3. Markov Processes

3.1 Stochastic Processes

All of the examples given in Chap. 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 $X(t)$ exists. We can measure values $x_1, x_2, x_3, \ldots$, etc., of $X(t)$ at times $t_1, t_2, t_3, \ldots$ and we assume that a set of joint probability densities exists

$$p(x_1, t_1; x_2, t_2; x_3, t_3; \ldots) \tag{3.1.1}$$

which describe the system completely.

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

$$\begin{aligned} &p(x_1, t_1, ; x_2, t_2; \ldots | y_1, \tau_1; y_2, \tau_2; \ldots) \\ &\quad = p(x_1, t_1; x_2, t_2; \ldots; y_1, \tau_1; y_2, \tau_2; \ldots)/p(y_1, \tau_1; y_2, \tau_2; \ldots). \end{aligned} \tag{3.1.2}$$

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.,

$$t_1 \geqslant t_2 \geqslant t_3 \geqslant \ldots \geqslant \tau_1 \geqslant \tau_2 \geqslant \ldots . \tag{3.1.3}$$

The concept of an evolution equation leads us to consider the conditional probabilities as predictions of the future values of $X(t)$ (i.e., $x_1, x_2, \ldots$ at times $t_1, t_2, \ldots$), given the knowledge of the past (values $y_1, y_2, \ldots$, at times $\tau_1, \tau_2, \ldots$).

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 in (3.1.1). If such knowledge does define the process, it is known as a *separable stochastic process.* All the processes considered in this book will be assumed to be separable.

The most simple kind of stochastic process is that of complete independence:

$$p(x_1, t_1; x_2, t_2; x_3, t_3; \ldots) = \prod_i p(x_i, t_i) \tag{3.1 4}$$

which means that the value of X at time t is completely independent of its values in the past (or future). An even more special case occurs when the $p(x_i, t_i)$ 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.

The next most simple idea is that of the *Markov process* in which knowledge o only the present determines the future.

3.2 Markov Process

The *Markov assumption* is formulated in terms of the conditional probabilities. W require that if the times satisfy the ordering (3.1.3), the conditional probability i determined entirely by the knowledge of the most recent condition, i.e.,

$$p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \ldots \,|\, \boldsymbol{y}_1, \tau_1; \boldsymbol{y}_2, \tau_2; \ldots) = p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \ldots \,|\, \boldsymbol{y}_1, \tau_1). \tag{3.2.1}$$

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 mean that we can define everything in terms of the simple conditional probabilitie $p(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{y}_1, \tau_1)$. For example, by definition of the conditional probability densit $p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2 \,|\, \boldsymbol{y}_1, \tau_1) = p(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2; \boldsymbol{y}_1, \tau_1) p(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{y}_1, \tau_1)$ and using the Marko assumption (3.2.1), we find

$$p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \boldsymbol{y}_1, \tau_1) = p(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) p(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{y}_1, \tau_1) \tag{3.2.2}$$

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

$$\begin{aligned} &p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \boldsymbol{x}_3, t_3; \ldots \boldsymbol{x}_n, t_n) \\ &\quad = p(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) p(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{x}_3, t_3) p(\boldsymbol{x}_3, t_3 \,|\, \boldsymbol{x}_4, t_4) \ldots \\ &\qquad \ldots p(\boldsymbol{x}_{n-1}, t_{n-1} \,|\, \boldsymbol{x}_n, t_n) p(\boldsymbol{x}_n, t_n) \end{aligned} \tag{3.2.3}$$

provided

$$t_1 \geqslant t_2 \geqslant t_3 \geqslant \ldots \geqslant t_{n-1} \geqslant t_n \,. \tag{3.2.4}$$

3.2.1 Consistency—the Chapman-Kolmogorov Equation

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

$$\sum_B P(A \cap B \cap C \ldots) = P(A \cap C \ldots) \,; \tag{3.2.5}$$

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

$$\begin{aligned} p(\boldsymbol{x}_1, t_1) &= \int d\boldsymbol{x}_2 \, p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) \\ &= \int d\boldsymbol{x}_2 \, p(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) p(\boldsymbol{x}_2, t_2) \,. \end{aligned} \tag{3.2.6}$$

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

$$p(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_3, t_3) = \int d\boldsymbol{x}_2\, p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2 | \boldsymbol{x}_3, t_3)$$
$$= \int d\boldsymbol{x}_2\, p(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2; \boldsymbol{x}_3, t_3) p(\boldsymbol{x}_2, t_2 | \boldsymbol{x}_3, t_3). \tag{3.2.7}$$

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

$$p(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_3, t_3) = \int d\boldsymbol{x}_2\, p(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2) p(\boldsymbol{x}_2, t_2 | \boldsymbol{x}_3, t_3) \tag{3.2.8}$$

which is the *Chapman-Kolmogorov equation.*

What is the essential difference between (3.2.8) and (3.2.6)? The obvious answer is that (3.2.6) is for unconditioned probabilities, whereas (3.2.7) is for conditional probabilities. Equation (3.2.8) is a rather complex nonlinear functional equation relating all conditional probabilities $p(\boldsymbol{x}_i, t_i | \boldsymbol{x}_j, t_j)$ to each other, whereas (3.2.6) simply constructs the one time probabilities in the future t_1 of t_2, given the conditional probability $p(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2)$.

The Chapman-Kolmogorov equation has many solutions. These are best understood by deriving the differential form which is done in Sect. 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 $\boldsymbol{N} = (N_1, N_2, N_3 \ldots)$, where the N_i are random variables which take on integral values. Clearly, we now replace

$$\int d\boldsymbol{x} \leftrightarrow \sum_{\boldsymbol{n}} \tag{3.2.9}$$

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

$$P(\boldsymbol{n}_1, t_1 | \boldsymbol{n}_3, t_3) = \sum_{\boldsymbol{n}_2} P(\boldsymbol{n}_1, t_1 | \boldsymbol{n}_2, t_2)\, P(\boldsymbol{n}_2, t_2 | \boldsymbol{n}_3, t_3)\,. \tag{3.2.10}$$

This is now a matrix multiplication, with possibly infinite matrices.

3.2.3 More General Measures

A more general formulation would assume a measure $d\mu(\boldsymbol{x})$ instead of $d\boldsymbol{x}$ where a variety of choices can be made. For example, if $\mu(\boldsymbol{x})$ is a step function with steps at integral values of $\boldsymbol{x}$, we recover the discrete state 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 favour a more specific notation.

3.3 Continuity in Stochastic Processes

Whether or not the random variable $X(t)$ has a continuous range of possible values is a completely different question from whether the sample path of $X(t)$ is a continuous function of t. For example, in a gas composed of molecules with velocities $V(t)$, it is clear that all possible values of $V(t)$ are in principle realisable, so that the *range* of $V(t)$ 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, v_i, will change instantaneously at the time of impact to another value v_f, so the sample path of $V(t)$ is not continuous. Nevertheless, in such a model, the *position* of a gas molecule $X(t)$ would be expected to be continuous.

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 idealisation 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 in all attempts to derive Markovian probabilistic equations from mechanics. Equations which are derived are rarely truly Markovian—rather there is a certain characteristic memory time during which the previous history is important (*Haake* [3.1]).

This means that 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 as being well approximated by a Markov process. But in this case, the question of whether the sample paths are actually 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 modelled 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 modelled 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

modelled as changing discontinuously by discrete jumps. Compared to the distances 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 of 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 its sign. The velocities simply cannot reach (with any significant probability) values for which the changes of velocity can be regarded as very small. Hence, there is no sense in a continuous path description of velocities in a gas.

3.3.1 Mathematical Definition of a Continuous Markov Process

For a *Markov process*, it can be shown [3.2] that with probability one, the sample paths are continuous functions of t, if for any $\varepsilon > 0$ we have

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x-z|>\varepsilon} dx\, p(x, t + \Delta t | z, t) = 0 \tag{3.3.1}$$

uniformly in z, t and Δt.

This means that the probability for the final position x to be finitely different from z goes to zero *faster* than Δt, as Δt goes to zero. [Equation (3.3.1) is sometimes called the Lindeberg condition.]

Examples

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

$$p(x, t + \Delta t | z, t) = (4\pi D \Delta t)^{-1/2} \exp\left[-(x - z)^2/4D\Delta t\right] \tag{3.3.2}$$

and it is easy to check that (3.3.1) is satisfied in this case. Thus, Brownian motion in Einstein's formulation has continuous sample paths.

ii) Cauchy Process: Suppose

$$p(x, t + \Delta t | z, t) = \frac{\Delta t}{\pi} / [(x - z)^2 + \Delta t^2]\,. \tag{3.3.3}$$

Then this does not satisfy (3.3.1) so the sample paths are discontinuous.

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

$$\lim_{\Delta t \to 0} p(x, t + \Delta t | z, t) = \delta(x - z)\,, \tag{3.3.4}$$

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

The difference between the two processes just described is illustrated in Fig. 3.1 in which simulations of both processes are given. The difference between the two is

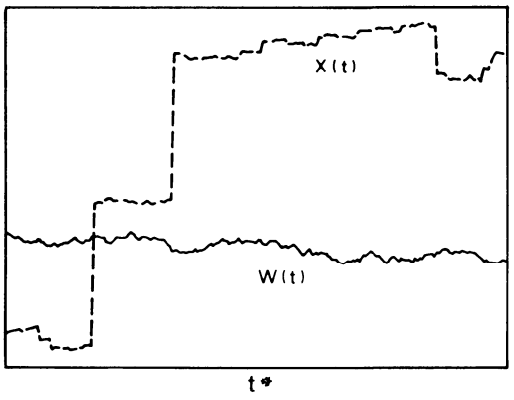

Fig. 3.1. Illustration of sample paths of the Cauchy process $X(t)$ (- - - - -) and Brownian motion $W(t)$ (———)

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 is, of course, wildly discontinuous.

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 (3.3.1), one is led to consider a method of dividing the differentiability conditions into parts, one corresponding to continuous motion of a representative point and the other to discontinuous motion.

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

$$\text{i)} \lim_{\Delta t \to 0} p(\boldsymbol{x}, t + \Delta t \,|\, \boldsymbol{z}, t)/\Delta t = W(\boldsymbol{x} \,|\, \boldsymbol{z}, t) \tag{3.4.1}$$

uniformly in $\boldsymbol{x}$, $\boldsymbol{z}$, and t for $|\boldsymbol{x} - \boldsymbol{z}| \geqslant \varepsilon$;

$$\text{ii)} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}\,(x_i - z_i) p(\boldsymbol{x}, t + \Delta t \,|\, \boldsymbol{z}, t) = A_i(\boldsymbol{z}, t) + O(\varepsilon)\ ; \tag{3.4.2}$$

$$\text{iii)} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}\,(x_i - z_i)(x_j - z_j) p(\boldsymbol{x}, t + \Delta t \,|\, \boldsymbol{z}, t) = B_{ij}(\boldsymbol{z}, t) + O(\varepsilon); \tag{3.4.3}$$

the last two being uniform in $\boldsymbol{z}$, ε, and t.

Notice that all higher-order coefficients of the form in (3.4.2,3) must vanish. For example, consider the third-order quantity defined by

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} dx\,(x_i - z_i)(x_j - z_j)(x_k - z_k)\, p(\boldsymbol{x}, t + \Delta t \,|\, \boldsymbol{z}, t)$$
$$\equiv C_{ijk}(\boldsymbol{z}, t) + O(\varepsilon)\,. \tag{3.4.4}$$

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

$$\sum_{i,j,k} \alpha_i \alpha_j \alpha_k C_{ijk}(z, t) \equiv \bar{C}(\alpha, z, t) \tag{3.4.5}$$

so that

$$C_{ijk}(z, t) = \frac{1}{3!} \frac{\partial^3}{\partial \alpha_i \partial \alpha_j \partial \alpha_k} \bar{C}(\alpha, z, t) . \tag{3.4.6}$$

Then,

$$\begin{aligned} |\bar{C}(\alpha, z, t)| &\leqslant \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x-z|<\varepsilon} |\alpha \cdot (x - z)| [\alpha \cdot (x - z)]^2 \, p(x, t + \Delta t | z, t) \, dx \\ &\quad + O(\varepsilon) \\ &\leqslant |\alpha| \varepsilon \lim_{\Delta t \to 0} \int [\alpha \cdot (x - z)]^2 p(x, t + \Delta t | z, t) dx + O(\varepsilon) \\ &= \varepsilon |\alpha| [\alpha_i \alpha_j B_{ij}(z, t) + O(\varepsilon)] + O(\varepsilon) \\ &= O(\varepsilon) \end{aligned} \tag{3.4.7}$$

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

According to the condition for continuity (3.3.1), the process can only have continuous paths if $W(x | z, t)$ vanishes for all $x \neq z$. 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 $f(z)$ which is twice continuously differentiable.

Thus,

$$\begin{aligned} &\partial_t \int dx \, f(x) p(x, t | y, t') \\ &\quad = \lim_{\Delta t \to 0} \{ \int dx \, f(x) [p(x, t + \Delta t | y, t') - p(x, t | y, t')] \} / \Delta t \qquad (3.4.8) \\ &\quad = \lim_{\Delta t \to 0} \{ \int dx \int dz \, f(x) p(x, t + \Delta t | z, t) p(z, t | y, t') \\ &\qquad - \int dz \, f(z) p(z, t | y, t') \} / \Delta t , \qquad (3.4.9) \end{aligned}$$

where we have used the Chapman-Kolmogorov equation in the positive term of (3.4.8) to produce the corresponding term in (3.4.9).

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

$$f(\boldsymbol{x}) = f(\boldsymbol{z}) + \sum_i \frac{\partial f(\boldsymbol{z})}{\partial z_i}(x_i - z_i) + \sum_{i,j} \frac{1}{2}\frac{\partial^2 f(\boldsymbol{z})}{\partial z_i \partial z_j}(x_i - z_i)(x_j - z_j) + |\boldsymbol{x} - \boldsymbol{z}|^2 R(\boldsymbol{x}, \boldsymbol{z}), \tag{3.4.10}$$

where we have (again by the twice continuous differentiability)

$$|R(\boldsymbol{x}, \boldsymbol{z})| \to 0 \quad \text{as} \quad |\boldsymbol{x} - \boldsymbol{z}| \to 0. \tag{3.4.11}$$

Now substitute in (3.4.9):

$$\begin{aligned}(3.4.9) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Bigg\{ & \iint_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}d\boldsymbol{z} \left[\sum_i (x_i - z_i)\frac{\partial f}{\partial z_i} + \sum_{i,j} \frac{1}{2}(x_i - z_i)(x_j - z_j)\frac{\partial^2 f}{\partial z_i \partial z_j} \right] \\ & \times p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t)\, p(\boldsymbol{z}, t | \boldsymbol{y}, t') \\ & + \iint_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}d\boldsymbol{z}\, |\boldsymbol{x} - \boldsymbol{z}|^2 R(\boldsymbol{x}, \boldsymbol{z}) p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t) p(\boldsymbol{z}, t | \boldsymbol{y}, t') \\ & + \iint_{|\boldsymbol{x}-\boldsymbol{z}|\geqslant\varepsilon} d\boldsymbol{x}d\boldsymbol{z}\, f(\boldsymbol{x}) p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t) p(\boldsymbol{z}, t | \boldsymbol{y}, t') \\ & + \iint_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}d\boldsymbol{z}\, f(\boldsymbol{z}) p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t) p(\boldsymbol{z}, t | \boldsymbol{y}, t') \\ & - \iint d\boldsymbol{x}d\boldsymbol{z}\, f(\boldsymbol{z}) p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t) p(\boldsymbol{z}, t | \boldsymbol{y}, t') \Bigg\}\end{aligned} \tag{3.4.12}$$

[notice that since $p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t)$ is a probability, the integral over $\boldsymbol{x}$ in the last term gives 1—this is simply the last term in (3.4.9)].

We now consider these line by line.

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

$$\int d\boldsymbol{z} \left[\sum_i A_i(\boldsymbol{z}) \frac{\partial f}{\partial z_i} + \frac{1}{2} \sum_{i,j} B_{ij}(\boldsymbol{z}) \frac{\partial^2 f}{\partial z_i \partial z_j} \right] p(\boldsymbol{z}, t | \boldsymbol{y}, t') + O(\varepsilon). \tag{3.4.13}$$

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

$$\begin{aligned} & \left| \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}\, |\boldsymbol{x} - \boldsymbol{z}|^2 R(\boldsymbol{x}, \boldsymbol{z}) p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t) \right| \\ & \leqslant \left[\frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}\, |\boldsymbol{x} - \boldsymbol{z}|^2 p(\boldsymbol{x}, t + \Delta t | \boldsymbol{z}, t) \right] \operatorname*{Max}_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} |R(\boldsymbol{x}, \boldsymbol{z})| \\ & \to [\sum_i B_{ii}(\boldsymbol{z}, t) + O(\varepsilon)] \ \{\operatorname*{Max}_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} |R(\boldsymbol{x}, \boldsymbol{z})|\}. \end{aligned} \tag{3.4.14}$$

From (3.4.11) we can see that as $\varepsilon \to 0$, the factor in curly brackets vanishes.

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

$$\iint_{|\boldsymbol{x}-\boldsymbol{z}|\geqslant\varepsilon} d\boldsymbol{x}d\boldsymbol{z}\, f(\boldsymbol{z}) [W(\boldsymbol{z} | \boldsymbol{x}, t) p(\boldsymbol{x}, t | \boldsymbol{y}, t') - W(\boldsymbol{x} | \boldsymbol{z}, t) p(\boldsymbol{z}, t | \boldsymbol{y}, t')]. \tag{3.4.15}$$

The whole right-hand side of (3.4.12) is independent of ε. Hence, taking the limit $\varepsilon \to 0$, we find

$$\partial_t \int dz\, f(z) p(z, t \mid y, t') = \int dz \left[\sum_i A_i(z, t) \frac{\partial f(z)}{\partial z_i} + \frac{1}{2} \sum_{i,j} B_{ij}(z) \frac{\partial^2 f(z)}{\partial z_i \partial z_j} \right] p(z, t \mid y, t') + \int dz\, f(z) \{ \int dx [W(x \mid z, t) p(x, t \mid y, t') - W(x \mid z, t) p(z, t \mid y, t')] \} . \quad (3.4.16)$$

Notice, however, that we use the definition

$$\lim_{\varepsilon \to 0} \int_{|x-z|>\varepsilon} dx\, F(x, z) \equiv \int dx\, F(x, z) \quad (3.4.17)$$

for a principal value integral of a function $F(x, z)$. For (3.4.16) to have any meaning, this integral should exist. Equation (3.4.1) defines $W(x \mid z, t)$ only for $x \neq z$ and hence leaves open the possibility that it is infinite at $x = z$, as is indeed the case for the Cauchy process, discussed in Sect. 3.3.1, for which

$$W(x \mid z, t) = 1/[\pi(x - z)^2] . \quad (3.4.18)$$

However, if $p(x, t \mid y, t')$ is continuous and once differentiable, then the principal value integral exists. In the remainder of the book we shall 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 now is to integrate by parts. We find

$$\int dz\, f(z) \partial_t p(z, t \mid y, t') = \int dz\, f(z) \Big\{ -\sum_i \frac{\partial}{\partial z_i} A_i(z, t) p(z, t \mid y, t') + \sum_{i,j} \frac{1}{2} \frac{\partial^2}{\partial z_i \partial z_j} B_{ij}(z, t) p(z, t \mid y, t') + \int dx [W(z \mid x, t) p(x, t \mid y, t') - W(x \mid z, t) p(z, t \mid y, t')] \Big| + \text{surface terms.} \quad (3.4.19)$$

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

$$p(x, t \mid z, t') = 0 \text{ unless both } x \text{ and } z \in R. \quad (3.4.20)$$

It is clear that by definition we have

$$W(x \mid z, t) = 0 \text{ unless both } x \text{ and } y \in R. \quad (3.4.21)$$

But the conditions on $A_i(z, t)$ and $B_{ij}(z, t)$ can result in discontinuities in these functions as defined by (3.4.2.3) since the conditional probability $p(x, t + \Delta t \mid z, t')$ can very reasonably change discontinuously as 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* of R,

$$\begin{aligned}\partial_t p(\mathbf{z}, t\,|\,\mathbf{y}, t') = &-\sum_i \frac{\partial}{\partial z_i}[A_i(\mathbf{z}, t)p(\mathbf{z}, t\,|\,\mathbf{y}, t')] \\ &+ \sum_{i,j} \frac{1}{2}\frac{\partial^2}{\partial z_i \partial z_j}[B_{ij}(\mathbf{z}, t)p(\mathbf{z}, t\,|\,\mathbf{y}, t')] \\ &+ \int d\mathbf{x}\,[W(\mathbf{z}\,|\,\mathbf{x}, t)p(\mathbf{x}, t\,|\,\mathbf{y}, t') - W(\mathbf{x}\,|\,\mathbf{z}, t)p(\mathbf{z}, t\,|\,\mathbf{y}, t')]\,.\end{aligned} \tag{3.4.22}$$

Surface terms do not arise, since they necessarily vanish.

This equation does not seem to have any agreed name in the literature. Since it is purely a differential form of the Chapman-Kolmogorov equation, I propose to call it the *differential Chapman-Kolmogorov* equation.

3.4.2 Status of the Differential Chapman-Kolmogorov Equation

From our derivation 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 [3.3] that, under certain conditions, if we specify $\boldsymbol{A}(\boldsymbol{x}, t)$, $\underline{\boldsymbol{B}}(\boldsymbol{x}, t)$ (which must be positive semi-definite), and $W(\boldsymbol{x}\,|\,\boldsymbol{y}, t)$ (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,*

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

which follows from the definition of the conditional probability density, and any appropriate boundary conditions. These are very difficult to specify in the full equation, but in the case of the *Fokker-Planck equation* (Sect. 3.5.2) are given in Chap. 5.

3.5 Interpretation of Conditions and Results

Each of the conditions (i), (ii), (iii) of Sect. 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

$$A_i(z, t) = B_{ij}(z, t) = 0 \tag{3.5.1}$$

so that we now have the *Master equation:*

$$\partial_t p(z, t \mid y, t') = \int dx\, [W(z \mid x, t)p(x, t \mid y, t') - W(x \mid z, t)p(z, t \mid y, t')]\,. \tag{3.5.2}$$

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

$$p(z, t \mid y, t) = \delta(y - z)\,. \tag{3.5.3}$$

Hence,

$$p(z, t + \Delta t \mid y, t) = \delta(y - z)[1 - \int dx\, W(x \mid y, t)\Delta t] + W(z \mid y, t)\Delta t\,. \tag{3.5.4}$$

We see that for any Δt there is a finite probability, given by the coefficient of the $\delta(y - z)$ in (3.5.4), for the particle to stay at the original position y. The distribution of those particles which do not remain at y is given by $W(z \mid y, t)$ after appropriate normalisation. Thus, a typical path $X(t)$ will consist of sections of straight lines $X(t) =$ constant, interspersed with discontinuous jumps whose distribution is given by $W(z \mid y, t)$. For this reason, the process is known as a jump process. The paths are discontinuous at discrete points.

In the case where the state space consists of integers only, the Master equation takes the form

$$\partial_t P(n, t \mid n', t') = \sum_m [W(n \mid m, t)P(m, t \mid n', t') - W(m \mid n, t)P(n, t \mid n', t')]. \tag{3.5.5}$$

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

3.5.2 Diffusion Processes—the Fokker-Planck Equation

If we assume the quantities $W(z \mid x, t)$ to be zero, the differential Chapman-Kolmogorov equation reduces to the *Fokker-Planck equation:*

$$\frac{\partial p(z, t \mid y, t')}{\partial t} = -\sum_i \frac{\partial}{\partial z_i}[A_i(z, t)p(z, t \mid y, t')] + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j}[B_{ij}(z, t)p\,(z, t \mid y, t')] \tag{3.5.6}$$

and the corresponding process is known mathematically as a *diffusion process.* The vector $A(z, t)$ is known as the drift vector and the matrix $\underline{B}(z, t)$ as the diffusion matrix. The diffusion matrix is positive semidefinite and symmetric as a result of its

definition in (3.4.3). It is easy to see that from the definition of $W(x|z, t)$ (3.4.1), the requirement (3.3.1) for continuity of the sample paths is satisfied if $W(x|z, t)$ is zero. Hence, the Fokker-Planck equation describes a process in which $X(t)$ has continuous sample paths. In fact, we can heuristically give a much more definite description of the process. Let us consider computing $p(z, t + \Delta t|y, t)$, given that

$$p(z, t|y, t) = \delta(z - y) . \tag{3.5.7}$$

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

$$\frac{\partial p(z, t|y, t')}{\partial t} = -\sum_i A_i(y, t)\frac{\partial p(z, t|y, t')}{\partial z_i} + \sum_{i,j}\frac{1}{2}B_{ij}(y, t)\frac{\partial^2 p(z, t|y, t')}{\partial z_i \partial z_j} , \tag{3.5.8}$$

where we have also neglected the time dependence of A_i and B_{ij} for small $t - t'$. Equation (3.5.8) can now be solved, subject to the initial condition (3.5.7), and we get

$$p(z, t + \Delta t|y, t) = (2\pi)^{-N/2}\{\det[\underset{\sim}{B}(y, t)]\}^{1/2}[\Delta t]^{-1/2} \times \exp\left\{-\frac{1}{2}\frac{[z - y - A(y, t)\Delta t]^T[\underset{\sim}{B}(y, t)]^{-1}[z - y - A(y, t)\Delta t]}{\Delta t}\right\}, \tag{3.5.9}$$

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

$$y(t + \Delta t) = y(t) + A(y(t), t)\Delta t + \eta(t)\Delta t^{1/2} , \tag{3.5.10}$$

where $\langle \eta(t) \rangle = 0$ (3.5.11)

$$\langle \eta(t)\eta(t)^T \rangle = \underset{\sim}{B}(y, t) . \tag{3.5.12}$$

It is easy to see that this picture gives

i) sample paths which are always continuous — for, clearly, as $\Delta t \to 0$, $y(t + \Delta t) \to y(t)$;

ii) sample paths which are nowhere differentiable, because of the $\Delta t^{1/2}$ occurring in (3.5.10).

We shall see later, in Chap. 4 that the heuristic picture of (3.5.10) 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 (3.4.22) only the first term is nonzero, so we are led to the special case of a *Liouville equation:*

$$\frac{\partial p(z, t \mid y, t')}{\partial t} = -\sum_i \frac{\partial}{\partial z_i} [A_i(z, t) p(z, t \mid y, t')] \tag{3.5.13}$$

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

$$\frac{dx(t)}{dt} = A[x(t), t] \tag{3.5.14}$$

with $x(y, t') = y$, (3.5.15)

then the solution to (3.5.13) with initial condition

$$p(z, t' \mid y, t') = \delta(z - y) \tag{3.5.16}$$

is

$$p(z, t \mid y, t') = \delta[z - x(y, t)] . \tag{3.5.17}$$

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

$$-\sum_i \frac{\partial}{\partial z_i} \{A_i(z, t) \delta[z - x(y, t)]\} \tag{3.5.18}$$

$$= -\sum_i \frac{\partial}{\partial z_i} \{A_i[x(y, t), t] \delta[z - x(y, t)]\} \tag{3.5.19}$$

$$= -\sum_i \left\{A_i[x(y, t), t] \frac{\partial}{\partial z_i} \delta[z - x(y, t)]\right\} \tag{3.5.20}$$

and

$$\frac{\partial}{\partial t} \delta[z - x(y, t)] = -\sum_i \frac{\partial}{\partial z_i} \delta[z - x(y, t)] \frac{dx_i(y, t)}{dt} \tag{3.5.21}$$

and by use of (3.5.14), we see that (3.5.20,21) are equal. Thus, if the particle is in a well-defined initial position y at time t', it stays on the trajectory obtained by solving the ordinary differential equation (3.5.14).

Hence, deterministic motion, as defined by a first-order differential equation of the form (3.5.14), is an elementary form of Markov process. The solution (3.5.17) is, of course, merely a special case of the kind of process approximated by equations like (3.5.9) in which the Gaussian part is zero.

3.5.4 General Processes

In general, none of the quantities in $A(z, t)$, $B(z, t)$ and $W(x \mid z, t)$ need vanish, and in this case we obtain a process whose sample paths are as illustrated in Fig. 3.2, 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.

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

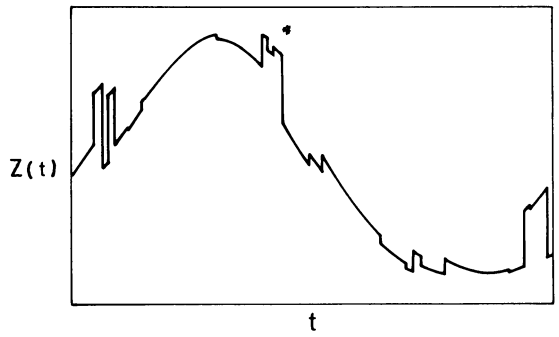

Fig. 3.3. Sample path of a Markov process with only drift and jumps

It is also possible that $A(z, t)$ is nonzero, but $B(z, t)$ is zero and here the sample paths are, as in Fig. 3.3, composed of pieces of smooth curve [solutions of (3.5.14)] 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 Sect. 3.4, some equations which give the time development with respect to the initial variables y, t' of $p(x, t|y, t')$.

We consider

$$\lim_{\Delta t' \to 0} \frac{1}{\Delta t'} [p(x, t|y, t' + \Delta t') - p(x, t|y, t')] \tag{3.6.1}$$

$$= \lim_{\Delta t' \to 0} \frac{1}{\Delta t'} \int dz\, p(z, t' + \Delta t'|y, t')[p(x, t|y, t' + \Delta t') - p(x, t|z, t' + \Delta t')] \tag{3.6.2}$$

by use of the Chapman-Kolmogrov equation in the second term and by noting that the first term gives $1 \times p(x, t|y, t' + \Delta t')$.

The assumptions that are necessary are now the existence of all relevant deriva-