 1 \right] \frac{\partial}{\partial x_i} (A_i p_1) + \frac{p_1}{p_2} \frac{\partial}{\partial x_i} (A_i p_2) \right\} \tag{3.105}$$

which can be rearranged to give

$$\left(\frac{dK}{dt} \right)_{drift} = \sum_i \int \mathbf{x} \frac{\partial}{\partial x_i} \left[-A_i p_1 \log \left(\frac{p_1}{p_2} \right) \right] \tag{3.106}$$

Similarly, we may calculate

$$\left(\frac{dK}{dt} \right)_{diff} = -\tfrac{1}{2} \sum_{i,j} \int \mathbf{x} \left\{ \left[\log \frac{p_1}{p_2} + 1 \right] \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij} p_1) - \frac{p_1}{p_2} \frac{\partial^2}{\partial x_i \partial x_j} (B N_{ij} p_2) \right\} \tag{3.107}$$

and after some rearranging we may write

$$\begin{aligned} \left(\frac{dK}{dt} \right)_{diff} = & -\tfrac{1}{2} \sum_{i,j} \int \mathbf{x}\, p_1 B_{ij} \left\{ \frac{\partial}{\partial x_i} \left[\log \frac{p_1}{p_2} \right] \right\} \left\{ \frac{\partial}{\partial x_i} \left[\log \frac{p_1}{p_2} \right] \right\} \\ & + \tfrac{1}{2} \sum_{i,j} \int \mathbf{x} \frac{\partial^2}{\partial x_i \partial x_j} \left[p_1 B_{ij} \log \frac{p_1}{p_2} \right] \end{aligned} \tag{3.108}$$

Finally, we may calculate the jump contribution similarly:

$$\begin{aligned} \left(\frac{dK}{dt} \right)_{jump} = \int d\mathbf{x} d\mathbf{x}' \{ & [W(\mathbf{x}|\mathbf{x}') p_1(\mathbf{x}', t) - W(\mathbf{x}'|\mathbf{x}) p_1(\mathbf{x}, t)] \\ & \times \left\{ \log \left[\frac{p_1(\mathbf{x}, t)}{p_2(\mathbf{x}, t)} \right] + 1 \right\} \\ & - [W(\mathbf{x}|\mathbf{x}') p_2(\mathbf{x}', t) - W(\mathbf{x}'|\mathbf{x}) p_2(\mathbf{x}, t)] \frac{p_1(\mathbf{x}, t)}{p_2(\mathbf{x}, t)} \Big\} \end{aligned} \tag{3.109}$$

and after some rearrangement,

$$\left(\frac{dK}{dt} \right)_{jump} = \int d\mathbf{x} d\mathbf{x}'\, W(\mathbf{x}|\mathbf{x}') \left\{ p_2(\mathbf{x}', t) \left[\phi' \log \frac{\phi}{\phi'} \right] - \phi + \phi' \right\} \tag{3.110}$$

where

$$\phi = \frac{p_1(\mathbf{x}, t)}{p_2(\mathbf{x}, t)} \tag{3.111}$$

and ϕ' is similarly defined in terms of $\mathbf{x}'$.

We now consider the simplest case. Suppose a stationary solution $p_s(\mathbf{x})$ exists which is nonzero everywhere, except at infinity, where it and its first derivative vanish. Then we may choose $p_2(\mathbf{x}, t) = p_s(\mathbf{x}$. The contribution to dK/dt from (3.106) and the second term in (3.108) can be integrated to give surface terms which vanish at infinity so we find

$$\left(\frac{dK}{dt}\right)_{drift} = 0 \tag{3.112a}$$

$$\left(\frac{dK}{dt}\right)_{diff} \leq 0 \tag{3.112b}$$

$$\left(\frac{dK}{dt}\right)_{jump} \leq 0 \tag{3.112c}$$

where the last inequality comes by setting $z = \phi'/\phi$ in (3.103).

We must now consider under what situations the *equalities* in (3.112c) are actually achieved. Inspection of (3.110) shows that this term will be zero if and only if $\phi = \phi'$ for almost all $\mathbf{x}$ and $\mathbf{x}'$ which are such that $W(\mathbf{x}|\mathbf{x}') \neq 0$. Thus, if $W(\mathbf{x}|\mathbf{x}')$ is never zero, i.e., if transitions can take place in *both directions* between any pair of states, the vanishing of the jump contribution implies that $\phi(x) = \phi(x')$ for all x anmd x', that is, $\phi(x)$ is independent of x, so that

$$\frac{p_1(\mathbf{x}, t)}{p_s(\mathbf{x})} = \text{ constant} \tag{3.113}$$

The constant must equal one since $p_1(\mathbf{x}, t)$ and $p_s(\mathbf{x})$ are both normalized.

The term arising from diffusion will be strictly negative if B_{ij} is almost everywhere positive definite. Hence, we have now shown that under rather strong conditions, namely,

$$\left.\begin{array}{ccc} p_s(\mathbf{x}) & \neq 0 & \text{with probability 1} \\ W(\mathbf{x}|\mathbf{x}') & \neq 0 & \text{with probability 1} \\ \mathbf{B}_{ij}(\mathbf{x}) & \text{positive definite} & \text{with probability 1} \end{array}\right\} \tag{3.114}$$

that any solution of the differential Chapman-Kolmogorov equation approaches the stationary solution $p_s(\mathbf{x})$ at $t \to \infty$.

This result fails in two basic kinds of systems.

(a) **Disconnected State Space**: The result is best illustrated when A_i and B_{ij} vanish so we have a pure jump system. Suppose the space divides into

two regions R_1 and R_2 such that transitions from R_1 to R_2 and back are impossible; hence $W(\mathbf{x}|\mathbf{x}') = 0$ if $\mathbf{x}$ and $\mathbf{x}'$ are not both in R_1 or R_2. Then it is possible to have $dK/dt = 0$ if

$$p_1(\mathbf{x}, t) = \begin{cases} \lambda_1 p_s(\mathbf{x}) & x \in R_1 \\ \lambda_2 p_s(\mathbf{x}) & x \in R_2 \end{cases} \tag{3.115}$$

so that there are no unique stationary distributions. The two regions are disconnected and separate stochastic processes take places in each, and in each of these, there is a unique stationary solution. The relative probability of being R_1 or R_2 is not changed by the process.

A similar result holds, in general, if as well we have B_{ij} and A_i vanishing on the boundary between R_1 and R_2.

(b) $p_s(\mathbf{x})$ **Vanishes in Some Definitie Region**: If we have

$$p_s(\mathbf{x}) \begin{cases} = 0 & \mathbf{x} \in R_1 \\ \neq 0 & \mathbf{x} \in R_2 \end{cases} \tag{3.116}$$

and again A_i and B_{ij} vanish, then jot follows that, since $p_s(\mathbf{x})$ satisfies the stationary equation (3.100),

$$W(\mathbf{x}|\mathbf{y}) = 0 \quad , \quad \mathbf{x} \in R_1, \mathbf{x} \in R_2 \tag{3.117}$$

In other words, no transitions are possible from the region R_2 where the stationary distribution is positive to R_1, where the stationary distribution vanishes.

3.7.4 Autocorrelation Function for Markov Processes

For any Markov process, we can write a very elegant formula for the autocorrelation function. We define

$$\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle = \int d\mathbf{x}\, \mathbf{x} p(\mathbf{x}, t|\mathbf{x}_0, t_0) \tag{3.118}$$

then the autocorrelation matrix is

$$\langle \mathbf{X}(t)\mathbf{X}(t_0)^T \rangle = \int d\mathbf{x} \int d\mathbf{x}_0\, \mathbf{x}\mathbf{x}_0^T p(\mathbf{x}, t|\mathbf{x}_0, t_0) \tag{3.119}$$

$$= \int d\mathbf{x}_0\, \langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle\, \mathbf{x}_0^T p(\mathbf{x}_0, t_0) \tag{3.120}$$

Thus, we see that (3.118) defines the mean of $\mathbf{X}(t)$ under the condition that $\mathbf{X}$ had the value $\mathbf{x}_0$ at time t_+0, and (3.120) tells us that the autocorrelation matrix is obtained by averaging this conditional average (multiplied by $\mathbf{x}_0^T$) at time t_0. These results are true buy definition for any stochastic process.

In a Markov process we have, however, a unique conditional probability which determines the whole process. Thus, for a Markov process, $\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle$ is a uniquely defined quantity, since the knowledge of $\mathbf{x}_0$ at time t_0 completely determines the future of the process.

(a) **Stationary Autocorrelation Function**: The most notable use of this property is in the computation of the stationary autocorrelation function. To illustrate how this uniqueness is important, let us consider a non-Markov stationary process with joint probabilities

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; .....\mathbf{x}_n, t_n) \tag{3.121}$$

which, of course, depend only on time differences. Let us now create a corresponding non stationary process by selecting only simple paths which pass through the point $x = a$ at time $t = 0$. Thus, we define

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; .....\mathbf{x}_n, t_n) = p_s(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; .....\mathbf{x}_n, t_n | \mathbf{a}, 0) \tag{3.122}$$

Then for this process we note that

$$\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle_{\mathbf{a}} = \int d\mathbf{x}\, \mathbf{x} p_s(\mathbf{x}, t | \mathbf{x}_0, t_0;\, mathbfa, 0) \tag{3.123}$$

which contains a dependence on $\mathbf{a}$ symbolized by the subscript $\mathbf{a}$ on the average bracket. If the original stationary process possesses appropriate ergodic properties then

$$\lim_{\tau \to \infty} p_s(\mathbf{x}, t + \tau | \mathbf{x}_0, t_0 + \tau; \mathbf{a}, 0) = p_s(\mathbf{x}, t - t_0 | \mathbf{x}_0, 0) \tag{3.124}$$

so that we will also have a stationary conditional average of $\mathbf{x}$

$$\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle_s = \lim_{\tau \to \infty} \langle \mathbf{X}(t + \tau) \,|\, [\mathbf{x}_0, t_0 + \tau] \rangle_{\mathbf{a}} \tag{3.125}$$

and the stationary autocorrelation matrix is given by

$$\langle \mathbf{X}(t) \mathbf{X}(t_0)^T \rangle_s = \int d\mathbf{x}_0\, \mathbf{x}_0^T \, \langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle_s \, p_s(\mathbf{x}_0) \tag{3.126}$$

$$= \lim_{\tau \to \infty} \langle \mathbf{X}(t + \tau) \,|\, \mathbf{X}(t_0 + \tau) \rangle_{\mathbf{a}}$$

$$= \lim_{\tau \to \infty} \int d\mathbf{x}_0\, \mathbf{x}_0^T \, \langle \mathbf{x}(t + \tau) \,|\, [\mathbf{x}_0, t_0 + \tau] \rangle_{\mathbf{a}} \, p_{\mathbf{a}}(\mathbf{x}_0, t_0 + \tau) \tag{3.127}$$

However, when the process is Markovian, this cumbersome limiting procedure is not necessary since

$$\begin{aligned} \text{Markov } \Rightarrow \langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle_s &= \langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle_{\mathbf{a}} \\ &= \langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \rangle \end{aligned} \tag{3.128}$$

(b) **Regression Theorem**: Equation (3.120) is a *regression theorem* when applied to a Markov process and is the basis of a more powerful regression theorem for *linear systems*. By this we mean systems such that a linear equation of motion exists for the mean, i.e.,

$$\frac{d\left\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \right\rangle}{dt} = -A \left\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \right\rangle \tag{3.129}$$

which is very often the case in systems of practical interest, either as an exact result or as an approximation. The initial conditions for (3.129) are clearly

$$\left\langle \mathbf{X}(t) \,|\, [\mathbf{x}_0, t_0] \right\rangle = \mathbf{x}_0 \tag{3.130}$$

Then from (3.128,3.129)

$$\frac{d}{dt}\langle \mathbf{X}(t)\mathbf{X}(t_0)^T \rangle = -A\langle \mathbf{X}(t)\mathbf{X}(t_0)^T \rangle \tag{3.131}$$

with the initial conditions $\langle \mathbf{X}(t_0)\mathbf{X}(t_0)^T \rangle$. The time correlation matrix

$$\langle \mathbf{X}(t)\mathbf{X}(t_0)^T \rangle - \langle \mathbf{X}(t) \rangle \langle \mathbf{X}(t_0)^T \rangle = \langle \mathbf{X}(t), \mathbf{X}(t_0)^T \rangle \tag{3.132}$$

obviously obeys the same equation, with the initial condition given by the covariance matrix at time t_0. In a stationary system, we have the result that if $G(t)$ is the stationary time correlation function and σ the stationary covariance matrix, then

$$\frac{dG(t)}{dt} = -AG(t) \tag{3.133}$$

$$G(0) = \sigma \tag{3.134}$$

so that

$$G(t) = \exp\left[-At\right]\sigma \tag{3.135}$$

which is the *regression theorem* in its simplest form. We again stress that it is valid for the *Markov processes* in which the mean values obey *linear* evolution equations like (3.129).

For non-Markov processes there is not simple procedure. We must carry out the complicated procedure implicit in (3.127).

3.8 Examples of Markov Processes

We present here for reference some fundamental solutions of certain cases of the differential Chapman-Kolmogorov equation.

3.8.1 The Wiener Process

This takes its name from *Norbert Wiener* who studied it extensively. From the point of view of this chapter, it is a solution of the Fokker-Planck equation as discussed in Section 3.5.2, in which there is only one variable $W(t)$, the drift coefficient is zero and the diffusion coefficient is 1. Thus, the Fokker-Planck equation for this case is

$$\frac{\partial}{\partial t}p(w,t|w_0,t_0) = \frac{1}{2}\frac{\partial^2}{\partial w^2}p(w,t|w_0,t_0) \tag{3.136}$$

Utilizing the initial condition

$$p(w,t_0|w_0,t_0) = \delta(w-w_0) \tag{3.137}$$

on the conditional probability, we solve (3.136) by use of the characteristic function

$$\phi(s,t) = \int dw\, p(w,t|w_o,t_0)\exp(isw) \tag{3.138}$$

which satisfies

$$\frac{\partial\phi}{\partial t} = -\frac{1}{2}s^2\phi \tag{3.139}$$

so that

$$\phi(s,t) = \exp\left[-\frac{1}{2}s^2(t-t_0)\right]\phi(s,t_0) \tag{3.140}$$

From (3.137), the initial condition is

$$\phi(s,t_0) = \exp(isw_o) \tag{3.141}$$

so that

$$\phi(s,t) = \exp\left[isw_0 - \frac{1}{2}s^2(t-t_0)\right] \tag{3.142}$$

Performing the Fourier inversion, we have the solution to (3.136):

$$p(w,t|w_0,t_0) = \frac{1}{\sqrt{2\pi(t-t_0)}}\exp\left(-\frac{(w-w_0)^2}{2(t-t_0)}\right) \tag{3.143}$$

This represents a Gaussian with

$$\langle W(t)\rangle = w_0 \tag{3.144}$$

$$\langle[W(t)-w_0]^2\rangle = t-t_0 \tag{3.145}$$

so that the initially sharp distribution spreads in time, as graphed in the figure below.

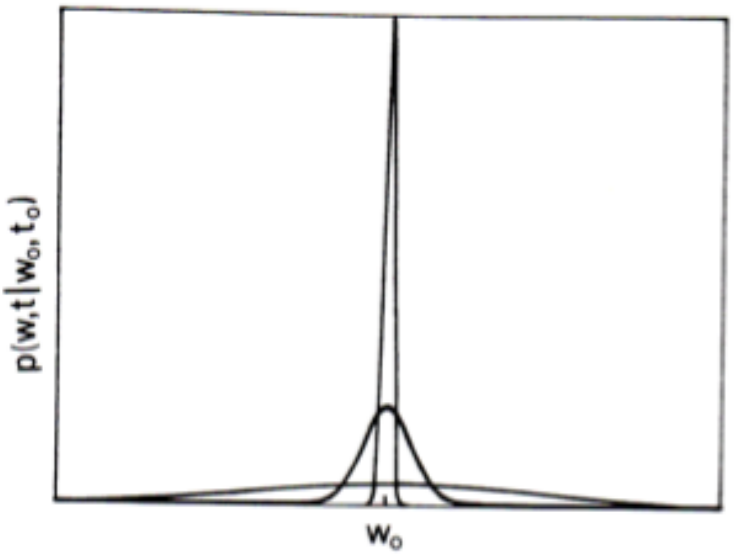

Figure 3.4: Wiener process: spreading of an initially sharp distribution $p(w, t|w_0, t_0)$ with increasing time $t - t_0$

A multivariate Wiener process can be defined as

$$\mathbf{W}(t) = [W_1(t), W_2(t), ..., W_n(t)] \tag{3.146}$$

which satisfies the multivariate Fokker-Planck equation

$$\frac{\partial}{\partial t} p(\mathbf{w}, t|\mathbf{w}_0, t_0) = \frac{1}{2} \sum_i \frac{\partial^2}{\partial w_i^2} p(\mathbf{w}, t|\mathbf{w}_0, t_0) \tag{3.147}$$

whose solution is

$$p(\mathbf{w}, t|\mathbf{w}_0, t_0) = \frac{1}{[2\pi(t - t_0)]^{n/2}} \exp\left(-\frac{(\mathbf{w} - \mathbf{w}_0)^2}{2(t - t_0)}\right) \tag{3.148}$$

a multivariate Gaussian with

$$\langle \mathbf{W}(t) \rangle = \mathbf{w}_0 \tag{3.149}$$

and

$$\langle [W_i(t) - w_{0i}][W_j(t) - w_{0j}] \rangle = (t - t_)\delta_{ij} \tag{3.150}$$

The one-variable Wiener process is often simply call Brownian motion, since the Wiener process equation (3.136) is exactly the same as the differential equation of diffusion, shown by Einstein to be obeyed by Brownian motion, as we noted in Section 1.2. The terminology is, however, not universal.

Points of note concerning the Wiener process are:

(a) **Irregularity of Sample Paths**: Although the mean value of $W(t)$ is zero, the mean square becomes infinite as $t \to infty$. This means that the sample paths of $W(t)$ are *very* variable, indeed surprisingly so. In the figure below, we have given a few different sample paths with the same initial point to illustrate the extreme non-reproducibility of the paths.

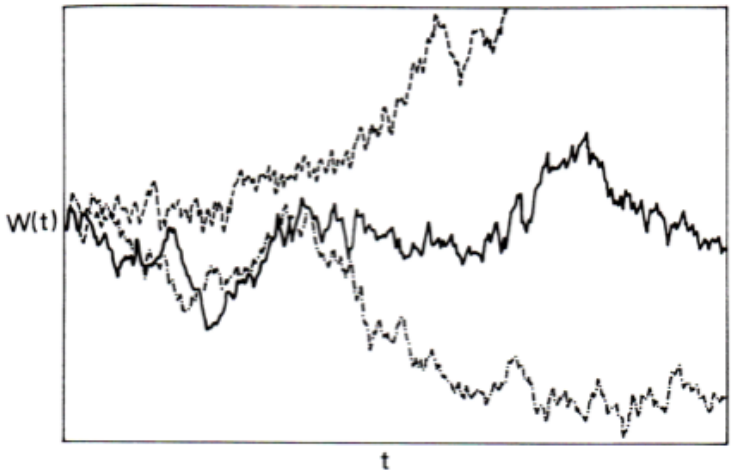

Figure 3.5: Three simulated sample paths of the Wiener process, illustrating their great variability

(b) **Non-differentiability of Sample Paths**: The Wiener process is a diffusion process and hence the sample paths of $W(t)$ are continuous. However, they are not differentiable. Consider

$$Prob\left\{\left|\frac{[W(t+h)-W(t)}{h}]\right| > k\right\} \tag{3.151}$$

From the solution for the conditional probability, this probability is

$$2\int_{kh}^{\infty} dw\, \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{w^2}{2h}\right) \tag{3.152}$$

and in the limit $h \to 0$, this is one. This means that no matter what value of k we choose, $|[W(t+h)-W(t)]/h|$ is almost certain to be greater than this, i.e., the derivative at any point is almost certainly infinite. This is in agreement with the similar intuitive picture presented in Section 3.5.2 and the simulated paths given in Figure 3.5 illustrate the point dramatically. This corresponds, of course, to the well-known experimental fact that the Brownian particles have an exceedingly irregular motion. However, this is clearly an idealization, since if $W(t)$ represents the position of the

Brownian particle, this means that its speed is almost certainly infinite. The *Ornstein-Uhlenbeck* process is a more realistic model of Brownian motion (Section 3.8.4).

(c) **Independence of Increment**: The Wiener process is fundamental to the study of diffusion processes, and by means of stochastic differential equations, we can express any diffusion process in terms of the Wiener process. Of particular importance is the statistical independence of the increments of $W(t)$. More precisely, since the Wiener process is a Markov process, the joint probability density can be written

$$\begin{aligned} &p(w_n, t_n; w_{n-1}, t_{n-1}; w_{n-2}, t_{n-2}; .....; w_0, t_0) \\ &\qquad = \prod_{i=0}^{n-1} p(w_{i+1}, t_{i+1} | w_i, t_i) p(w_0, t_0) \end{aligned} \tag{3.153}$$

and using the explicit form of the conditional probabilities (3.143), we see that

$$\begin{aligned} &p(w_n, t_n; w_{n-1}, t_{n-1}; w_{n-2}, t_{n-2}; .....; w_0, t_0) \\ &\quad = \prod_{i=0}^{n} \left\{ \frac{1}{\sqrt{2\pi(t_{i+1} - t_i)}} \exp\left(-\frac{(w_{i+1} - w_i)^2}{2(t_{i+1} - t_i)} \right) \right\} p(w_0, t_0) \end{aligned} \tag{3.154}$$

If we define the variables

$$\Delta W_i \equiv W(t_i) - W(t_{i-1}) \tag{3.155}$$

$$\Delta t_i = t_i - t_{i-1} \tag{3.156}$$

then the joint probability density for ΔW_i is

$$\begin{aligned} &p(\Delta w_n; \Delta w_{n-1}; \Delta w_{n-2}; .....\Delta w_1; w_0) \\ &\qquad = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \Delta t_i}} \exp\left(-\frac{\Delta w_i^2}{2\Delta t_i} \right) \right\} p(w_0, t_0) \end{aligned} \tag{3.157}$$

which shows, from the definition of statistical independence given in Section 2.3.4, that the variables ΔW_i are independent of each other and of $W(t_0)$.

The aspect of having independent increments ΔW_i is very important in the definition of stochastic integration which is carried out in Section 4.2.

(d) **Autocorrelation Functions**: A quantity of great interest is the autocorrelation function, already discussed in Sections 1.5.2 and 3.7.4. The formal definition is

$$\langle W(t) W(s) \,|\, [W_0, t_0] \rangle = \int dW - 1 dw_2 \, w_1 w_2 p(w_1, t; w_2, s | w_0, t_0) \tag{3.158}$$

which is the mean product of $W(t)$ and $W(s)$ on the condition that the initial value is $W(t_0) = w_0$, and we can see, assuming $t > s$, that

$$\langle W(t)W(s) \,|\, [W_0, t_0]\rangle = \langle [W(t) - W(s)]W(s)\rangle + \langle [W(s)]^2\rangle \tag{3.159}$$

Using the independence of increments, the first average is zero and the second is given by (3.145) so that we have, in general,

$$\langle W(t)W(s) \,|\, [W_0, t_0]\rangle = min(t - t_0), s - t_0) + w_0^2 \tag{3.160}$$

which is correct for both $t > s$ and $t < s$.

3.8.2 The Random Walk in One Dimension

A man moves along a line, taking, at random, steps to the left or the right with equal probability. The steps are of length l so that his position can take on only the value nl, where n is integral. We want to know the probability that he reaches a given point a distance nl from the origin after a given elapsed time.

The problem can be defined in two ways. The first, which is more traditional, is to allow the walker to take steps at times $N\tau$ (N integral) at which times he *must* step either left or right, with equal probability. The second is to allow the walker to take steps left or right with probability per unit time d which means that the walker waits at each point for a *variable* time. The second method is describable by a master equation.

(a) **Continuous Time Random Walk**: To do a master equation treatment of the problem, we consider that the transition probability per unit time is given by the form

$$W(n + 1|n, t) = W(n - 1|n, t) = d \tag{3.161}$$

otherwise, $W(n|m, t) = 0$ so that, according to Section 3.5.1, the master equation for the man to be at the position nl, given that he started at $n'l$, is

$$\partial_t P(n, t|n', t') = d[P(n + 1, t|n', t') + P(n - 1, t|n', t')] \tag{3.162}$$

(b) **Discrete Time Random Walk**: the more classical form of the random walk does not assume that the man makes his jump to the left or right according to a master equation, but that he jumps left or right with equal probability at times $N\tau$, so that time is a discrete variable. In this case, we can write

$$P(n, (N+1)\tau|n', N'\tau) = \frac{1}{2}\{P(n + 1, N\tau|n', N'\tau) + (P(n - 1, N\tau|n', N'\tau)\} \tag{3.163}$$

If τ is small, we can view (3.162) and (3.163) as approximations to each other by writing

$$P(n, (N + 1)\tau|n', N'\tau) \simeq P(n, N\tau|n', N'\tau) + \tau\partial_t P(n, t|n', t') \tag{3.164}$$

with $t = N\tau$, $t' = N'\tau$ and $d = 1/2\tau.$, so that the transition probability per unit time in the master equation model corresponds to half the inverse waiting time τ in the discrete time model.

(c) **Solutions Using the Characteristic Function**: Both systems can be easily solved by introducing the characteristic function

$$G(s,t) = \langle e^{ins} \rangle = \sum_n P(n,t|n',t') e^{ins} \tag{3.165}$$

in which case the master equation gives

$$\partial_t G(s,t) = d(e^{is} + e^{-is} - 2)G(s,t) \tag{3.166}$$

and the discrete time equation becomes

$$G(s,(N+1)\tau) = \frac{1}{2}(e^{is} + e^{-is})G(s,N\tau) \tag{3.167}$$

Assuming the man starts at the origin $n' = 0$ at time $t' = 0$, we find

$$G(s,0) = 1 \tag{3.168}$$

in both cases, so that the solution to (3.166) is

$$G_1(s,t) = \exp\left[(e^{is} + e^{-is} - 2)td\right] \tag{3.169}$$

and to (3.167)

$$G_2(s,N\tau) = \left[\frac{1}{2}(e^{is} + e^{-is})\right]^N \tag{3.170}$$

The appropriate probability distributions can be obtained by expanding $G_1(s,t)$ and $G_2(s,t)$ in powers of $\exp(is)$; we find

$$P_1(n,t|0,0) = e^{-2td} I_n(4td) \tag{3.171}$$

$$P_2(n,N\tau|0,0) = \left(\frac{1}{2}\right)^N N! \left[\left(\frac{N-n}{2}\right)! \left(\frac{N+n}{2}\right)!\right]^{-1} \tag{3.172}$$

The discrete time distribution is also known as the Bernoulli distribution; it gives the probability of a total of n heads in tossing an unbiased coin N times.

(d) **Continuous Space Limit**: For both kinds of random walk, the limit of continuous space - that is, very many steps of very small size - gives the Wiener process. If we set the distance travelled as

$$x = nl \tag{3.173}$$

so that the characteristic function of the distribution of x is

$$\phi_1(s,t) = \langle e^{isx} \rangle = G_1(ls,t) = \exp\left[(e^{ils} + e^{-ils}) - 2\right]td \tag{3.174}$$

Then the limit of infinitesimally small steps $l \to 0$ is

$$\phi_1(s,t) \to \exp\left(-s^2 t D\right) \tag{3.175}$$

where

$$D = \lim_{l\to 0}(l^2 d) \tag{3.176}$$

This is the characteristic function of a Gaussian (Section 2.8.1) of the form

$$P(x,t|0,0) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right) \tag{3.177}$$

and is of course the distribution for the Wiener process (Section 3.81.) or Brownian motion, as mentioned in Section 1.2. Thus, the Wiener process can be regarded as the limit of a continuous time random walk in the limit of infinitesimally small step size.

The limit

$$l \to 0 \ , \ \tau \to 0 \ , \ \text{with} \ D = \lim_{l\to 0}(l^2/\tau) \tag{3.178}$$

of the discrete time random walk gives the same result. From this form, we see clearly the expression of D as the mean square distance travelled per unit time.

We can also see more directly that expanding the right-hand side of (3.162) as a function of x up to second order in l gives

$$\partial_t p(x,t|0,0) = l^2 d \partial_x^2 P(x,t|0,0) \tag{3.179}$$

The three processes are thus intimately connected with each other at two levels, namely, under the limits considered, the stochastic equations approach each other and under those same limits, the solutions to these equations approach each other. These limits are exactly those used by Einstein. Comparison with Section 1.2 shows that he modeled Brownian motion by a discrete time and space random walk, but nevertheless, derived the Wiener process model by expanding the equations for time development of the distribution function.

The limit results of this section are a slightly more rigorous version of Einstein's method. There are generalizations of these results to less specialized situations and it is a fair statement to make that almost any jump process has some kind of limit which is a diffusion process. However, the precise limits are not always so simple, and there are limits in which certain jump processes become deterministic and are governed by Liouville's equation (Section 3.5.3) rather than the full Fokker-Planck equation.

3.8.3 Poisson Process

We have already noted the Poisson process in Section 1.5.1. The process in which electrons arrive at an anode or customers arrive at a shop with probability per unit time λ of arriving, is governed by the master equation for which

$$W(n+1|n,t) = \lambda \tag{3.180}$$

otherwise,

$$W(n|m,t) = 0 \tag{3.181}$$

This master equation becomes

$$\partial_t P(n,t|n',t') = \lambda[P(n-1,t|n',t') - P(n,t|n',t')] \tag{3.182}$$

and by comparison with (3.162) also represents a "one-sided" random walk, in which the walker steps to the right only with probability per unit time equal to λ.

The characteristic function equation is similar to (3.166):

$$\partial_t G(s,t) = \lambda[e^{is} - 1]G(s,t) \tag{3.183}$$

with the solution

$$G(s,t) = e^{\lambda t[e^{is}-1]} \tag{3.184}$$

for the initial condition that there are initially no customers (or electrons) at time $t = 0$, yielding

$$P(n,t|0,0) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \tag{3.185}$$

a Poisson distribution with mean given by

$$\langle N(t)\rangle = \lambda t \tag{3.186}$$

(a) **Continuous Space Limit**: In contrast to the random walk, the only limit that exists is $l \to 0$, with

$$\lambda l \equiv v \tag{3.187}$$

held fixed, and the limiting characteristic function is

$$\lim_{l\to 0}\{\lambda t(e^{ils} - 1)\} \tag{3.188}$$

with the solution

$$p(x,t|0,0) = \delta(x - vt) \tag{3.189}$$

We also see that in this limit the master equation (3.182) would become the Liouville equation, whose solution would be the deterministic motion we have derived.

(b) **Asymptotic Approximation**: We can do a slightly more refined analysis. We expand the characteristic function up to second order in s in the exponent and find

$$\phi(s,t) = G(ls,t) \simeq e^{\left[t\left(ivs - \frac{s^2 D}{2}\right)\right]} \tag{3.190}$$

where, as in the previous section,

$$D = l^2\lambda \tag{3.191}$$

This is the characteristic function of a Gaussian with variance Dt and mean vt, so that we now have

$$p(x,t|0,0) \simeq \frac{1}{\sqrt{2\pi Dt}} \exp\left(-\frac{(x-vt)^2}{2Dt}\right) \tag{3.192}$$

It is also clear that this solution is the solution of

$$\partial_t p(x,t|0,0) = -v\partial_x p(x,t|0,0) + \frac{1}{2}D\partial_x^2 p(x,t|0,0) \tag{3.193}$$

which is obtained by expanding the master equation (3.182) to order l^2, by writing

$$\begin{aligned} p(n-1,t|0,0) &= \lambda p(x-l,t|0,0) \\ &\simeq \lambda p(x,t|0,0) - l\lambda\partial_x p(x,t|0,0) + \frac{1}{2}l^2\lambda\partial_x^2 p(x,t|0,0) \end{aligned} \tag{3.194}$$

However, this is an *approximation* or an *expansion* and not a limit. The limit $l \to 0$ gives Liouville's equation with the purely deterministic solution (3.189) . Effectively, the limit $l \to 0$ with well-defined v corresponds to $D = 0$.

3.8.4 The Ornstein-Uhlenbeck Process

All the examples so far had no stationary distribution, that is, as $t \to \infty$, the distribution at any finite point approaches zero and we see that, with probability one, the point moves to infinity.

If we add a linear drift term to the Wiener process, we have a Fokker-Planck equation of the form

$$\partial_t p = \partial_x(kxp) + \frac{1}{2}D\partial_x^2 p \tag{3.195}$$

where by p we mean $p(x,t|x_0,0)$. This is the *Ornstein-Uhlenbeck process.*

(a) **Characteristic Function Solution**: The equation for the characteristic function

$$\phi(s) = \int_{-\infty}^{\infty} e^{isx} p(x,t|x_0,0)dx \tag{3.196}$$

is

$$\partial_t \phi + ks\partial_s \phi = -\frac{1}{2} D s^2 \phi \tag{3.197}$$

The method of characteristics can be used to solve this equation, namely, if

$$u(s,t,\phi) = a \quad , \quad v(s,t,\phi) = b \tag{3.198}$$

are two integrals of the subsidiary equation (with a and b arbitrary constants)

$$\frac{dt}{1} = \frac{ds}{ks} = -\frac{d\phi}{\frac{1}{2} D s^2 \phi} \tag{3.199}$$

then a general solution of (3.197) is given by

$$f(u,v) = 0 \tag{3.200}$$

The particular integrals are readily found by integrating the equation involving dt and ds and that involving ds and $d\phi$; they are

$$u(s,t,\phi) = se^{-kt} \tag{3.201}$$

$$v(s,t,\phi) = \phi e^{Ds^2/4k} \tag{3.202}$$

The general solution can clearly be put in the form $v = g(u)$ with $g(u)$ an arbitrary function of u. Thus, the general solution is

$$\phi(s,t) = e^{-Ds^2/4k} g[se^{-kt}] \tag{3.203}$$

The boundary condition

$$p(x,0|x_0,0) = \delta(x - x_0) \tag{3.204}$$

clearly requires

$$\phi(s,0) = e^{ix_0 s} \tag{3.205}$$

and gives

$$g(s) = \exp\left(\frac{Ds^2}{4k} + ix_0 s\right) \tag{3.206}$$

Hence

$$\phi(s,t) = \exp\left(\frac{Ds^2}{4k}(1 - e^{-2kt}) + isx_0 e^{-kt}\right) \tag{3.207}$$

From Section 2.8.1 this corresponds to a Gaussian with

$$\langle X(t) \rangle = x_0 e^{-kt} \tag{3.208}$$

$$var[X(t)] = \frac{D}{2k}(1 - e^{-2kt}) \tag{3.209}$$

(b) **Stationary Solution**: Clearly, as $t \to \infty$, the mean and variance approach limits 0 and $D/2k$, respectively, which gives a limiting stationary solution. This solution can also be obtained directly by requiring $\partial_t p = 0$, so that p satisfies the stationary Fokker-Planck equation

$$\partial_x \left[kxp + \frac{1}{2} D \partial_x p \right] = 0 \tag{3.210}$$

and integrating once, we find

$$\left[kxp + \frac{1}{2} D \partial_x p \right]_{-\infty}^{x} = 0 \tag{3.211}$$

The requirement that p vanish at $-\infty$ together with its derivative, is necessary for normalization. Hence, we have

$$\frac{1}{p} \partial_x p = -\frac{2kx}{D} \tag{3.212}$$

so that

$$p_s(x) = \sqrt{\frac{k}{\pi D}} \exp\left(-\frac{kx^2}{D} \right) \tag{3.213}$$

This is a Gaussian with mean 0 and variance $D/2k$, as predicted from the time-dependent solution.

It is clear that a stationary solution can always be obtained for a one variable system by this integration method if such a stationary solution exists. If a stationary solution does not exist, this method gives an unnormalizable solution.

(c) **Time Correlation Functions**: The time correlation function analogous to that mentioned in connection with the Wiener process can be calculated and is a measurable piece of data in most stochastic systems. However, we have no easy way of computing it other than by the definition.

$$\langle X(t)X(s) \,|\, [x_0, t_0] \rangle = \int\int dx_1 dx_2 \, x_1 x_2 p(x_1, t; x_2, s | x_0, t_0) \tag{3.214}$$

and using the Markov property

$$= \int\int dx_1 dx_2 \, x_1 x_2 p(x_1, t | x_2, s) p(x_2, s | x_0, t_0) \tag{3.215}$$

on the assumption that

$$t \geq s \geq t_0 \tag{3.216}$$

The correlation function with a definite initial condition is not normally of as much interest as the *stationary correlation function*, which is obtained by allowing the system to approach the stationary distribution. It

is achieved by putting the initial condition in the remote past, as pointed out in Section 3.7.2. Letting $t_0 \to -\infty$, we find

$$\lim_{t_0 \to -\infty} p(x_2, s|x_0, t_0) = p_s(x_2) = \sqrt{\frac{k}{\pi D}} \exp\left(-\frac{kx_2^2}{D}\right) \tag{3.217}$$

and by straightforward substitution and integration and noting that the stationary mean is zero, we get

$$\langle X(t)X(s)\rangle_s = \langle X(t), X(s)\rangle_s = \frac{D}{2k} \exp\left(-k|t-s|\right) \tag{3.218}$$

This result demonstrates the general property of stationary processes: that the correlation functions depend only on the time differences. It is also a general result that the process we have described in this section is the only stationary Gaussian Markov process in one real variable.

The results of this section are very easily obtained by the stochastic differential equation methods which will be developed in Chapter 4.

The Ornstein-Uhlenbeck process is a simple, explicitly representable process, which has a stationary solution. In its stationary state, it is often used to model a realistic noise signal, in which $X(t)$ and $X(s)$ are only significantly correlated if

$$|t-s| \sim \frac{1}{k} \equiv \tau \tag{3.219}$$

More precisely, τ, known as the *correlation time* can be defined for arbitrary processes $X(s)$ by

$$\tau = \int_0^\infty dt\, \frac{\langle X(t), X(s)\rangle_s}{var[X]_s} \tag{3.220}$$

which is independent of the precise functional form of the correlation function.

3.8.5 Random Telegraph Process

We consider a signal $X(t)$ which can have either of two values a and b and switches from one to the other with certain probabilities per unit time. Thus, we have a master equation

$$\left.\begin{array}{l} \partial_t P(a,t|x,t_0) = -\lambda P(a,t|x,t_0) + \mu P(b,t|x,t_0) \\ \partial_t P(ab,t|x,t_0) = \lambda P(a,t|x,t_0) - \mu P(b,t|x,t_0) \end{array}\right\} \tag{3.221}$$

(a) **Time-Dependent Solutions**: These can simply be found by noting that

$$P(a,t|x,t_0) + P(b,t|x,t_0) = 1 \tag{3.222}$$

and using the initial condition

$$P(x', t_0|x, t_0) = \delta_{x,x'} \tag{3.223}$$

A simple equation can then be derived for $\lambda P(a, t|x, t_0) - \mu P(b, t|x, t_0)$, whose solution is

$$\lambda P(a, t|x, t_0) - \mu P(b, t|x, t_0) = \exp\left[-(\lambda + \mu)(t - t_0)\right](\lambda\delta_{a,x} - \mu\delta_{b,x}) \tag{3.224}$$

The solution for the probabilities then takes the form

$$\left.\begin{aligned} P(a, t|x, t_0) &= \tfrac{\mu}{\lambda+\mu} + e^{-(\lambda+\mu)(t-t_0)}\left(\tfrac{\lambda}{\lambda+\mu}\delta_{a,x} - \tfrac{\mu}{\lambda+\mu}\delta_{b,x}\right) \\ P(b, t|x, t_0) &= \tfrac{\mu}{\lambda+\mu} - e^{-(\lambda+\mu)(t-t_0)}\left(\tfrac{\lambda}{\lambda+\mu}\delta_{a,x} - \tfrac{\mu}{\lambda+\mu}\delta_{b,x}\right) \end{aligned}\right\} \tag{3.225}$$

The mean of $X(t)$ is straightforwardly computed:

$$\begin{aligned} \langle X(t) \,|\, [x_0, t_0]\rangle &= \sum xP(x, t|x_0, t_0) \\ &= \frac{a\mu + b\lambda}{\mu + \lambda} + e^{-(\lambda+\mu)(t-t_0)}\left(x_0 - \frac{a\mu + b\lambda}{\mu + \lambda}\right) \end{aligned} \tag{3.226}$$

The variance can also be computed but is a very messy expression.

(b) **Stationary Solutions**: This process has the stationary solution obtained by letting $t_0 \to -\infty$:

$$P_s(a) = \frac{\mu}{\lambda + \mu} \quad , \quad P_s(b) = \frac{\lambda}{\lambda + \mu} \tag{3.227}$$

which is obvious from the master equation.

The stationary mean and variance are

$$\langle X\rangle_s = \frac{a\mu + b\lambda}{\mu + \lambda} \tag{3.228}$$

$$var[X]_s = \frac{(a - b)^2\mu\lambda}{(\lambda + \mu)^2} \tag{3.229}$$

(c) **Stationary Correlation Functions**: To compute the stationary time correlation function, let $t \geq s$, and write

$$\begin{aligned} \langle X(t)X(s)\rangle_s &= \sum_{xx'} xx' P(x, t|x', s)P_s(x') \\ &= \sum_{x'} x' \left\langle X(t) \,|\, [x', s]\right\rangle P_s(x') \end{aligned} \tag{3.230}$$

Now use (3.226-3.229) to obtain

$$\begin{aligned}\langle X(t)X(s)\rangle = &\ angleX\rangle_s^2 + \exp\left[-(\lambda+\mu)(t-s)\right](\langle X^2\rangle_s - \langle X\rangle_s^2) \\ &= \left(\frac{a\mu+b\lambda}{\mu+\lambda}\right)^2 + \exp\left[-(\lambda+\mu)(t-s)\right]\frac{(a-b)^2\mu\lambda}{(\lambda+\mu)^2} \quad (3.231)\end{aligned}$$

Hence,

$$\langle X(t), X(s)\rangle_s = \langle X(t)X(s)\rangle_s - \langle X\rangle_s^2 = \frac{(a-b)^2\mu\lambda}{(\lambda+\mu)^2}e^{(\lambda+\mu}|t-s| \quad (3.232)$$

Notice that this time correlation function is exactly the same form as that of the Ornstein-Uhlenbeck process. Higher-order correlation functions are not the same of course, but because of this simple correlation function and the simplicity of the two state process, the random telegraph signal also finds wide application in model building.

Chapter 4

The Ito Calculus and Stochastic Differential equations

4.1 Motivation

In Section 1.2.2 we met for the first time the equation which is the prototype of what is now known as a Langevin equation, which can be described heuristically as an ordinary differential equation in which a rapidly and irregularly fluctuating random function of time [the term $X(t)$ in Langevin's original equation] occurs. The simplicity of Langevin's derivation of Einstein's results is in itself sufficient motivation to attempt to put the concept of such an equation on a reasonably precise footing.

The simple-minded Langevin equation that turns up most often can be written in the form

$$\frac{dx}{dt} = a(x,t) + b(x,t)\xi(t) \tag{4.1}$$

where x is the variable of interest, $a(x,t)$ and $b(x,t)$ are certain known functions and $\xi(t)$ is the rapidly fluctuating random term. An idealized mathematical formulation of the concept of a "rapidly varying, highly irregular function" is that for $t \neq t'$, $\xi(t)$ and $\xi(t')$ are statistically independent. We also require $\langle \xi(t) \rangle = 0$, since any nonzero mean can be absorbed into the definition of $a(x,t)$, and thus require that

$$\langle \xi(t)\xi(t') \rangle = \delta(t - t') \tag{4.2}$$

which satisfies the requirement of no correlation at different times and furthermore, has the rather pathological result that $\xi(t)$ has infinite variance. From a realistic point of view, we know that no quantity can have such an infinite variance, but the concept of *white noise* as an *idealization* of a realistic fluctuating

signal does have some meaning, and has already been mentioned in section 1.5.2 in connection with Johnson noise in electrical circuits. We have already met two sources which might be considered realistic version of almost uncorrelated noise, namely, the Ornstein-Uhlenbeck process and the random telegraph signal. For both of these the second-order correlation function can, up to a constant factor, be put into the form

$$\langle X(t), X(t')\rangle = \frac{\gamma}{2} e^{-\gamma|t-t'|} \tag{4.3}$$

Now the essential difference between these two is that the sample paths of the random telegraph signal are discontinuous, while those of the Ornstein-Uhlenbeck process are not. If (4.1) is to be regarded as a real differential equation, in which $\xi(t)$ is not white noise with a delta function correlation, but rather a noise with a finite correlation time, then the choice of a continuous function for $\xi(t)$ seems essential to make this equation realistic: we do not expect dx/dt to change discontinuously. The limit as $\gamma \to \infty$ of the correlation function (4.3) is clearly the Dirac delta function since

$$\int_{-\infty}^{\infty} \tfrac{\gamma}{2} e^{-\gamma|t-t'|} dt' = 1 \tag{4.4}$$

and for $t \neq t'$

$$\lim_{\gamma\to\infty} \frac{\gamma}{2} e^{-\gamma|t-t'|} = 0 \tag{4.5}$$

This means that a possible model of the $\xi(t)$ could be obtained by taking some kind of limit as $\gamma \to \infty$ of the Ornstein-Uhlenbeck process. This would correspond in the notation of Section 3.8.4, to the limit $k \to \infty$ with $D = k^2$.

This limit simply does not exist. Any such limit must clearly be taken after calculating measurable quantities. Such a procedure is possible but too cumbersome to use as a calculational tool.

An alternative approach is called for. Since we write the differential equation (4.1), we must expect it to be integrable and hence must expect that

$$u(t) = \int_0^t dt'\, \xi(t') \tag{4.6}$$

exists.

Suppose we now demand the ordinary property of an integral, that $u(t)$ is a continuous function of t. This implies that $u(t)$ is a Markov process since we can write

$$\begin{aligned} u(t') &= \int_0^t ds\, \xi(s) + \int_t^{t'} ds\, \xi(s) \\ &= \lim_{\epsilon\to 0} \left[\int_0^{t-\epsilon} ds\, \xi(s)\right] + \int_t^{t'} ds\, \xi(s) \end{aligned} \tag{4.7}$$

and for any $\epsilon > 0$, the $\xi(s)$ in the first integral are independent of the $\xi(s)$ in the second integral. Hence, by continuity, $u(t)$ and $u(t') - u(t)$ are statistically independent and further, $u(t') - u(t)$ is independent of $u(t'')$ for all $t'' < t$. This means that $u(t')$ is fully determined (probabilistically) from the knowledge of the value of $u(t)$ and not by any past values. Hence, $u(t)$ is a Markov process.

Since the sample functions of $u(t)$ are continuous, we must be able to describe $u(t)$ by a Fokker-Planck equation. We can compute the drift and diffusion coefficients for this process by using the formulae of Section 3.5.2. We can write

$$\langle u(t+\Delta t) - u_0 \,|\, [u_0, t] \rangle = \left\langle \int_t^{t+\Delta t} \xi(s) ds \right\rangle = 0 \tag{4.8}$$

and

$$\begin{aligned} \left\langle [u(t+\Delta t) - u_0]^2 \,\middle|\, [u_0, t] \right\rangle &= \int_t^{t+\Delta t} ds \int_t^{t+\Delta t} ds' \langle \xi(s)\xi(s') \rangle \\ &= \int_t^{t+\Delta t} ds \int_t^{t+\Delta t} ds' \delta(s - s') = \Delta t \end{aligned} \tag{4.9}$$

This means that the drift and diffusion coefficients are

$$A(u_0, t) = \lim_{\Delta t \to 0} \frac{\left\langle [u(t+\Delta t) - u_0]^2 \,\middle|\, [u_0, t] \right\rangle}{\Delta t} = 0 \tag{4.10}$$

$$B(u_0, t) = \lim_{\Delta t \to 0} \frac{\left\langle [u(t+\Delta t) - u_0]^2 \,\middle|\, [u_0, t] \right\rangle}{\Delta t} = 1 \tag{4.11}$$

The corresponding Fokker-Planck equation is that of the Wiener process and we can write

$$\int_o^t \xi(t') dt' = u(t) = W(t) \tag{4.12}$$

Thus, we have the paradox that the integral of $\xi(t)$ is $W(t)$, which is itself not differentiable, as shown in Section 3.8.1. This means that mathematically speaking, the Langevin equation (4.1) does not exist. However, the corresponding *integral equation*

$$x(t) - x(0) = \int_0^t a[x(s), s] ds + \int_0^t b[x(s), s] \xi(s) ds \tag{4.13}$$

can be interpreted consistently.

We make the replacement, which follows directly from the interpretation of the integral of $\xi(t)$ as the Wiener process $W(t)$, that

$$dW(t) \equiv W(t + dt) - W(t) = \xi(t) dt \tag{4.14}$$

and thus write the second integral as

$$\int_o^t b[x(s), s]dW(s) \tag{4.15}$$

which is a kind of Stochastic Stieltjes integral with respect to the sample function $W(t)$. Such an integral can be defined and we will carry this out in the next section.

Before doing so, it should be noted that the requirement that $u(t)$ be continuous, while very natural, can be relaxed to yield a way of defining jump processes as stochastic differential equations. This has already been hinted at in the treatment of shot noise in Section 1.5.1., However, it does not seem to be nearly so useful and will not be treated in these notes.

As a final point, we note that one normally *assumes* that $\xi(t)$ is Gaussian, and satisfies (4.2) as well. The above discussion did not require this: the Gaussian nature follows in fact from the *assumed* continuity of $u(t)$. Which of these assumptions is made is, in a strict sense, a matter of taste. However, the continuity of $u(t)$ seems a much more natural assumption to make than the Gaussian nature of $\xi(t)$, which involves in principle the determination of moments of arbitrarily high order.

4.2 Stochastic Integration

4.2.1 Definition of the Stochastic Integral

Suppose $G(t)$ is an arbitrary function of time and $W(t)$ is the Wiener process. We define the stochastic integral $\int_{t_0}^t G(t')dW(t')$ as a kind of Riemann-Stieltjes integral. Namely, we divide the interval $[t_0, t]$ into n subintervals by means of partitioning points (as in the figure below)

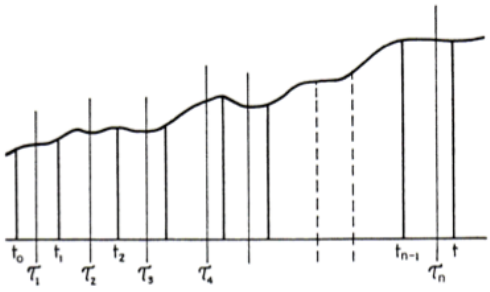

Figure 4.1: Partitioning of the time interval used in the definition of stochastic integration

with

$$t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_{n-1} \leq t \tag{4.16}$$

and define intermediate points τ_i such that

$$t_{i-1} \leq \tau_i \leq t_i \tag{4.17}$$

The stochastic integral $\int_{t_0}^{t} G(t')dW(t')$ is defined as the limit of the partial sums

$$S_n = \sum_{i=1}^{n} G(\tau_i)[W(t_i) - W(t_{i-1})] \tag{4.18}$$

It is heuristically quite easy to see that, in general, the integral defined as the limit of S_n *depends on the particular choice of intermediate pint* τ_i. For example, if we take the choice of $G(\tau_i) = W(\tau_i)$,

$$\begin{aligned} \langle S_n \rangle &= \left\langle \sum_{i=1}^{n} W(\tau_i)[W(t_i) - W(t_{i-1})] \right\rangle \\ &= \sum_{i=1}^{n} [min(\tau_,, t_i) - min(\tau_i, t_{i-1})] \\ &= \sum_{i=1}^{n} (\tau_i - t_{i-1}) \end{aligned} \tag{4.19}$$

If, for example, we choose for all i

$$\tau_i = \alpha t_i + (1-\alpha)t_{i-1} \qquad (0 < \alpha < 1) \tag{4.20}$$

then

$$\langle S_n \rangle = \sum_{i=1}^{n} (t_i - t_{i-1})\alpha = (t - t_0)\alpha \tag{4.21}$$

So that the mean value of the integral can be anything between zero and $(t-t_0)$, depending on the choice of intermediate points.

4.2.2 Ito Stochastic Integral

The choice of intermediate points characterized by $\alpha = 0$, that is by the choice

$$\tau_i = t_{i-1} \tag{4.22}$$

defines the *Ito stochastic integral* of the function $G(t)$ by

$$\int_{t_0}^{t} G(t')dW(t') = ms - \lim_{n\to\infty} \left\{ \sum_{i=1}^{n} G(t_{i-1}[W(t_i) - W(t_{i-1})] \right\} \tag{4.23}$$

By $ms - lim$ we mean the *mean square limit*, as defined in Section 2.9.2.

4.2.3 Example $\int_{t_)}^{t} W(t')dW(t')$

An exact calculation is possible. We write [writing W_i for $W(t_i)$]

$$S_n = \sum_{i=1}^{n} W_{i-1}(W_i - W_{i-1}) \equiv \sum_{i=1}^{n} W_{i-1}\Delta W_i \tag{4.24}$$

$$= \frac{1}{2}\sum_{i=1}^{n}[(W_{i-1} + \Delta W_i)^2 - (W_{i-1})^2 - (\Delta W_i)^2] \tag{4.25}$$

$$= \frac{1}{2}[W(t)^2 - W(t_0)^2] - \frac{1}{2}\sum_{i=1}^{n}\Delta W_i)^2 \tag{4.26}$$

We can calculate the mean square limit of the last term. Notice that

$$\left\langle \sum \Delta W_i^2 \right\rangle = \sum_i \langle (W_i - W_{i-1})^2 \rangle = \sum_i (t_i - t_{i-1}) = t - t_0 \tag{4.27}$$

Because of this

$$\left\langle \left[\sum_i (W_i - W_{i-1})^2 - (t - t_0)^2 \right] \right\rangle = \left\langle \sum_i (W_i - W_{i-1})^4 \right.$$
$$\left. 2\sum_{i<j}(W_i - W_{i-1})^2(W_j - W_{j-1})^2 - 2(t - t_0)\sum_i (W_i - W_{i-1})^2 + (t - t_0)^4 \right\rangle \tag{4.28}$$

Notice that $W_i - W_{i-1}$ is a Gaussian variable and is independent of $W_j - W_{j-1}$. hence, we can factorize. Thus

$$\langle (W_i - W_{i-1})^2 (W_j - W_{j-1})^2 \rangle = (t_i - t_{i-1})(t_j - t_{j-1}) \tag{4.29}$$

and also using formula (2.82) for the fourth moment of a Gaussian variable

$$\langle (W_i - W_{i-1})^4 \rangle = 3\langle (W_i - W_{i-1})^2 \rangle^2 = 3(t_i - t_{i-1})^2 \tag{4.30}$$

which combined with (4.29) gives

$$\left\langle \left[\sum_i (W_i - W_{i-1})^2 - (t - t_0)^2 \right] \right\rangle$$
$$= 2\sum_i (t_i - t_{i-1})^2 + \sum_{i,j}[(t_i - t_{i-1} - (t - t_0)][(t_j - t_{j-1} - (t - t_0)]$$
$$= 2\sum_i (t_i - t_{i-1})^2 \to 0 \text{ as } n \to \infty \tag{4.31}$$

Thus,

$$ms - \lim_{n\to\infty} \sum_i (W_i - W_{i-1})^2 = t - t_0 \tag{4.32}$$

by definition of the mean square limit, so

$$\boxed{\int_{t_0}^{t} W(t')dW(t') = \frac{1}{2}[W(t)^2 - W(t_0)^2 - (t - t_0)]} \tag{4.33}$$

Comments

(i)

$$\left\langle \int_{t_0}^{t} W(t')dW(t') \right\rangle = \frac{1}{2}[\langle W(t)^2\rangle - \langle W(t_0)^2\rangle - (t - t_0)] = 0 \tag{4.34}$$

This is also obvious by definition, since the individual terms are $(W_{i-1}\Delta W_i)$, which vanishes because ΔW_i is statistically independent of W_{i-1}, as was demonstrated in Section 3.8.1.

(ii) The result for the integral is no longer the same as for the ordinary Riemann-Stieltjes integral in which the term $(t - t_0)$ would be absent. The reason for this is that $[W(t + \Delta t) - W(t)]$ is almost always of order $\sqrt{t}$, so that in contrast to ordinary integration, terms of second-order in $\Delta W(t)$ do not vanish on taking the limit.

4.2.4 The Stratonovich Integral

An alternative definition was introduced by Stratonovich as a stochastic integral in which the anomalous term $(t-t_0)$ does not occur. We define this fully in Section 4.4 - in the cases considered so far, it amounts to evaluating the integrand as a function of $W(t)$ at the value $\frac{1}{2}[W(t_i) + W(t_{i-1})]$. It is straightforward to show that

$$\begin{aligned}(S)\int_{t_0}^{t} W(t')dW(t') &= \underset{n\to\infty}{ms-lim}\sum_{i=1}^{n}\frac{W(t_i)+W(t_{i-1})}{2}[W(t_i)-W(t_{i-1})] \\ &= \frac{1}{2}[W(t)^2 - W(t_0)^2]\end{aligned} \tag{4.35}$$

However, the integral defined by Stratonovich [which we will always designate by a prefixed (S) as in (4.35)] has no general relationship with that defined by Ito. That is, for *arbitrary* functions $G(t)$, there is no connection between the two integrals. In the case, however, where we can specify that $G(t)$ is related to some stochastic differential equation, a formula can be given relating one to the other, see Section 4.4.

4.2.5 Nonanticipating Functions

The concept of a nonanticipating function can be easily made quite obscure by complex notation, but it is really quite simple. We have in mind a situation in which all functions can be expressed as functions or functionals of a certain

Wiener process $W(t)$ through the mechanism of a stochastic differential (or integral) equation of the form

$$x(t) - x(0) = \int_{t_0}^{t} a[x(t'), t']dt' + \int_{t_0}^{t} b[x(t'), t']dW(t') \tag{4.36}$$

A function $G(T)$ is called a *nonanticipating function of* t if $G(t)$ is statistically independent of $W(s) - W(t)$ for all s and b such that $t < s$. This means that $G(t)$ is independent of the behavior of the Wiener process in the future of t. This is clearly a rather reasonable requirement for a physical function which could be the solution of an equation like (4.36) in which it seems heuristically obvious that $x(t)$ involves $W(t')$ only for $t' \leq t$.

For example, specific non anticipating functions of t are:

(i)
$$W(t)$$

(ii)
$$\int_{t_0}^{t} F[W(t')]dt'$$

(iii)
$$\int_{t_0}^{t} F[W(t')]dW(t')$$

(iv)
$$\int_{t_0}^{t} G(t')dt' \tag{4.37}$$

(v)
$$\int_{t_0}^{t} G(t')dW(t') \tag{4.38}$$

where in the last two items $G(t)$ itself isa non anticipating function. Results (iii) and (v) depend on the fact that the Ito stochastic integral, as defined in (), is the limit of the sequence in which only $G(t')$ for $t' < t$ and $W(t')$ for $t' \leq t$ are involved

The reasons for considering nonanticipating functions specifically are:

(i) Many results can be derived, which are only true for such functions.

(ii) They occur naturally in situations, such as in the study of differential equations involving time, in which some kind of *causality* is expected in the sense that the unknown future cannot affect the present.

(iii) The definition of stochastic differential equations requires such functions.

4.2.6 Proof that $dW(t)^2 = dt$ and $dW(t)^{2+N} = 0$

The formulae in the heading are the key to the use of the Ito calculus as an ordinary computational tool. However, as write they are not very precise and what is really meant is that for an arbitrary *nonanticipating function* $G(t)$

$$\int_{t_0}^{t} [dW(t')]^{2+N} G(t') \equiv \underset{n\to\infty}{ms-lim} \sum_i G_{i-1} \Delta W_i^{2+N} = \begin{cases} \int_{t_0}^{t} dt'\, G(t') & \text{for } N = 0 \\ 0 & \text{for } N > 0 \end{cases} \tag{4.39}$$

The proof is quite straightforward. For $N = 0$, let us define

$$I = \lim_{n\to\infty} \left\langle \left[\sum_i G_{i-1} (\Delta W_i^2 - \Delta t_i) \right]^2 \right\rangle \tag{4.40}$$

$$= \lim_{n\to\infty} \left\{ \left\langle \underbrace{(G_{i-1})^2} \underbrace{(\Delta W_i^2 - \Delta t_i)^2} + \sum_{i>j} \underbrace{2 G_{i-1} G_{j-1} (\Delta W_j^2 - \Delta t_j)^2} \underbrace{(\Delta W_i^2 - \Delta t_i)^2} \right\rangle \right\} \tag{4.41}$$

The horizontal braces indicate factors which are statistically independent of each other because of the properties of the Wiener process, and because the G_i are values of a nonanticipating function which are independent of all ΔW_i for $j > i$.

Using this independence, we can factorize the means, and also using

(i) $\langle \Delta W_i^2 \rangle = \Delta t_i$

(ii) $\langle (\Delta W_i^2 - \Delta t_i)^2 \rangle = 2\Delta t_i^2$ (from Gaussian nature of ΔW_i)

we find

$$I = 2 \lim_{n\to\infty} \left[\sum_i \Delta t_i^2 \langle (G_{i-1})^2 \rangle \right] \tag{4.42}$$

Under reasonably mild conditions on $G(t)$ (e.g., boundedness), this means that

$$\underset{n\to\infty}{ms-lim} \left(\sum_i G_{i-1} \Delta W_i^2 - \sum_i G_{i-1} \Delta t_i \right) = 0 \tag{4.43}$$

and since

$$\underset{n\to\infty}{ms-lim} \sum_i G_{i-1} \Delta t_i = \int_{t_0}^{t} G(t') \tag{4.44}$$

we have

$$\boxed{\int_{t_0}^{t} [dW(t')]^2 G(t') = \int_{t_0}^{t} G(t')} \tag{4.45}$$

Comments

(i) The proof $\int_{t_0}^{t} G(t)[dW(t)]^{2+N} = 0$ for $N > 0$ is similar and uses the explicit expressions for the higher moments of a Gaussian given in Section 2.8.1.

(ii) $dW(t)$ only occurs in integrals so that when we restrict ourselves to nonanticipating functions, we can simply write

$$dW(t)^2 \equiv dt \tag{4.46}$$

$$dW(t)^{2+N} \equiv 0 \quad (N > 0) \tag{4.47}$$

(iii) The results are only valid for the Ito integral, since we have used the fact that ΔW_i is independent of G_{i-1}. In the Stratonovich integral,

$$\Delta W_i = W(t_i) - W(t_{i-1}) \tag{4.48}$$

$$G_{i-1} = G\left(\tfrac{1}{2}(t_i + t_{i-1})\right) \tag{4.49}$$

and although $G(T)$ is nonanticipating, this is not sufficient to guarantee the independence of ΔW_i and G_{i-1} as thus defined.

(iv) By similar methods one can prove that

$$\int_{t_0}^{t} G(t')\, dt'\, dW(t') \equiv \underset{n\to\infty}{ms-lim} \sum G_{i-i} \Delta W_i \Delta t_i = 0 \tag{4.50}$$

and similarly for higher powers. The simplest way of characterizing these results is to say that $dW(t)$ is an infinitesimal order of $\frac{1}{2}$ and that in calculating differentials, infinitesimals of higher order than 1 are discarded.

4.2.7 Properties of the Ito Stochastic Integral

(a) **Existence**: One can show that the Ito stochastic integral $\int_{t_0}^{t} G(t')dW(t')$ exists whenever the function $G(t')$ is *continuous* and *nonanticipating* on the classed interval $[t_0, t]$.

(b) **Integration of Polynomials**: We can formally use the result of Section 4.2.6:

$$d[W(t)]^n = [W(t) + dW(t)]^n - W(t)^n = \sum_{r=1}^{n} \binom{n}{r} W(t)^{n-r} dW(t)^r \tag{4.51}$$

and using the fact that $dW(t)^r \to 0$ for all $r > 2$,

$$= nW(t)^{n-1} dW(t) + \frac{n(n-1)}{2} W(t)^{n-2} dt \tag{4.52}$$

so that

$$\boxed{\int_{t_0}^{t} W(t')^n dW(t') = \frac{1}{n+1}[W(t)^{n+1} - W(t_0)^{n+1}] - \frac{n}{2}\int_{t_0}^{t} W(t')^{n-1} dt'} \tag{4.53}$$

(c) **Two Kinds of Integral**: We note that for each $G(t)$ there are two kinds of integrals, namely,

$$\int_{t_0}^{t} G(t')dt' \quad \text{and} \quad \int_{t_0}^{t} G(t')dW(t') \tag{4.54}$$

both of which occur in the previous equation. There is, in general, no connection between these two kinds of integral.

(d) **General Differentiation Rules**: In forming differentials, as in (b) above, one must keep all terms up to second order in $dW(t)$. This means that, for example,

$$d\{\exp[W(t)]\} = \exp[W(t) + dW(t)] - \exp[W(t)] \tag{4.55}$$

$$= \exp[W(t)]\left[dW(t) + \frac{1}{2}dW(t)^2\right] \tag{4.56}$$

$$= \exp[W(t)]\left[dW(t) + \frac{1}{2}dt\right] \tag{4.57}$$

For an arbitrary function

$$df[W(t),t] = \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial f}{\partial W}dW(t) + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}[dW(t)]^2 + \frac{\partial^2 f}{\partial W \partial t}dt\, dW(t) + ... \tag{4.58}$$

and we use

$$[dW(t)]^2 \to dt \tag{4.59}$$

$$dt\, dW(t) \to 0 \quad \text{Section 4.2.6, comment (iv)}] \tag{4.60}$$

$$(dt)^2 \to 0 \tag{4.61}$$

and all higher powers, to arrive at

$$\boxed{df[W(t),t] = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right)dt + \frac{\partial f}{\partial W}dW(t)} \tag{4.62}$$

(e) **Mean Value Formula**: For nonanticipating $G(t)$.

$$\boxed{\left\langle \int_{t_0}^{t} G(t')dW(t') \right\rangle = 0} \tag{4.63}$$

Proof: Since $G(t)$ is nonanticipating, in the definition of the stochastic integral,

$$\left\langle \sum_i G_{i-1}\Delta W_i \right\rangle = \sum_i \langle G_{i-1}\rangle\langle \Delta W_i\rangle = 0 \tag{4.64}$$

We know from Section 2.9.5 that operations of $ms - lm$ and $\langle\ \rangle$ may be interchanged. Hence, taking the limit of (4.64), we have the result.

This result is *not true* for Stratonovich's integral, since the value of G_{i-1} is chosen in the middle of the interval, and may be correlated with ΔW_i.

(f) **Correlation Formula**: Ig $G(t)$ and $H(t)$ are arbitrary continuous nonanticipating functions,

$$\boxed{\left\langle \int_{t_0}^{t} G(t')dW(t') \int_{t_0}^{t} H(t')dW(t') \right\rangle = \int_{t_0}^{t} \langle G(t')H(t')\rangle dt'} \tag{4.65}$$

Proof: Notice that

$$\left\langle \sum_i G_{i-1}\Delta W_i \sum_j H_{j-1}\Delta W_j \right\rangle$$
$$\left\langle \sum_i G_{i-1}H_{i-1}(\Delta W_i)^2 \right\rangle + \left\langle \sum_{i>j} (G_{i-1}H_{j-1} + G_{j-1}H_{i-1})\Delta W_i \Delta W_j \right\rangle \tag{4.66}$$

In the second term, ΔW_i is independent of all other terms since $j < i$, and G and H are nonanticipating. Hence, we may factorize out the term $\langle \Delta W_i\rangle = 0$ so that this term vanishes. Using

$$\langle \Delta W_i^2\rangle = \Delta t_i \tag{4.67}$$

and interchanging mean and limit operations, the result follows.

(g) **Relation to Delta-Correlated White Noise**: Formally, this is equivalent to the idea that Langevin terms $\xi(t)$ are delta correlated and uncorrelated with $F(t)$ and $G(t)$. For, rewriting

$$dW(t) \to \xi(t)dt \tag{4.68}$$

it is clear that if $F(t)$ and $G(t)$ are nonanticipating, $\xi(t)$ is independent of them, and we get

$$\int_{t_0}^{t} dt' \int_{t_0}^{t} ds' \langle G(t')H(s')\xi(t')\xi(s')\rangle = \int_{t_0}^{t}\int_{t_0}^{t} dt'\, ds'\, \langle G(t')H(s')\rangle\langle \xi(t')\xi(s')\rangle \tag{4.69}$$

which implies

$$\langle \xi(t)\xi(s)\rangle = \delta(t-s) \tag{4.70}$$

An important point of definition arises here, however. In integrals involving delay functions, it frequently occurs in the study of stochastic differential equations that the argument of the delay function is equal to either the upper or lower limit of the integral, that is, we find integrals like

$$I_1 = \int_{t_1}^{t_2} dt\, f(t)\delta(t-t_1) \tag{4.71}$$

or

$$I_2 = \int_{t_1}^{t_2} dt\, f(t)\delta(t-t_2) \tag{4.72}$$

Various conventions can be made concerning the value of such integrals. We will show that in the present context, we must always make the interpretation

$$I_1 = f(t_1) \tag{4.73}$$

$$I_2 = 0 \tag{4.74}$$

corresponding to counting all the weight of the delay fundtion at the lower limit of an integral, and none of the weight at the upper limit. To demonstrate this, note that

$$\left\langle \int_{t_0}^{t} G(t')dW(t') \left[\int_{t_0}^{t} H(s')dW(s') \right] \right\rangle = 0 \tag{4.75}$$

This follows, since the function defined by the integral inside the square bracket is, by Section 4.2.5 comment (v), a nonanticipating function and hence the complete integrand, [obtained by multiplying by $G(t')$ which is also nonanticipating] is itself nonanticipating. Hence the average vanishes by the result of Section4.27e.

Now using the formulation in terms of the Langevin source $\xi(t)$, we can rewrite (4.75) as

$$\int_{t_0}^{t}\int_{t_0}^{t} \langle G(t')H(s')\rangle \delta(t'-s') \tag{4.76}$$

which corresponds to not counting the weight of the delta function at the upper limit. Consequently, the full weight must be counted at the lower limit.

This property is a direct consequence of the definition of the Ito integral as in (4.23), in which the increment points "towards the future". That is, we can interpret

$$dW(t) = W(t+dt) - W(t) \tag{4.77}$$

In the case of the Stratonovich integral, we get quite a different formula, which is by no means as simple to prove as in the Ito case, but which amounts to choosing

$$\left.\begin{aligned} I_1 &= \tfrac{1}{2} f(t_1) \\ I_2 &= \tfrac{1}{2} f(t_2) \end{aligned}\right\} \quad \text{(Stratonovich)} \tag{4.78}$$

This means that in both cases, the delta function occurring at the limit of an integral has half its weight counted. This formula, although intuitively more satisfying than the Ito form, is more complicated to use, especially in the perturbation theory of stochastic differential equations, where the Ito method makes very many terms vanish.

4.3 Stochastic Differential equations (SDE)

We concluded in Section 4.1, that the most satisfactory interpretation of the Langevin equation

$$\frac{dx}{dt} = a(x,t) + b(x,t)\xi(t) \tag{4.79}$$

is a stochastic differential equation

$$x(t) - x(0) = \int_0^t dt'\, a[x(t'),t'] + \int_0^t dW(t')\, b[x(t'),t'] \tag{4.80}$$

Unfortunately, the kind of stochastic integral to be used is not given by the reasoning of Section 4.1. The Ito integral is mathematically and technically the most satisfactory, but it is not always the most natural choice physically. The Stratonovich integral is the natural choice for an interpretation which assumes $\xi(t)$ is real noise(not white noise) with finite correlation time, which is then allowed to become infinitesimally small after calculating measurable quantities. Furthermore, a Stratonovich interpretation enables us to use ordinary calculus, which is not possible for an Ito interpretation.

From a mathematical point of view, the choice is made clear by the near impossibility of carrying out proofs using the Stratonovich integral. We will therefore define the Ito SDE, develop its equivalence with the Straonovich SDE, and use either form depending on the circumstances.

4.3.1 Ito Stochastic Differential equation: Definition

A stochastic quantity $x(t)$ obeys an Ito SDE written as

$$dx(t) = a[x(t),t]dt + b[x(t),t]dW(t) \tag{4.81}$$

if for all t and t_0,

$$x(t) - x(0) = \int_0^t a[x(t'),t']dt' + \int_0^t b[x(t'),t']dW(t') \tag{4.82}$$

Before considering what conditions must be satisfied by the coefficients in (4.82), it is wise to consider what one means by a solution of such an equation and what the uniqueness of the solution would mean in this context. For this purpose, we can consider a discretized version of the SDE obtained by taking a mesh of points (as illustrated in the figure below) such that

$$t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = t \tag{4.83}$$

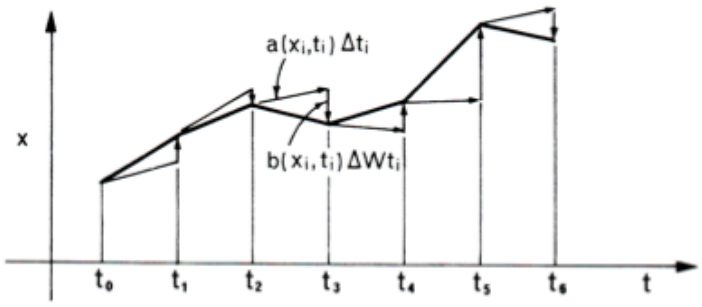

Figure 4.2: Illustration of the Cauchy-Euler procedure for constructing an approximate solution of the stochastic differential equation $dx(t) = a[x(t), t]dt + b[x(t), t]dW(t)$

and writing the equation as

$$x_{i+1} = x_i + a(x_i, t_i)\Delta t_i + b(x_i, t_i)\Delta W_i \tag{4.84}$$

Here,

$$\left.\begin{array}{c} x_i = x(t_i) \\ \Delta t_i = t_{i+1} - t_i \\ \Delta W_i = W(t_{i+1}) - W(t_i) \end{array}\right\} \tag{4.85}$$

(a) **Cauchy-Euler Construction of the Solution of an Ito SDE**: We see from (4.84) that an approximate procedure for solving the equation is to calculate x_{i+1} from the knowledge of x_i by adding a deterministic term

$$a(x_i, t_i)\Delta t_i \tag{4.86}$$

and a stochastic term

$$b(x_i, t_i)\Delta W_i \tag{4.87}$$

The stochastic term contains an element ΔW_i, which is the increment of the Wiener process, but is statistically independent of x_i if

(i) x_0 is itself independent of all $W(t) - W(t_0)$ for $t > t_)$ (thus, the initial conditions if considered random, must be nonanticipating), and

(ii) $a(x,t)$ is a nonanticipating function of t for any fixed x.

Constructing an approximate solution iteratively by use of (4.84), we see that x_i is always independent of ΔW_j for $j \geq i$.

The solution is then formally constructed by letting the mesh size go to zero. To say that the solution is unique means that for a given sample function $\widetilde{W}(t)$ of the random Wiener process $W(t)$, the particular solution of the equation which arises is unique. To say the solution exists means that with probability one, a solution exists for any choice of sample function $\widetilde{W}(t)$ of the Wiener process $W(t)$.

(b) **Existence and Uniqueness of Solutions of an Ito SDE**: Existence and uniqueness will not be proved here. The conditions which are required for existence and uniqueness in a time interval $[t_0, T]$ are:

(i) *Lipschitz condition*: a K exists such that

$$|a(x, t - a(y,t)| + |b(x,t) - b(y,t)| \leq K|x - y| \tag{4.88}$$

for all x and y, and all t in the range $[t_0, T]$.

(ii) *Growth condition*: a K exists such that for all t in the range $[t_0, T]$,

$$|a(x,t)|^2 + |b(x,t)|^2 \leq K^2(1 + |x|^2) \tag{4.89}$$

Under these conditions there will be a unique nonanticpating solution $x(t)$ in the range $[t_0, T]$.

Almost every stochastic differential equation encountered in practice satisfies the Lipschitz condition since it is essentially a smoothness condition. However, the growth condition is often violated. This does not mean that no solution exists, rather, it means the solution may "explode" to infinity, that is, the value of x can become infinite in a finite time; in practice, a finite random time. This phenomenon occurs in ordinary differential equations, for example,

$$\frac{dx}{dt} = \frac{1}{2}ax^3 \tag{4.90}$$

has the general solution with an initial condition $x = x_0$ at $t = 0$

$$x(t) = \frac{1}{\sqrt{-at + \frac{1}{x_0^2}}} \tag{4.91}$$

If a is positive, this becomes infinite when $x_0 = (at)^{-1/2}$, but if a is negative, the solution never explodes. Failing to satisfy the Lipschitz condition does not guarantee the solution will explode.

4.3.2 Dependence on Initial Conditions and Parameters

In exactly the same way as in the case of deterministic differential equation, if the functions which occur in a stochastic differential equation depend continuously on parameters, then the solution normally depends continuously on that parameter. Similarly, the solution depends continuously on initial conditions. Let us formulate this more precisely. Consider a one-variable equation

$$dx = a(\lambda, x, t)dt + b(\lambda, x, t)dW(t) \tag{4.92}$$

with initial condition

$$x(t_0) = c(\lambda) \tag{4.93}$$

where λ is a parameter. Let the solution of (4.92) be $x(\lambda, t)$. Suppose

(i)
$$st - \lim_{\lambda \to \lambda_0} c(\lambda) = c(\lambda_0) \tag{4.94}$$

(ii) For every $N > 0$

$$\lim_{\lambda \to \lambda_0} \left\{ \sup_{t \in [t_0, T], |x| < N} [|a(\lambda, x, t) - a(\lambda_0, x, t)| + |b(\lambda, x, t) - b(\lambda_0, x, t)|] \right\} = 0 \tag{4.95}$$

(iii) There exists a K independent of λ such that

$$|a(\lambda, x, t)|^2 + |b(\lambda, x, t)|^2 \leq K^2(1 + |x|^2) \tag{4.96}$$

Then,

$$st - \lim_{\lambda \to \lambda_0} \left\{ \sup_{t \in [t_0, T]} |x(\lambda, t) - x(\lambda_0, t)| \right\} = 0 \tag{4.97}$$

Comments

(i) Recalling the definition of stochastic limit, the interpretation of the limit (4.97) is that as $\lambda \to \lambda_0$, the probability that the maximum deviation over any finite interval $[t_0, T]$ between $x(\lambda, t)$ and $x(\lambda_0, t)$ is nonzero, goes to zero.

(ii) Dependence on the initial condition is achieved by letting a and b be independent of λ.

(iii) The result will be very useful in justifying perturbation expansions.

(iv) Condition (ii) is written in the most natural form for the case that the functions $a(x, t)$ and $b(x, t)$ are not themselves stochastic. It often arises that $a(x, t)$ and $b(x, t)$ are themselves stochastic (nonanticipating) functions. In this case, condition (ii) must be replaced by a probabilistic statement. It is, in fact, sufficient to replace $\lim_{\lambda \to \lambda_0}$ by $st - \lim_{\lambda \to \lambda_0}$.

4.3.3 Markov Property of the Solution of an Ito SDE

We now show that $x(t)$, the solution to the stochastic differential equation (4.82), is a Markov process. Heuristically, the result is obvious, since with a given initial condition $x(t_0)$, the future time development is uniquely (stochastically) determined, that is, $x(t)$ for $t > t_0$ is determined only by

(i) The particular sample path of $W(t)$ for $t > t_0$

(ii) The value of $x(t_0)$

Since $x(t)$ is a nonanticipating function of t, $W(t)$ for $t > t_0$ is independent of $x(t)$ for $t < t_0$. Thus, the time development of $x(t)$ for $t > t_0$ is independent of $x(t)$ for $t < t_0$ provided $x(t_0)$ is known. Hence, $x(t)$ is a Markov process.

4.3.4 Change of variables: Ito's Formula

Consider and arbitrary function of $x(t) : f[x(t)]$. What stochastic differential equation does it obey? We use the results of Section 4.2.6 to expand $df[x(t)]$ to second order in $dW(t)$:

$$df[x(t)] = f[x(t) + dx(t)] - f[x(t)] \tag{4.98}$$

$$= f'[x(t)]dx(t) + \tfrac{1}{2}f''[x(t)]dx(t)^2 + \cdots \tag{4.99}$$

$$= f'[x(t)]\{a[x(t),t]dt + b[x(t),t]dW(t)\} + \tfrac{1}{2}f''[x(t)]b[x(t),t]^2 dW(t)^2 \tag{4.100}$$

where all other terms have been discarded since they are higher order. Now use $dW(t)^2 = dt$ to obtain

$$\boxed{\begin{aligned} df[x(t)] = \{a[x(t),t]f'[x(t)] + \tfrac{1}{2}b[(x(t),t]^2 f''[x(t)]\}dt \\ + b[x(t),t]f'[x(t)]dW(t) \end{aligned}} \tag{4.101}$$

This formula is known as Ito's formula and shows that changing variables is not given by ordinary calculus unless $f[x(t)]$ is merely linear in $x(t)$.

Many Variables: In practice, Ito's formula becomes very complicated and the easiest method is to simply use the multivariate form of the rule that $dW(t)$ is and infinitesimal of order $\frac{1}{2}$. By similar methods to those used in Section 4.2.6, we can show that for an n dimensional Wiener process $W(t)$,

$$dW_i(t)dW_j(t) = \delta_{ij}dt \tag{4.102a}$$

$$dW_i(t)^{N-2} = 0 \quad (N > 0) \tag{4.102b}$$

$$dW_i(t)dt = 0 \tag{4.102c}$$

$$dt^{1+N} = 0 \quad (N > 0) \tag{4.102d}$$

which imply that $dW_i(t)$ is an infinitesimal number of order $\frac{1}{2}$. Note, however, that (4.102a) is a consequence of the independence of $dW_i(t)$ and $dW_j(t)$. To

develop Ito's formula for functions of an n dimensional vector $\mathbf{x}(t)$ satisfying the stochastic differential equation

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{W}(t) \tag{4.103}$$

we simply follow this procedure. The result is

$$\boxed{\begin{aligned} df[\mathbf{x}] = \Bigg\{ & \sum_i A_i(\mathbf{x}, t)\partial_i f(\mathbf{x}) \\ & + \tfrac{1}{2}\sum_{i,j} \left[\mathbf{B}(\mathbf{x}, t)\mathbf{B}^T(\mathbf{x}, t)\right]_{ij} \partial_i\partial_j f(\mathbf{x}) \Bigg\} dt \\ & + \sum_{i,j} B_{ij}(\mathbf{x}, t)\partial_i f(\mathbf{x}) dW_j(t) \end{aligned}} \tag{4.104}$$

4.3.5 Connection Between Fokker-Planck Equation and Stochastic Differential Equation

(a) **Forward Fokker-Planck Equation**: We now consider the time development of an arbitrary $f(x(t))$. Using Ito's formula

$$\begin{aligned} \frac{\langle df[x(t)]\rangle}{dt} &= \left\langle \frac{df[x(t)]}{dt} \right\rangle = \frac{d}{dt}\langle f[x(t)]\rangle \\ &= \left\langle a[x(t), t]\partial_x f + \tfrac{1}{2}b[x(t), t]^2\partial_x^2 f \right\rangle \end{aligned} \tag{4.105}$$

However, $x(t)$ has a conditional probability density $p(x, t|x_0, t_0)$ and

$$\begin{aligned} \frac{d}{dt}\langle f[x(t)]\rangle &= \int dx\, f(x)\partial_t p(x, t|x_0, t_0) \\ &= \int dx\, \left[a(x, t)\partial_x f + \tfrac{1}{2}b(x, t)^2\partial_x^2 f\right] p(x, t|x_0, t_0) \end{aligned} \tag{4.106}$$

This is now of the same form as (3.36) of Section 3.4.1. Under the same conditions as there, we integrate by parts and discard the surface terms to obtain

$$\int dx\, f(x)\partial_t p = \int dx\, f(x) \left\{-\partial_x[a(x, t)p] + \tfrac{1}{2}\partial_x^2[b(x, t)^2 p]\right\} \tag{4.107}$$

and hence, since $f(x)$ is arbitrary,

$$\boxed{\partial_t p(x, t|x_0, t_0) = \partial_x[a(x, t)p(x, t|x_0, t_0)] + \tfrac{1}{2}\partial_x^2[b(x, t)^2 p(x, t|x_0, t_0)]} \tag{4.108}$$

We have thus a complete equivalence to a diffusion process defined by a drift co0efficient $a(x, t)$ and a diffusion coefficient $b(x, t)^2$.

The results are precisely analogous to Section 3.5.2, in which it was shown that the diffusion process could be locally approximated by an equation resembling an Ito stochastic differential equation.

(b) **Backward Fokker-Planck Equation - the Feynman-Kac Formula**: Suppose a function $g(x,t)$ obeys the backward Fokker-Planck equation

$$\partial_t g = -a(x,t)\partial_g - \tfrac{1}{2}b(x,t)^2\partial_x^2 g \tag{4.109}$$

with the initial condition

$$g(x,T) = G(x) \tag{4.110}$$

If $x(t)$ obeys the stochastic differential equation (4.81), then using Ito's rule (adapted appropriately to account for explicit time dependence), the function $g[x(t),t]$ obeys the stochastic differential equation

$$\begin{aligned} dg[x(t),t] = \left\{\partial_t g = a[(x(t),t]\partial_x g g[(x(t),t] + \tfrac{1}{2}b[(x(t),t]^2\partial_x^2 g[(x(t),t]\right\} dt \\ + b[(x(t),t]\partial_x g[(x(t),t]dW(t) \end{aligned} \tag{4.111}$$

and using (4.109) this becomes

$$dg[(x(t),t] = b[(x(t),t]\partial_x g[(x(t),t]dW(t) \tag{4.112}$$

Now integrate from t to T, and take the mean

$$\langle g[x(T),T]\rangle - \langle g[x(t),t]\rangle = \left\langle \int_t^T b[x(t'),t']\partial_x g[x(t'),t']dW(t') \right\rangle = 0 \tag{4.113}$$

Let the initial condition of the stochastic differential equation for $x(t')$ and $t' = t$ be

$$x(t) = x \tag{4.114}$$

where x is a non-stochastic value, so that

$$\langle g[x(t),t]\rangle = g(x,t) \tag{4.115}$$

At the other end of the interval, use the final condition (4.110) to write

$$\langle g[x(T),T]\rangle = \langle G[x(T)]|x(t) = x\rangle \tag{4.116}$$

where the notation on the right hand side indicates the mean conditional on the initial condition (4.114).

Putting these two together, the *Feynman-Kac* formula results:

$$\langle G[x(T)]|x(t) = x\rangle = g(x,t) \tag{4.117}$$

where $g(x,t)$ is the solution of the backward Fokker-Planck equation (4.109) with initial condition (4.110).

This formula is essentially equivalent to the fact that $p(x,t|x_0,t_0)$ obeys the backward Fokker-Planck equation in the arguments x_0, t_0 as shown in Section 3.6, since

$$\langle G[x(T)]|x(t_0)=x_0\rangle = \int dx\, G(x)p(x,T|x_0,t_0) \tag{4.118}$$

4.3.6 Multivariable Systems

In general, many variable systems of stochastic differential equations can be defined for n variables by

$$d\mathbf{x} = \mathbf{A}(\mathbf{x},t)dt\mathbf{B}(\mathbf{x},t)d\mathbf{W}(t) \tag{4.119}$$

where $d\mathbf{W}(t)$ is an n variable Wiener process, as defined in Section 3.8.1. The many variable version of the reasoning used in Section 4.3.5 shows that the Fokker-Planck equation for the conditional probability density $p(\mathbf{x},t|x_0,t_0)\equiv p$ is

$$\partial_t p = \sum_i \partial_t[A_i(\mathbf{x},t)p] + \tfrac{1}{2}\sum_{i,j}\partial_i\partial_j\left\{\left[\mathbf{B}(\mathbf{x},t)\mathbf{B}^T(\mathbf{x},t)\right]\right\} \tag{4.120}$$

Notice that the same Fokker-Planck equation arises from all matrices $\mathbf{B}$ such that $\mathbf{B}\mathbf{B}^T$ is the same. This means that we can obtain the same Fokker-Planck equation by replacing $\mathbf{B}$ by $\mathbf{B}\mathbf{S}$ where $\mathbf{S}$ is orthogonal, i.e., $\mathbf{S}\mathbf{S}^T = 1$. Notice that $\mathbf{S}$ may depend on $\mathbf{x}(t)$.

This can be seen directly from the stochastic differential equation. Suppose $\mathbf{S}(t)$ is an orthogonal matrix with an arbitrary *nonanticipating* dependence on t. Then define

$$d\mathbf{V}(t) = \mathbf{S}(t)d\mathbf{W}(t) \tag{4.121}$$

Now the vector $d\mathbf{V}(t)$ is a linear combination of Gaussian variables $d\mathbf{W}(t)$ with coefficients $\mathbf{S}(t)$ which are independent of $d\mathbf{W}(t)$, since $\mathbf{S}(t)$ is nonanticipating. For any fixed value of $\mathbf{S}(t)$, the $d\mathbf{V}(t)$ are thus Gaussian and their correlation matrix is

$$\begin{aligned}\langle dV_i(t)dV_j(t)\rangle &= \sum l,m S_{il}(t)S_{jm}(t)\langle dW_l(t)dW_m(t)\rangle \\ &= \sum_l S_{il}(t)S_{jl}(t)dt = \delta_{ij}dt\end{aligned} \tag{4.122}$$

since $\mathbf{S}(t)$ is orthogonal. Hence, all the moments are independent of $\mathbf{S}(t)$ and are the same as those of $d\mathbf{W}(t)$, so $d\mathbf{V}(t)$ is itself Gaussian with the same correlation matrix as $d\mathbf{W}(t)$. Finally, averages at different times factorize, for example, if $t > t'$ in

$$\sum_{i,k}\langle[dW_i(t)S_{ij}(t)]^m[dW_k(t')S_{kl}(t')]^n\rangle \tag{4.123}$$

we can factorize out the averages of $dW_i(t)$ to various powers since $dW_i(t)$ is independent of all other terms. Evaluating these we will find that the orthogonal

nature of $\mathbf{S}(t)$ gives, after averaging over $dW_i(t)$, simply

$$\sum_k \langle [dW_j(t)]^m \rangle \langle [dW_k(t')S_{kl}(t')]^n \rangle \tag{4.124}$$

which similarly gives $\langle [dW_j(t)]^m [dW_l(t')]^n \rangle$. Hence, the $d\mathbf{V}(t)$ are also increments of a Wiener process. The orthogonal transformation simply mixes up different sample paths of the process, without changing its stochastic nature.

Hence, instead of (4.119) we can write

$$\begin{aligned} d\mathbf{x} &= \mathbf{A}(\mathbf{x},t)dt\mathbf{B}(\mathbf{x},t)\mathbf{S}^T(t)\mathbf{S}(t)d\mathbf{W}(t) \\ &= \mathbf{A}(\mathbf{x},t)dt\mathbf{B}(\mathbf{x},t)\mathbf{S}^T(t)d\mathbf{V}(t) \end{aligned} \tag{4.125}$$

and since $\mathbf{V}(t)$ is itself simply a Wiener process, this equation is equivalent to

$$d\mathbf{x} = \mathbf{A}(\mathbf{x},t)dt\mathbf{B}(\mathbf{x},t)\mathbf{S}^T(t)d\mathbf{W}(t) \tag{4.126}$$

which is exactly the same Fokker-Planck equation (4.120).

We will return to some examples in which this identity is relevant in Section 4.5.5.

4.4 The Stratonovich Stochastic Integral

The Stratonovich stochastic integral is an alternative to the Ito definition, in which Ito's formula, developed in Section 4.3.4, is replaced by the ordinary chain rule for change of variables. This apparent advantage does not come without cost, since in Stratonovich's definition the independence of a non-anticipating integrand $G(t)$ and the increment $dW(t)$ in a stochastic integral no longer holds. This means that increment and the integral are correlated, and therefore to give a full definition of the Stratonovich integral requires some way of specifying what this correlation is.

The correlation is implicitly specified in the situation of most interest, the case in which the integrand is a function whose stochastic nature arises from a dependence on a variable $x(t)$ which obeys a stochastic differential equation. Since the aim is to recover the chain rule for change of variables in a stochastic differential equation this seems a reasonable restriction.

4.4.1 Definition of the Stratonovich Stochastic Integral

Stratonovich defined a stochastic integral of an integrand which is a function of $x(t)$ and t by

$$(S)\int_{t_0}^{t} G[x(t'),t']dW(t') = \underset{n\to\infty}{ms-lim}\sum_{i=1}^{n} G\left\{\tfrac{1}{2}\left(x(t_i)+x(t_{i-1}),t_{i-1}\right)\right\}\left[W(t_i)-W(t_{i-1})\right] \tag{4.127}$$

The Stratonovich integral is clearly *related* to a mid-loint choice of τ_i in the definition of stochastic integration as given in Section 4.2.1, but clearly is *not* necessarily equivalent to that definition. Rather, instead of everlasting x at the midpoint $\frac{1}{2}(t_i+t_{i-1})$, the average of the values at the two time points is taken. Furthermore it is only the dependence on $x(t)$ that is averaged in this way, and not the explicit dependence on t. However, if $G(z,t)$ is differentiable in t, the integral can be shown to be independent of the particular choice of value for t in the range $[t_{i-1},t_i]$.

4.4.2 Stratonovich Stochastic Differential Equation

It is possible to write a stochastic differential equation (SDE) using Stratonovich's integral

$$x(t)=x(t_0)+\int_{t_0}^{t} dt'\,\alpha[x(t'),t']+(S)\int_{t_0}^{t} dt'\,dW(t')\beta[x(t'),t'] \tag{4.128}$$

(a) **Change of Variables for the Straonovich SDE**: The definition of the Stratonovich integral is such as to make the ordinary rules of calculus valid for change of variables. This means, that for the Stratonovich integral, Ito's formula is replaced by the simple calculus rule

$$\boxed{(S)\,df[x(t)]-f'[x(t)]\{a[x(t),t]dt+b[x(t),t]dW(t)\}} \tag{4.129}$$

This can be proved quite simply from the definition (4.127). The essence of the proof can be explained by using the simple SDE

$$x_{i+1}=x_i+B\left[\tfrac{1}{2}(x_{i+1}+x_i)\right](W_{i+1}-W_i) \tag{4.130}$$

To find the Stratonovich SDE for $f[x(t)]$, we need only use the Taylor series expansion of a function about a midpoint in the form

$$f(x+1)=f(x-a)+\sum_{n=0}^{\infty}\frac{f^{2n+1}(x)a^{2n+1}}{(2n+1)!} \tag{4.131}$$

In expanding $f(x_{i+1})$ we only need to keep terms up to second order, so we drop all but the first two terms and write

$$\begin{aligned} f(x_{i+1}) &= f(x_i) + f'\left[\tfrac{1}{2}(x_{i+1}+x_i)\right](x_{i+1}-x_i) \\ &= f'\left[\tfrac{1}{2}(x_{i+1}+x_i)\right] B\left[\tfrac{1}{2}(x_{i+1}+x_i)\right](W_{i+1}-W_i) \end{aligned} \tag{4.132}$$

This means that the Stratonovich SDE for $f[x(t)]$ is

$$(S)\, df[x(t)] = f'[x(t)]B[x(t)]dW(t) \tag{4.133}$$

which is the ordinary calculus rule. The extension the the general case (4.128) is straightforward.

(b) **Equivalent to Ito SDE**: We will show that the Stratonovich SDE is in fact equivalent to an appropriate Ito SDE. Let us assume that $x(t)$ is a solution of the Ito SDE

$$dx(t) = a[x(t),t]dtb[x(t),t]dW(t) \tag{4.134}$$

and deduce the α and β for a corresponding Stratonovich equation of the form (4.128). In both cases, the solution $x(t)$ is the same function.

We first compute the connection between the Ito integral $\int_{t_0}^{t} dW(t')\, b[x(t'),t']$ and the Stratonovich integral $(S)\int_{t_0}^{t} dW(t')\,\beta[x(t'),t']$:

$$(S)\int_{t_0}^{t} dW(t')\,\beta[x(t'),t'] \simeq \sum_i \beta\left[\tfrac{1}{2}(x(t_i)+x(t_{i-1})),t_{i-1}\right]\Delta W(t_{i-1}) \tag{4.135}$$

In (4.135) we write

$$x(t_i) = x(t_{i-1}) + \Delta x(t_{i-1}) \tag{4.136}$$

and use the Ito SDE (4.134) to write

$$\Delta x(t_i) = a[x(t_{i-1}),t_{i-1}]\Delta t_{i-1} + b[x(t_{i-1}),t_{i-1}]\Delta W(t_{i-1}) \tag{4.137}$$

Then applying Ito's formula, we can write

$$\begin{aligned} \beta\left[\tfrac{1}{2}(x(t_i)+x(t_{i-1})),t_{i-1}\right] &= \beta\left[x(t_{i-1})+\tfrac{1}{2}\Delta x(t_{i-1}),t_{i-1}\right] \\ &= \beta(t_{i-1}) + \left[a(t_{i-1})\partial_x\beta(t_{i-1}) + \tfrac{1}{4}b^2(t_{i-1})\right]\frac{1}{2}\Delta t_{i-1} \\ &\quad \tfrac{1}{2}b(t_{i-1})\partial_x\beta(t_{i-1})\Delta W(t_{i-1}) \end{aligned} \tag{4.138}$$

(For simplicity, we write $\beta(t_i)$ etc, instead of $\beta[x(t_i),t_i]$ wherever possible). Putting all these back in the original equation (4.134) and dropping as usual dt^2, $dtdW(t)$, and setting $dW(t)^2 = dt$, we find

$$(S)\int = \sum_i \beta(t_{i-1})\{W(t_i)-W(t_{i-1})\} + \tfrac{1}{2}\sum_i b[x(t'),t']\partial_x\beta(t_{i-1})(t_i-t_{i-1}) \tag{4.139}$$

Hence we derive

$$\boxed{\begin{aligned}(S)\int_{t_0}^{t}\beta[x(t'),t']dW(t') &= \int_{t_0}^{t}\beta[x(t'),t']dW(t') \\ &\quad + \tfrac{1}{2}\int_{t_0}^{t} b[x(t'),t']\partial_x\beta[x(t'),t']dt'\end{aligned}} \tag{4.140}$$

This formula gives a connection between the Ito and Stratonovich integrals of functions $\beta[x(t'),t']$, in which $x(t')$ is the solution of the Ito SDE (4.128). It does not give a general connection between the Ito and Stratonovich integrals of arbitrary functions.

If we now make the choice

$$\begin{aligned}\alpha(x,t) &= a(x,t) - \tfrac{1}{2}b(x,t)\partial_x b(x,t) \\ \beta(x,t) &= b(x,t)\end{aligned} \tag{4.141}$$

we see that: the Ito SDE:

$$\boxed{dx = adt + bdW(t)} \tag{4.142}$$

is the same as the Stratonovich SDE

$$\boxed{(S)\,dx = \left(a - \tfrac{1}{2}b\partial_x b\right)dt + bdW(t)} \tag{4.143}$$

and conversely, the Stratononvich SDE

$$\boxed{(S)\,dx = \alpha dt + \beta dW(t)} \tag{4.144}$$

is the same as the Ito SDE

$$\boxed{dx = \left(\alpha + \tfrac{1}{2}\beta\partial_x\beta\right)dt + \beta dW(t)} \tag{4.145}$$

(c) **ManyVariables**: If a many variable Ito equation is

$$d\mathbf{x} = \mathbf{A}(\mathbf{x},t) + \mathbf{B}(\mathbf{x},t)d\mathbf{W}(t) \tag{4.146}$$

then the corresponding Stratonovich equation can be shown similarly to be given by replacing

$$\begin{aligned}A_i^s &= A_i - \tfrac{1}{2}\sum_{j,k} B_{kj}\partial_k B_{ij} \\ \mathbf{B}^s &= \mathbf{B}\end{aligned} \tag{4.147}$$

(d) **Fokker-Planck Equation**: Corresponding to the Stratonovich SDE,

$$(s)\,d\mathbf{x} = \mathbf{A}^s(\mathbf{x},t) + \mathbf{B}^s(\mathbf{x},t)d\mathbf{W}(t) \tag{4.148}$$

we can, by use of (4.147) and the known correspondence (Section 4.3.6) between the Ito stochastic differential equation and the Fokker-Planck equation, show that the equivalent Fokker-Planck equation is

$$\boxed{\partial_t p = -\sum_i \partial_i \{A_i^s p\} + \tfrac{1}{2} \sum_{i,j,k} \partial_, \{B_{ik}^s \partial_j [B_{jk}^s p]\}} \tag{4.149}$$

which is often known as the "Stratonovich form" of the Fokker-Planck equation. In contrast to the two forms of the stochastic differential equation, the two forms of the Fokker-Planck equation have a different appearance but are (of course) interpreted with the same rules - those of ordinary calculus.

4.5 Some Examples and Solutions

4.5.1 Coefficients without x Dependence

The simple equation

$$dx = a(t)dt + b(t)dW(t) \tag{4.150}$$

with (t) and $b(t)$ nonrandom functions of time, is solved by integrating

$$x(t) = x_0 + \int_{t_0}^{t} a(t')dt' + \int_{t_0}^{t} b(t')dW(t') \tag{4.151}$$

Here, x_0 can be either a nonrandom initial condition or may be random, but must be independent of $W(t) - W(t_0)$ for $t > t_0$; otherwise, $x(t)$ is not nonanticipating.

As constructed, $x(t)$ is Gaussian, provided x_0 is either nonrandom or itself gaussian, since

$$\int_{t_0}^{t} b(t')dW(t') \tag{4.152}$$

is simply a linear combination of infinitesimal Gaussian variables. Further,

$$\langle x(t) \rangle = \langle x_0 \rangle + \int_{t_0}^{t} a(t')dt' \tag{4.153}$$

(since the mean of the Ito integral vanishes), and

$$\begin{aligned} \langle [x(t) - \langle x(t) \rangle][x(s) - \langle x(s) \rangle] \rangle &\equiv \langle x(t), x(s) \rangle \\ &= var[x_0] + \left\langle \int_{t_0}^{t} b(t')dW(t') \int_{t_0}^{s} b(s')dW(s') \right\rangle \\ &= var[x_0] + \int_{t_0}^{min(t,s)} [b(t')]^2 dt' \end{aligned} \tag{4.154}$$

where we have used the result (4.65) with, however

$$G(t') = \begin{cases} b(t') & t' < t \\ 0 & t' \geq t \end{cases} \tag{4.155}$$

$$H(t') = \begin{cases} b(t') & t' < s \\ 0 & t' \geq s \end{cases} \tag{4.156}$$

The process is thus completely determined.

4.5.2 Multiplicative Linear White Noise Process - Geometric Brownian Motion

The equation

$$dx = cx dW(t) \tag{4.157}$$

is known as a multiplicative white noise process because it is linear in x, but the "noise term" $dW(t)$ multiplies x. It is also commonly known as *geometric Brownian motion*.

We can solve this exactly by using Ito's formula. Let us define a new variable by

$$y = \log x \tag{4.158}$$

so that

$$dy = \frac{1}{x}dx - \frac{1}{2x^2}(dx)^2 = c dW(t) - \frac{1}{2}c^2 dt \tag{4.159}$$

This equation can now be directly integrated, so we obtain

$$y(t) = y(t_0) + c[W(t) - W(t_0)] - \frac{1}{2}c^2(t - t_0) \tag{4.160}$$

and hence

$$x(t) = x(t_0) \exp\left\{c[W(t) - W(t_0)] - \tfrac{1}{2}c^2(t - t_0)\right\} \tag{4.161}$$

(a) **Mean value**: We can calculate the mean by using the formula for any Gaussian variable z with zero mean

$$\langle \exp z \rangle \exp\left(\tfrac{1}{2}\langle z^2 \rangle\right) \tag{4.162}$$

so that

$$\langle x(t) \rangle = \langle x(t_0) \rangle \exp\left[\tfrac{1}{2}c^2(t - t_0) - \tfrac{1}{2}c^2(t - t_0)\right] = \langle x(t_0) \rangle \tag{4.163}$$

The result is also obvious from definition, since

$$d\langle x(t) \rangle = \langle dx(t) \rangle = \langle c x(t) dW(t) \rangle \tag{4.164}$$

(b) **Autocorrelation Function**: We can also calculate the autocorrelation function

$$\begin{aligned}\langle x(t)x(s)\rangle &= \langle x(t_0)\rangle^2 \left\langle \exp c[W(t)+W(s)=W(t_0)] - \tfrac{1}{2}c^2(t+s-2t_0)\right\rangle \\ &= \langle x(t_0)\rangle^2 \exp \tfrac{1}{2}c^2[\langle [W(t)+W(s)-2W(t_0)]^2\rangle - (t+s-2t_0)] \\ &= \langle x(t_0)\rangle^2 \exp \tfrac{1}{2}c^2[t+s-4t_0-2\,min(t,s)-(t+s-2t_0)] \\ &= \langle x(t_0)\rangle^2 \exp [c^2 min(t-t_0,s-t_0) \end{aligned} \tag{4.165}$$

(c) **Stratonovich Interpretation**: The solution of this equation interpreted as a Stratonovich equation can also be obtained, but ordinary calculus would then be valid. Thus, instead of (4.159) we would obtain

$$(S)\,dy = cdW(t) \tag{4.166}$$

and hence

$$x(t) = x(t_0)\exp c[W(t)-W(t_0)] \tag{4.167}$$

In this case,

$$\langle x(t)\rangle = \langle x(t_0)\rangle \exp\left[\tfrac{1}{2}c^2(t-t_0)\right] \tag{4.168}$$

and

$$\langle x(t)x(s)\rangle = \langle x(t_0)\rangle^2 \exp\left\{\tfrac{1}{2}c^2[t+s-t_0+2\,min(t-t_0,s-t_0)]\right\} \tag{4.169}$$

One sees that there is a clear difference between these two answers.

4.5.3 Complex Oscillator with Noisy frequency

This a simplification of a model due to Kubo and is a slight generalization of the previous example for complex variables. We consider

$$\frac{dz}{dt} = i\left(\omega + \sqrt{2\gamma}\xi(t)\right)z \tag{4.170}$$

which formally represents a simple model of an oscillator with a mean frequency ω perturbed by a noise term $\xi(t)$.

Physically, this is best modeled by writing a Stratonovich equation

$$(S)\,dz = i\left(\omega dt + \sqrt{2\gamma}dW(t)\right)z \tag{4.171}$$

which is equivalent to the Ito equation (from Section 4.4)

$$dz = [(i\omega-\gamma)t]\langle z(0)\rangle \tag{4.172}$$

Taking the mean value, we see immediately that

$$\frac{d\langle z(t)\rangle}{dt} = (i\omega-\gamma)\langle z\rangle \tag{4.173}$$

with the *damped* oscillatory solution

$$\langle z(t)\rangle = \exp\left[(i\omega - \gamma)t\right]\langle z(0)\rangle \tag{4.174}$$

One can show fully why the Stratonovich model is more appropriate. The most obvious way to see this is to note that $\xi(t)$ would, in practice, be somewhat smoother than a white noise and ordinary calculus would apply, as in the case of the Stratonovich interpretation.

Now in this case, the correlation function obtained from solving the original Stratonovich equation is

$$\langle z(t)z(s)\rangle = \langle z(0)^2\rangle \exp\left[(i\omega - \gamma)(t+s) - 2\gamma\, min(t,s)\right] \tag{4.175}$$

In the limit $t.s \to \infty$, with $t + \tau = s$,

$$\lim_{t\to infty} \langle x(t+\tau)z(t)\rangle = 0 \tag{4.176}$$

However, the correlation function of physical interest is the complex correlations

$$\begin{aligned}\langle z(t)z^*(s)\rangle &= \langle |z(0)|^2\rangle\langle \exp\left\{i\omega(t-s) + i\sqrt{2\gamma}[W(t) - W(s)]\right\}\rangle \\ &= \langle |z(0)|^2 \exp\left\{i\omega(t-s) - \gamma[t+s-2\,min(t,s)]\right\}\rangle \\ &= \langle |z(0)|^2\rangle \exp\left[i\omega(t-s) - \gamma|t-s|\right]\end{aligned} \tag{4.177}$$

Thus, the complex correlation function has a damping term which arises purely from the noise. It may be thought of as a noise induced dephasing effect, whereby the phases of an ensemble of initial states with initial phases diffuse away from the value ωt arising from the deterministic motion, as illustrated in the figure below.

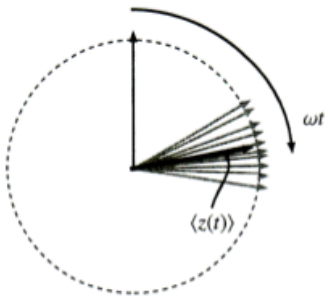

Figure 4.3: Illustration of the decay of the mean aptitude of a complex oscillator as a result of dephasing.

The mean of the ensemble consequently decays, although the amplitude $|z(t)|$ of any member of the ensemble is unchanged. For large time differences, $z(t)$ and $z^*(s)$ become independent.

A realistic oscillator cannot be described by this model of a complex oscillator. However, the qualitative behavior is very similar, and this model may be regarded as a prototype model of oscillators with noisy frequency.

4.5.4 Ornstein-Uhlenbeck Process

Taking the Fokker-Planck equation given for the Ornstein-Uhlenbeck process in Section 3.8.4, we can immediately write down the SDE using the result of Section 4.3.5:

$$dx = -kxdt + \sqrt{D}dW(t) \tag{4.178}$$

and solve this directly. Putting

$$y = x\, e^{kt} \tag{4.179}$$

then

$$\begin{aligned} dy &= (dx)d(e^{kt}) + (dx)e^{kt} + xd(e^{kt}) \\ &= [-kxdt + \sqrt{D}dW(t)]ke^{kt}dt + [-kxdt + \sqrt{D}dW(t)]e^{kt} + kxe^{kt}dt \end{aligned} \tag{4.180}$$

We note that the first product vanishes, involving only dt^2, and $dW(t)dt$ (in fact, it can be seen that this will always happen if we simply multiply x by a deterministic function of time). We get

$$dy = \sqrt{D}e^{kt}dW(t) \tag{4.181}$$

so that integrating and resubstituting for y, we get

$$x(t) = x(0)e^{-kt} + \sqrt{D}\int_0^t e^{-k(t-t')}dW(t') \tag{4.182}$$

If the initial condition is deterministic or Gaussian distributed, then $x(t)$ is clearly Gaussian, with mean and variance

$$\langle x(t)\rangle = \langle x(0)\rangle e^{-kt} \tag{4.183}$$

$$var[x(t)] = \left\langle \left\{ [x(0) - \langle x(0)\rangle]e^{-kt} + \sqrt{D}\int_0^t e^{-k(t-t')}dW(t') \right\}^2 \right\rangle \tag{4.184}$$

Taking the initial condition to be nonanticipating, that is, independent of $dW(t)$ for $t > 0$, we can write using the result of Section 4.4f

$$\begin{aligned} var[x(t)] &= var[x(0)]e^{-2kt} + D\int_0^t e^{-2k(t-t')}dt' \\ &= \left(var[x(0)] - \frac{D}{2k}\right)e^{-2kt} + \frac{D}{2k} \end{aligned} \tag{4.185}$$

These equations are the same as those obtained directly by solving the Fokker-Planck equation in Section 3.8.4, with the added generalization of a nonanticipating random initial condition. Added to the fact that the solution is a Gaussian variable, we also have the correct conditional probability.

The time correlation function can also be calculated directly and is,

$$\begin{aligned}\langle x(t), x(s)\rangle &= var[x(0)]e^{-k(t+s)} + D\left\langle \int_0^t e^{-k(t-t')}dW(t') \int_0^s e^{-k(s-s')}dW(s')\right\rangle \\ &= var[x(0)]e^{-k(t+s)} + D\int_0^{min(t,s)} e^{-k(t+s-2t')}dt' \\ &= \left[var[x(0)] - \frac{D}{2k}\right] e^{-k(t+s)} + \frac{D}{2k}e^{-k|t-s|} \end{aligned} \tag{4.186}$$

Notice that if $k > 0$, as $t, s \to \infty$ with finite $|t - s|$, the correlation function becomes stationary and the form deduced in Section 3.8.4.

In fact, if we set the initial time at $-\infty$ rather than 0, the solution (4.182) becomes

$$x(t) = \sqrt{D}\int_{-\infty}^t e^{-k(t-t')}dW(t') \tag{4.187}$$

in which the correction function and the mean obviously assume their stationery values. Since the process is Gaussian, this makes it stationary.

4.5.5 Conversion from Cartesian to Polar Coordinates

A model often used to describe an optical field is given by a pair of Ornstein-Uhlenbeck processes describing the real and imaginary components of the electric field, i.e.,

$$dE_1(t) = -\gamma E_1(t)dt + \epsilon dW_1(t) \tag{4.188}$$

$$dE_2(t) = -\gamma E_2(t)dt + \epsilon dW_2(t) \tag{4.189}$$

It is of interest to convert to polar coordinates. We set

$$E_1(t) = a(t)\cos\phi(t) \tag{4.190}$$

$$E_2(t) = a(t)\sin\phi(t) \tag{4.191}$$

and for simplicity, also define

$$\mu(t) = \log a(t) \tag{4.192}$$

so that

$$\mu(t) + i\phi(t) = \log\left[E_1(t) + iE_2(t)\right] \tag{4.193}$$

We then use the Ito calculus to derive

$$\begin{aligned} d(\mu + i\phi) &= \frac{d(E_1(t) + iE_2(t))}{E_1(t) + iE_2(t)} - \frac{[d(E_1(t) + iE_2(t))]^2}{2(E_1(t) + iE_2(t))^2} \\ &= -\frac{\gamma(E_1(t) + iE_2(t))}{E_1(t) + iE_2(t)}dt + \frac{\epsilon(dW_1(t) + idW_2(t))}{E_1(t) + iE_2(t)} \\ &\qquad - \frac{\epsilon^2[dW_1(t) + idW_2(t)]^2}{2(E_1(t) + iE_2(t))^2} \end{aligned} \tag{4.194}$$

and noting $dW_1(t)dW_2(t) = 0$, and $dW_1(t)^2 = dW_2(t)^2 = dt$, it can be seen that the last term vanishes, so we find

$$d[\mu(t) + i\phi(t)] = -\gamma dt + \epsilon \exp[-\mu(t) - i\phi(t)]\{dW_1(t) + idW_2(t))\} \tag{4.195}$$

We now take the real part, set $a(t) = \exp[\mu(t)]$ and using the Ito calculus find

$$da(t) = \left(-\gamma a(t) + \frac{\epsilon^2}{2a(t)}\right)dt + \epsilon(dW_1(t)\cos\phi(t) + dW_2(t)\sin\phi(t)) \tag{4.196}$$

The imaginary part yields

$$d\phi(t) = \frac{\epsilon}{a(t)}(-dW_1(t)\cos\phi(t) + dW_2(t)\sin\phi(t)) \tag{4.197}$$

We now define

$$\left.\begin{aligned} dW_a(t) &= dW_1(t)\cos\phi(t) + dW_2(t)\sin\phi(t) \\ dW_\phi(t) &= -dW_1(t)\cos\phi(t) + dW_2(t)\sin\phi(t) \end{aligned}\right\} \tag{4.198}$$

We note that this is an orthogonal transformation of the kind mentioned in Section 4.3.6, so that we may take $dW_a(t)$ and $dW_\phi(t)$ as increments of independent Wiener processes $W_a(t)$ and $W_\phi(t)$.

Hence, the stochastic differential equations for phase and amplitude are

$$d\phi(t) = \frac{\epsilon}{a(t)}dW_\phi(t) \tag{4.199}$$

$$da(t) = \left(-\gamma a(t) + \frac{\epsilon^2}{2a(t)}\right)dt + \epsilon dW_a(t) \tag{4.200}$$

Comment. Using the rules given in Section 4.4 (ii), it is possible to convert both the Cartesian equations (4.188, 4.189) and the polar equations (4.199, 4.200) to the Stratonovich form, and find that both are exactly the same as the Ito form. Nevertheless, a direct conversion using ordinary calculus is not possible. Doing so would get the same result until (4.195) where the term $[\epsilon^2/2a(t)]dt$ would not be found. This must be compensated by an extra term which arises from the fact that the Stratonovich increments $dW_i(t)$ are *correlated* with $\phi(t)$ and thus, $dW_a(t)$ and $dW_\phi(t)$ cannot simply be defined by (4.197). We

see the advantage of the Ito method which retains the statistical independence of $d\mathbf{W}(t)$ and variables evaluated at time t.

Unfortunately, the equations in Polar form are not soluble, as the corresponding Cartesian equations are. There is advantage, however, in dealing with polar equations in the laser, whose equations are similar, but have an added term proportional to $a(t)^2dt$ in (4.200).

4.5.6 Multivariate Ornstein-Uhlenbeck Process

We define the process by the stochastic differential equation

$$d\mathbf{x}(t) = -A\mathbf{x}(t)dt + Bd\mathbf{W}(t) \tag{4.201}$$

(A and B are constant matrices) for which the solution is easily obtained (as in Section 4.5.4):

$$\mathbf{x}(t) = \exp(-At)\mathbf{x}(0) + \int_0^t \exp[-A(t-t')]BdW(t') \tag{4.202}$$

The mean is

$$\langle \mathbf{x}(t)\rangle = \exp(-At)\langle \mathbf{x}(0)\rangle \tag{4.203}$$

The correlation function follows similarly

$$\begin{aligned}\langle \mathbf{x}(t), \mathbf{x}^T(s)\rangle &\equiv \langle [\mathbf{x}(t) - \langle \mathbf{x}(t)\rangle][\mathbf{x}(s) - \langle \mathbf{x}(s)\rangle]^T\rangle \\ &= \exp(-At)\langle \mathbf{x}(0), \mathbf{x}^T(0)\rangle \exp(-As) \\ &\quad + \int_0^{min(t,s)} \exp[-A(t-t')]BB^T \exp[-A^T(s-s)]dt'\end{aligned} \tag{4.204}$$

The integral can be explicitly evaluated in certain special cases, and for particular low-dimensional problems, it is possible to simply multiply everything out term by ten. In the remainder, we set $\langle \mathbf{x}(0), \mathbf{x}^T(0)\rangle = 0$, corresponding to a deterministic initial condition, and evaluate a few special cases.

(a) **The Case $\mathbf{AA}^T = \mathbf{A}^T\mathbf{A}$**: In this case (for real A) we can find a unitary matrix S such that

$$\begin{aligned}&SS^\dagger = 1 \\ &SAS^\dagger = SA^TS^\dagger = diag(\lambda_1, \lambda_2, ..., \lambda_n)\end{aligned} \tag{4.205}$$

For simplicity, assume $t \geq s$. Then

$$\langle \mathbf{x}(t), \mathbf{x}^T(s)\rangle = S^\dagger G(t,s)S \tag{4.206}$$

where

$$[G(t,s)]_{ij} = \frac{(BB^T)_{ij}}{\lambda_i + \lambda_j}[\exp(-\lambda_i|t-s|) - \exp(-\lambda_i t - \lambda_j s)] \tag{4.207}$$

(b) **Stationary Variance**: If A has only eigenvalues with positive real part, a stationary solution exists of the form

$$\mathbf{x}_s(t) = \int_{-\infty}^{t} \exp\left[-A(t-t')\right] B dW(t') \tag{4.208}$$

We have of course

$$\langle \mathbf{x}_s(t) \rangle = 0 \tag{4.209}$$

$$\langle \mathbf{x}_s(t), \mathbf{x}_s^T(s) \rangle = \int_0^{min(t,s)} \exp\left[-A(t-t')\right] BB^T \exp\left[-A^T(s-s')\right] dt' \tag{4.210}$$

Let us define the stationary covariance matrix σ by

$$\sigma = \langle \mathbf{x}_s(t), \mathbf{x}_s^T(t) \rangle \tag{4.211}$$

This can be evaluated by means of an algebraic equation thus:

$$\begin{aligned} A\sigma + \sigma A^T &= \int_{-\infty}^{t} A \exp\left[-A(t-t')\right] BB^T \exp\left[-A^T(t-t')\right] dt' \\ &\quad + \int_{-\infty}^{t} \exp\left[-A(t-t')\right] BB^T \exp\left[-A^T(t-t')\right] A^T dt' \\ &= \int_{-\infty}^{t} \frac{d}{dt'} \{\exp\left[-A(t-t')\right] BB^T \exp\left[-A^T(t-t')\right]\} dt' \end{aligned} \tag{4.212}$$

Carrying out the integral, we find that the lower limit vanishes by the assumed positivity of the eigenvalues of A and hence only the upper limit remains, giving

$$\boxed{A\sigma + \sigma A^T = BB^T} \tag{4.213}$$

as and algebraic equation for the stationary covariance matrix.

(c) **Stationary Variance for Two Dimensions**: We note that if A is a 2×2 matrix, it satisfies the characteristic equation

$$A^2 - (TrA)A + (\det A) = 0 \tag{4.214}$$

and form (4.209) and the fact that (4.214) implies $\exp(-At)$ is a polynomial of degree 1 in A, we must be able to write

$$\sigma = \alpha BB^T + \beta(ABB^T + BB^T A^T) + \gamma ABB^T A^T \tag{4.215}$$

Using (4.214), we find (4.213) is satisfied if

$$\alpha + (TrA)\beta - (\det A)\gamma = 0 \tag{4.216}$$

$$2\beta(\det A) + 1 = 0 \tag{4.217}$$

$$\beta + (TrA)\gamma = 0 \tag{4.218}$$

From which we have

$$\boxed{\sigma = \frac{(\det A)BB^T + [A - (TrA)BB^T[A - (TrA)]^T}{2(TrA)(\det A)}} \tag{4.219}$$

(d) **Time Correlation Matrix in the Stationary State**: From the solution of (4.209), we see that if $t > s$,

$$\begin{aligned}\langle \mathbf{x}_s(t), \mathbf{x}_s^T(s)\rangle &= \int_{-\infty}^{s} \exp[-A(t-t')]BB^T \exp[-A^T(t-t')]dt' \\ &= \exp[-A(t-s)]\sigma \quad , \quad t > s \end{aligned} \tag{4.220}$$

and similarly,

$$= \sigma \exp[-A^T(s-t)] \quad , \quad t < s \tag{4.221}$$

This depends only on $s-t$, as expected of a stationary solution. Defining then

$$G_s(t-s) = \langle \mathbf{x}_s(t), \mathbf{x}_s^T(s)\rangle \tag{4.222}$$

we see (remembering $\sigma = \sigma^T$) that

$$G_s(t-s) = [G_s(s-t)]^T \tag{4.223}$$

(e) **Spectrum Matrix in Stationary State**: The spectrum matrix turns out to be rather simple. We define similarly to Section 1.5.2:

$$\begin{aligned} S(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau} G_s(\tau) d\tau \\ &= \frac{1}{2\pi}\left\{\int_0^{\infty} \exp[-(i\omega + A)\tau]\sigma d\tau + \int_{-\infty}^{0} \sigma \exp[(-i\omega + A^T)\tau] d\tau\right\} \\ &= \frac{1}{2\pi}[(A + i\omega)^{-1})\sigma + \sigma(A^T - i\omega)^{-1}] \end{aligned} \tag{4.224}$$

Hence,

$$(A + i\omega)S(\omega)(A^T - i\omega) = \frac{1}{2\pi}(\sigma A^T + A\sigma) \tag{4.225}$$

and using (4.213) we get

$$\boxed{S(\omega) = \frac{1}{2\pi}(A + i\omega)^{-1}BB^T(A^T - i\omega)^{-1}} \tag{4.226}$$

(f) **Regression Theorem**: The result (4.220) is also known as a regression theorem in that it states that the time development $G_s(\tau)$ is for $\tau > 0$ governed by the same law of time development as the mean, as in (4.203).

It is a consequence of the Markovian linear nature of the problem. The time derivative of the stationary correlation function is

$$\begin{aligned}\frac{d}{d\tau}[G_s(\tau)]d\tau &= \frac{d}{d\tau}\langle \mathbf{x}_s(\tau), \mathbf{x}_s^T(0)\rangle d\tau \\ &= \langle [-A\mathbf{x}_s(\tau)d\tau + Bd\mathbf{W}(\tau)], \mathbf{x}_s^T(0)\rangle \end{aligned} \tag{4.227}$$

Since $\tau > 0$, the increment $d\mathbf{W}(\tau)$ is uncorrelated with $\mathbf{x}_s^T(0)$, this means that

$$\boxed{\frac{d}{d\tau}[G_s(\tau)] = -AG_s(\tau)} \tag{4.228}$$

Thus, computation of $G_s(\tau)$ requires the knowledge of $G_s(0) = \sigma$ and the time development equation of the mean. This result is similar to those of Section 3.7.4.

4.5.7 The General Single Variable Linear Equation

(a) **Homogeneous Case**: We consider firstly the homogeneous case

$$dx = [b(t)dt + g(t)dW(t)]x \tag{4.229}$$

and using the usual Ito rules, write

$$y = \log x \tag{4.230}$$

so that

$$dy = \frac{dx}{x} - \frac{dx^2}{2x^2} = b(t)dt + g(t)dW(t) - \tfrac{1}{2}g(t^2)dt \tag{4.231}$$

and integrating and inverting (4.230), we get

$$\begin{aligned} x(t) &= x(0)\exp\left\{\int_0^t \left[b(t') - \tfrac{1}{2}g(t')^2\right]dt' + \int_0^t g(t')dW(t')\right\} \\ &\equiv x(0)\phi(t) \end{aligned} \tag{4.232}$$

which serves to define $\phi(t)$.

We note that [using (4.162)

$$\begin{aligned}\langle [x(t)]^n\rangle &= \langle [x(0)]^n\rangle \left\langle \exp\left\{n\int_0^t \left[b(t') - \tfrac{1}{2}g(t')^2\right]dt' + n\int_0^t g(t')dW(t')\right\}\right\rangle \\ &= \langle [x(0)]^n\rangle \exp\left\{n\int_0^t b(t')dt' + \tfrac{1}{2}n(n-1)\int_0^t g(t')^2dt'\right\} \end{aligned} \tag{4.233}$$

(b) **Inhomogeneous Case**: Now consider

$$dx = [a(t) + b(t)x]dt + [f(t) + g(t)x]dW(t) \tag{4.234}$$

and write

$$z(t) = x(t)[\phi(t)]^{-1} \tag{4.235}$$

with $\phi(t)$ as defined in (4.232) and a solution of the inhomogeneous equation (4.229). Then we write

$$dz = dx[\phi(t)]^{-1} + xd[\phi(t)^{-1}] + dxd[\phi(t)^{-1}] \tag{4.236}$$

Noting that $d[\phi(t)^{-1}] = -d\phi(t)[\phi(t)]^{-2} + [d\phi(t)]^2[\phi(t)]^{-3}$ and using Ito rules, we find

$$dz = \{[a(t) - f(t)g(t)]dt + f(t)dW(t)\}\phi(t)^{-1} \tag{4.237}$$

which is directly integrable. Hence, the solution is

$$x(t) = \phi(t)\left\{x(0) + \int_0^t \phi(t')^{-1}\{[a(t') - f(t')g(t')]dt' + f(t')dW(t')\right\} \tag{4.238}$$

(c) **Moments and Autocorrelation**: It is better to derive equations for the moments from (4.234) rather than calculate moments and autocorrelation directlyy from the solution (4.238).

For we have

$$\begin{aligned} d[x(t)^n] &= nx(t)^{n-1}dx(t) + \tfrac{1}{2}n(n-1)x(t)^{n-2}[dx(t)]^2 \\ &= nx(t)^{n-1}dx(t) + \tfrac{1}{2}n(n-1)x(t)^{n-2}[f(t) + g(t)x(t)]^2dt \end{aligned} \tag{4.239}$$

Hence,

$$\begin{aligned} \frac{d}{dt}\langle x(t)^n\rangle = \langle x(t)^n\rangle &\left[nb(t) + \tfrac{1}{2}n(n-1)g(t)^2\right] \\ &+ \langle x(t)^{n-1}\rangle[na(t) + n(n-1)f(t)g(t)] \\ &+ \langle x(t)^{n-2}\rangle\tfrac{1}{2}n(n-1)f(t)^2 \end{aligned} \tag{4.240}$$

These equations form a hierarchy in which the n^{th} equation involves the solutions of the previous two, and can be integrated successively.

4.5.8 Multivariable Linear Equations

(a) **Homogeneous Case**: The equation is

$$d\mathbf{x}(t) = \left[B(t)dt + \sum_i G_i(t)dW_i(t)\right]\mathbf{x}(t) \tag{4.241}$$

where $B(t)$, $G_i(t')$ are matrices. The equation is not, in general. soluble in closed form unless the matrices $B(t)$, $G_i(t')$ commute at all times with each other, i.e.,

$$\left.\begin{array}{l} G_i(t)G_j(t') = G_j(t')G_i(t) \\ B(t)G_i(t') = G_i(t')B(t) \\ B(t)B(t') = B(t')B(t) \end{array}\right\} \tag{4.242}$$

In this case, the solution is completely analogous to the one variable case and we have

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) \tag{4.243}$$

with

$$\Phi(t) = \exp\left\{\int_0^t \left[B(t) - \tfrac{1}{2}\sum_i G_i(t)^2\right] dt + \int_0^t \sum_i G_i(t)dW_i(t)\right\} \tag{4.244}$$

(b) **Inhomogeneous Case**: We can reduce the inhomogeneous case to the homogeneous case in exactly the same way as in one dimension. Thus, we consider

$$d\mathbf{x}(t) = [\mathbf{A}(t) + B(t)\mathbf{x}]\,dt + \sum_i [\mathbf{F}_i(t) + G_i(t)\mathbf{x}]dW(t) \tag{4.245}$$

and write

$$\mathbf{y}(t) = \psi(t)^{-1}\mathbf{x}(t) \tag{4.246}$$

where $\psi(t)$ is a matrix solution of the homogeneous equation (4.241). We first have to evaluate $d[\psi^{-1}]$. For any matrix M we have $MM^{-1} = 1$, so expanding to second order, $Md[M^{-1}] + dMM^{-1} + dMdM^{-1}] = 0$.

Hence, $dM^{-1}] = -[M + dM]^{-1}dMM^{-1}$ and again to second order

$$dM^{-1}] = -M^{-1}dMM^{-1} + M^{-1}dMM^{-1}dMM^{-1} \tag{4.247}$$

and thus, since $\psi(t)$ satisfies the homogeneous equation,

$$d[\psi(t)^{-1}] = \psi(t)^{-1}\left\{\left[-B(t) + \sum_i G_i(t)^2\right] dt - \sum_i G_i(t)dW_i(t)\right\} \tag{4.248}$$

and, again taking differentials

$$d\mathbf{y}(t) = \psi(t)^{-1}\left\{\left[\mathbf{A}(t) - \sum_i G_i(t)\mathbf{F}_i(t)\right] dt + \sum_i \mathbf{F}_i(t)dW_i(t)\right\} \tag{4.249}$$

Hence,

$$\mathbf{x}(t) = \psi(t)\left\{\mathbf{x}(0) + \int_0^t \psi(t')^{-1}\left\{\left[\mathbf{A}(t') - \sum_i G_i(t')\mathbf{F}_i(t')\right] dt' + \sum_i \mathbf{F}_i(t')dW_i(t')\right\}\right\} \tag{4.250}$$

This solution is not very useful for practical purposes, even when the solution for the homogeneous equation is known, because of the difficulty in evaluating means and correlation functions.

4.5.9 Time-Dependent Ornstein-Uhlenbeck Process

This is a particular case of the previous general linear equation which is soluble. It is a generalization of the multivariate Ornstein-Uhlenbeck process (Section 4.5.6) to include time-dependent parameters, namely,

$$d\mathbf{x}(t) = -A(t)\mathbf{x}(t) + B(t)d\mathbf{W}(t) \tag{4.251}$$

This is clearly of the same form as (4.245) with the replacements

$$\left.\begin{array}{l} A(t) \to 0 \\ B(t) \to -A(t) \\ \sum_i F_i(t)dW_i(t) \to B(t)d\mathbf{W}(t) \\ G_i(t) \to 0 \end{array}\right\} \tag{4.252}$$

The corresponding homogeneous equation is simply the deterministic equation

$$d\mathbf{x}(t) = -A(t)\mathbf{x}(t)dt \tag{4.253}$$

which is soluble provided $A(t)A(t') = A(t')A(t)$ and has the solution

$$\mathbf{x}(t) = \psi(t)\mathbf{x}(0) \tag{4.254}$$

with

$$\psi(t) = \exp\left[-\int_0^t A(t')dt'\right] \tag{4.255}$$

Thus, applying (4.252),

$$\mathbf{x}(t) = \exp\left[-\int_0^t A(t')dt'\right]\mathbf{x}(0) + \int_0^t \left\{\exp\left[-\int_{t'}^t A(s)ds\right]\right\} B(t')d\mathbf{W}(t') \tag{4.256}$$

This is very similar to the solution of the time-independent Ornstein-Uhlenbeck process, as derived in Section 4.5.6, equation (4.202).

From this we have

$$\langle\mathbf{x}(t)\rangle = \exp\left[-\int_0^t A(t')dt'\right]\langle\mathbf{x}(0)\rangle \tag{4.257}$$

$$\begin{aligned}\langle\mathbf{x}(t),\mathbf{x}^T(t)\rangle = \exp\left[-\int_0^t A(t')dt'\right]\langle\mathbf{x}(0),\mathbf{x}^T(0)\rangle \exp\left[-\int_0^t A(t')dt'\right] \\ + \int_0^t dt' \exp\left[-\int_{t'}^t A(s)ds\right]B(t')B^T(t')\exp\left[-\int_{t'}^t A(s)ds\right]\end{aligned} \tag{4.258}$$

The time-dependent Ornstein-Uhlenbeck process will arise very naturally in connection with the development of asymptotic methods in low-noise systems.