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 $\boldsymbol{X}(t)$ exists. We can measure values $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$, etc., of $\boldsymbol{X}(t)$ at times t_1, $t_2, t_3, \ldots$ and we assume that a set of joint probability densities exists

$$p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \boldsymbol{x}_3, t_3; \ldots) \tag{3.1.1}$$

which describe the system completely.

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

$$\begin{aligned} &p(\boldsymbol{x}_1, t_1,; \boldsymbol{x}_2, t_2; \ldots | \boldsymbol{y}_1, \tau_1; \boldsymbol{y}_2, \tau_2; \ldots) \\ &\quad = p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \ldots; \boldsymbol{y}_1, \tau_1; \boldsymbol{y}_2, \tau_2; \ldots)/p(\boldsymbol{y}_1, \tau_1; \boldsymbol{y}_2, \tau_2; \ldots). \end{aligned} \tag{3.1.2}$$

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 $\boldsymbol{X}(t)$ (i.e., $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ at times $t_1, t_2, \ldots$), given the knowledge of the past (values $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$, 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:

$$p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \boldsymbol{x}_3, t_3; \ldots) = \prod_i p(\boldsymbol{x}_i, t_i) \tag{3.1 4}$$

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

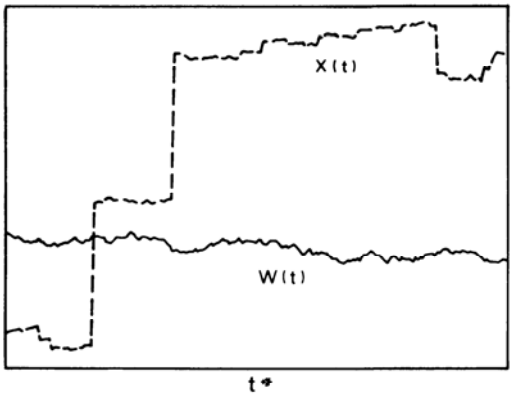

Fig. 3.1. Illustration of sample paths of the Cauchy process $X(t)$ (- - - - -) and Brownian motion $W(t)$ (———)

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 $|\boldsymbol{x} - \boldsymbol{z}| \geqslant \varepsilon$;

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

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

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

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

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,

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

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 $W(\boldsymbol{x} \mid \boldsymbol{y}, t)$ (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,*

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

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:*

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

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

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

Hence,

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

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 $W(z|y, t)$ 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 $W(z|y, t)$. 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

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

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 $W(z|x, t)$ to be zero, the differential Chapman-Kolmogorov equation reduces to the *Fokker-Planck equation:*

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

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

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

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

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

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

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

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

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

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

$$\langle \boldsymbol{\eta}(t)\boldsymbol{\eta}(t)^{\mathrm{T}} \rangle = \underline{\boldsymbol{B}}(\boldsymbol{y}, t) . \tag{3.5.12}$$

It is easy to see that this picture gives

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

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 équation.*

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 $A(z, t)$, $\underline{B}(z, t)$ 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.

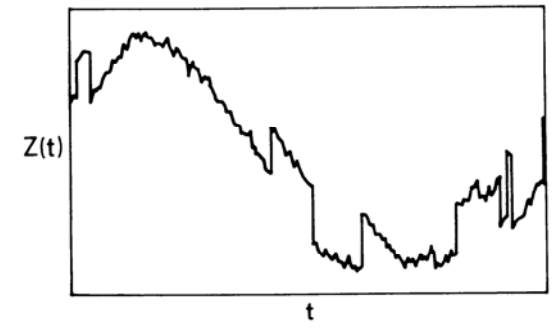

Fig. 3.2. Illustration of a sample path of a general Markov process, in which drift, diffusion and jumps exist

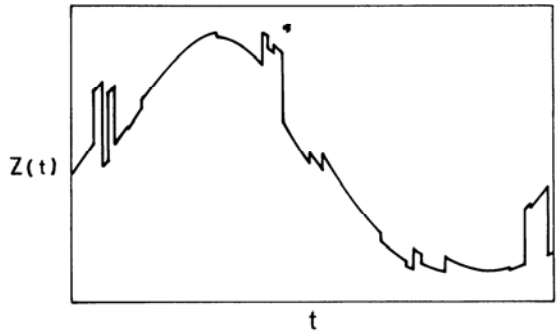

Fig. 3.3. Sample path of a Markov process with only drift and jumps

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-

tives, and that $p(x, t|y, t')$ is continuous and bounded in x,t,t' for some range $t - t' > \delta > 0$. We may then write

$$= \lim_{\Delta t' \to 0} \frac{1}{\Delta t'} \int dz\, p(z, t' + \Delta t'|y, t')\,[p(x, t|y, t') - p(x, t|z, t')] \tag{3.6.3}$$

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

$$\frac{\partial p(x, t|y, t')}{\partial t'} = -\sum_i A_i(y, t') \frac{\partial p(x, t|y, t')}{\partial y_i} - \frac{1}{2} \sum_{ij} B_{ij}(y, t') \frac{\partial^2 p(x, t|y, t')}{\partial y_i \partial y_j} + \int dz\, W(z|y, t')\,[p(x, t|y, t') - p(x, t|z, t')] \tag{3.6.4}$$

which will be called the backward differential Chapman-Kolmogorov equation. In a mathematical sense, it is better defined than the corresponding forward equation (3.4.22). The appropriate initial condition for both equation is

$$p(x, t|y, t) = \delta(x - y) \text{ for all } t, \tag{3.6.5}$$

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

The forward and the backward equations are equivalent to each other. For, solutions of the forward equation, subject to the initial condition (3.6.5) [or 3.5.4], and any appropriate boundary conditions, yield solutions of the Chapman-Kolmogorov equation, as noted in Sect. 3.4.2. But these have just been shown to yield the backward equation. (The relation between appropriate boundary conditions for the Fokker-Planck equations is dealt with in Sect. 5.2.1,4). The basic difference is which set of variables is held fixed. In the case of the forward equation, we hold (y, t') fixed, and solutions exist for $t \geqslant t'$, so that (3.6.5) is an *initial condition* for the forward equation. For the backward equation, solutions exist for $t' \leqslant t$, so that since the backward equation expresses development in t', (3.6.5) is really better termed *final 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

In Sect. 1.4.3 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 shall reserve the term "*stationary process*"

for a strict definition, namely, a stochastic process $\boldsymbol{X}(t)$ is stationary if $\boldsymbol{X}(t)$ and the process $\boldsymbol{X}(t+\varepsilon)$ have the same statistics for any ε. This is equivalent to saying that all joint probability densities satisfy time translation invariance, i.e.,

$$p(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \boldsymbol{x}_3, t_3; \ldots; \boldsymbol{x}_n, t_n) = p(\boldsymbol{x}_1, t_1 + \varepsilon; \boldsymbol{x}_2, t_2 + \varepsilon; \boldsymbol{x}_3, t_3 + \varepsilon; \ldots; \boldsymbol{x}_n, t_n + \varepsilon) \tag{3.7.1}$$

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 be simply written as

$$p_s(\boldsymbol{x}) \tag{3.7.2}$$

and the two-time joint probability as

$$p_s(\boldsymbol{x}_1, t_1 - t_2; \boldsymbol{x}_2, 0). \tag{3.7.3}$$

Finally, the conditional probability can also be written as

$$p_s(\boldsymbol{x}_1, t_1 - t_2 | \boldsymbol{x}_2, 0). \tag{3.7.4}$$

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 two-time probabilities in the forms given in (3.7.1–3).

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 $\boldsymbol{x}$ at successive times, and averaging various functions of these. This is effectively a belief that the law of large numbers (as explained in Sect. 2.5.2) applies to the variables defined by successive measurements in a stochastic process.

Let us define the variable $\bar{X}(T)$ by

$$\bar{X}(T) = \frac{1}{2T} \int_{-T}^{T} dt\, x(t)\,, \tag{3.7.5}$$

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. Clearly

$$\langle \bar{X}(t) \rangle = \langle x \rangle_s\,. \tag{3.7.6}$$

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

$$\langle \bar{X}(T)^2 \rangle = \frac{1}{4T^2} \int_{-T}^{T} \int_{-T}^{T} dt_1 dt_2\, \langle x(t_1) x(t_2) \rangle \tag{3.7.7}$$

and if the process is stationary,

$$\langle x(t_1)\, x(t_2)\rangle \equiv R(t_1 - t_2) + \langle x\rangle^2 , \tag{3.7.8}$$

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

$$\langle \bar{X}(T)^2\rangle - \langle x\rangle^2 = \frac{1}{4T^2} \int_{-2T}^{2T} d\tau\; R(\tau)(2T - |\tau|) \tag{3.7.8}$$

where the last factor follows by changing variables to

$$\tau = t_1 - t_2 \tag{3.7.9}$$
$$t = t_1$$

and integrating t.

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

$$\lim_{T\to\infty} \frac{1}{T} \int_{-2T}^{2T} d\tau \left(1 - \frac{|\tau|}{2T}\right) R(\tau) = 0 \tag{3.7.10}$$

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

$$\int_0^\infty d\tau\; |R(\tau)| < \infty , \tag{3.7.11}$$

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

$$R(\tau) \sim \mathrm{Re}\{A \exp(-\tau/\tau_c)\} , \tag{3.7.12}$$

where τ_c is a (possibly complex) parameter known as the *correlation time*. Clearly the criterion of (3.7.11) is satisfied, and we find in this case that the variance in $\bar{X}(T)$ approaches zero so that using (3.7.6) and (2.9.4), we may write

$$\underset{T\to\infty}{\text{ms-lim}}\; \bar{X}(T) = \langle x\rangle_s . \tag{3.7.13}$$

This means that the averaging procedure (3.7.5) 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 most interest are the autocorrelation function and the distribution function. As already mentioned in Sect. 1.4.2, the most natural way of measuring an autocorrelation function is through the definition

$$G(\tau, T) = \frac{1}{T}\int_0^T dt\, x(t)x(t+\tau) \tag{3.7.14}$$

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

$$\underset{T\to\infty}{\text{ms-lim}}\, G(\tau, T) = \langle x(t)x(t+\tau)\rangle_s\,, \tag{3.7.15}$$

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

$$\langle x(t+\lambda+\tau)x(t+\lambda)x(t+\tau)x(t)\rangle_s = \rho(\tau, \lambda) + \langle x(t+\tau)x(t)\rangle_s^2\,. \tag{3.7.16}$$

Then we require

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|\lambda|}{2T}\right)\rho(\tau, \lambda)d\lambda = 0\,. \tag{3.7.17}$$

We can see that this means that for sufficiently large λ, the four-time average (3.7.16) factorises into a product of two-time averages, and that the "error term" $\rho(\tau, \lambda)$ must vanish sufficiently rapidly for $\lambda \to \infty$. Exponential behaviour, such as given in (3.7.12) is sufficient, and usually found.

We similarly find that the spectrum, given by the Fourier transform

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}G(\tau)d\tau \tag{3.7.18}$$

as in Sect. 1.4, is also given by the procedure

$$S(\omega) = \lim_{T\to\infty}\frac{1}{2\pi T}\left|\int_0^T dt\, e^{-i\omega t}x(t)\right|^2\,. \tag{3.7.19}$$

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 this 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

$$\begin{aligned}\chi(x) &= 1 \qquad x_1 < x < x_2 \\ &= 0 \qquad \text{otherwise,}\end{aligned} \tag{3.7.20}$$

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

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T} d\tau\left(1 - \frac{|\tau|}{2T}\right)\int_{x_1}^{x_2} dx' p_s(x')\left\{\int_{x_1}^{x_2} dx\,[p(x, \tau\,|\,x', 0) - p_s(x)]\right\} = 0\,. \tag{3.7.21}$$

The most obvious sufficient condition here is that

$$\lim_{\tau\to\infty} p(x, \tau | x', 0) = p_s(x) \tag{3.7.22}$$

and that this limit is approached sufficiently rapidly. In practice, an exponential approach is 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 (3.7.22) can be readily seen to be sufficient to guarantee both (3.7.17) and (3.7.10). We will call a Markov process simply "*ergodic*" if this rather strong condition is satisfied.

3.7.2 Homogeneous Processes

If the condition (3.7.22) is satisfied for a stationary Markov process, then we clearly have a way of constructing from the stationary Markov process a nonstationary process whose limit as time becomes large is the stationary process. We simply define the process for

$$t, t' > t_0 \tag{3.7.23}$$

by

$$p(\boldsymbol{x}, t) = p_s(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) \qquad \text{and} \tag{3.7.24}$$

$$p(\boldsymbol{x}, t | \boldsymbol{x}', t') = p_s(\boldsymbol{x}, t | \boldsymbol{x}', t') \tag{3.7.25}$$

and all other joint probabilities are obtained from these in the usual manner for a Markov process. Clearly, if (3.7.22) is satisfied, we find that as $t \to \infty$ or as $t_0 \to -\infty$,

$$p(\boldsymbol{x}, t) \to p_s(\boldsymbol{x})$$

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 $\boldsymbol{x}$ is by some external agency fixed to have a value $\boldsymbol{x}_0$ at time t_0. 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

$$\begin{aligned} 0 = &-\sum_i \frac{\partial}{\partial z_i}[A_i(\boldsymbol{z})p(\boldsymbol{z}, t | \boldsymbol{y}, t')] + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial z_i \partial z_j}[B_{ij}(\boldsymbol{z})p(\boldsymbol{z}, t | \boldsymbol{y}, t')] \\ &+ \int d\boldsymbol{x}\,[W(\boldsymbol{z} | \boldsymbol{x})p(\boldsymbol{x}, t | \boldsymbol{y}, t') - W(\boldsymbol{x} | \boldsymbol{z})p(\boldsymbol{z}, t | \boldsymbol{y}, t')], \end{aligned} \tag{3.7.26}$$

where we have used the fact that the process is homogeneous to note that $\boldsymbol{A}$, $\underline{\boldsymbol{B}}$ and W, as defined in (3.4.1–3), 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, $\underset{\sim}{B}$ and W are independent of time and $p_s(z)$ satisfies (3.7.26). Under what conditions does a solution of the differential Chapman-Kolmogorov equation approach the stationary solution $p_s(z)$?

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-Kolgomorov equation by

$$K = \int dx\, p_1(x, t) \log [p_1(x, t)/p_2(x, t)] \tag{3.7.27}$$

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 $p_2(x, t)$ and $p_1(x, t)$ are normalised to one, we write

$$\begin{aligned} K[p_1, p_2, t] = \int dx\, p_1(x, t)\, \{&\log [p_1(x, t)/p_2(x, t)] \\ &+ p_2(x, t)/p_1(x, t) - 1\} \end{aligned} \tag{3.7.28}$$

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

$$-\log z + z - 1 \geqslant 0\,, \tag{3.7.29}$$

to show that $K \geqslant 0$.

Let us now show that $dK/dt \leqslant 0$. We can write (using an abbreviated notation)

$$\frac{dK}{dt} = \int dx \left\{ \frac{\partial p_1}{\partial t} [\log p_1 + 1 - \log p_2] - \frac{\partial p_2}{\partial t} [p_1/p_2] \right\}. \tag{3.7.30}$$

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

$$\begin{aligned} \left(\frac{dK}{dt}\right)_{\text{drift}} = \sum_i \int dx \Big\{ &- [\log (p_1/p_2) + 1] \frac{\partial}{\partial x_i} (A_i p_1) \\ &+ (p_1/p_2) \frac{\partial}{\partial x_i} (A_i p_2) \Big\} \end{aligned} \tag{3.7.31}$$

which can be rearranged to give

$$\left(\frac{dK}{dt}\right)_{\text{drift}} = \sum_i \int dx \frac{\partial}{\partial x_i} [- A_i p_1 \log (p_1/p_2)]. \tag{3.7.32}$$

Similarly, we may calculate

$$\begin{aligned} \left(\frac{dK}{dt}\right)_{\text{diff}} = -\frac{1}{2} \sum_{i,j} \int dx \Big\{ &[\log (p_1/p_2) + 1] \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij} p_1) \\ &- (p_1/p_2) \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij} p_2) \Big\} \end{aligned} \tag{3.7.33}$$

and after some rearranging we may write

$$\left(\frac{dK}{dt}\right)_{\text{diff}} = -\frac{1}{2}\sum_{i,j}\int dx\, p_1 B_{ij}\left\{\frac{\partial}{\partial x_i}[\log(p_1/p_2)]\right\}\left\{\frac{\partial}{\partial x_j}[\log(p_1/p_2)]\right\}$$
$$+\frac{1}{2}\sum_{i,j}\int dx\,\frac{\partial^2}{\partial x_i \partial x_j}\Big[p_1 B_{ij}\log(p_1/p_2)\Big]. \tag{3.7.34}$$

Finally, we may calculate the jump contribution similarly:

$$\left(\frac{dK}{dt}\right)_{\text{jump}} = \int dx dx'\Big\{[W(x\,|\,x')p_1(x',t) - W(x'\,|\,x)p_1(x,t)]$$
$$\times\{\log[p_1(x,t)/p_2(x,t)]+1\}$$
$$-[W(x\,|\,x')p_2(x',t) - W(x'\,|\,x)p_2(x,t)]p_1(x,t)/p_2(x,t)\Big\} \tag{3.7.35}$$

and after some rearrangement,

$$\left(\frac{dK}{dt}\right)_{\text{jump}} = \int dx dx'\, W(x\,|\,x')\{p_2(x',t)[\phi'\log[\phi/\phi'] - \phi + \phi']\}, \tag{3.7.36}$$

where

$$\phi = p_1(x,t)/p_2(x,t) \tag{3.7.37}$$

and ϕ' is similarly defined in terms of x'.

We now consider the simplest case. Suppose a stationary solution $p_s(x)$ exists which is nonzero everywhere, except at infinity, where it and its first derivative vanish. Then we may choose $p_2(x,t) = p_s(x)$. The contributions to dK/dt from (3.7.32) and the second term in (3.7.34) can be integrated to give surface terms which vanish at infinity so that we find

$$\left(\frac{dK}{dt}\right)_{\text{drift}} = 0 \tag{3.7.38a}$$

$$\left(\frac{dK}{dt}\right)_{\text{diff}} \leqslant 0 \tag{3.7.38b}$$

$$\left(\frac{dK}{dt}\right)_{\text{jump}} \leqslant 0, \tag{3.7.38c}$$

where the last inequality comes by setting $z = \phi'/\phi'$ in (3.7.29).

We must now consider under what situations the *equalities* in (3.7.38) are actually achieved. Inspection of (3.7.36) shows that this term will be zero if and only if $\phi = \phi'$ for almost all x and x' which are such that $W(x\,|\,x') \neq 0$. Thus, if $W(x\,|\,x')$ 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 $\phi(x)$ is independent of x, i.e.,

$$p_1(\boldsymbol{x}, t)/p_s(\boldsymbol{x}) = \text{constant}. \tag{3.7.39}$$

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

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,

$$\begin{aligned} &p_s(\boldsymbol{x}) \neq 0 \text{ with probability 1} \\ &W(\boldsymbol{x}|\boldsymbol{x}') \neq 0 \text{ with probability 1,} \\ &B_{ij}(\boldsymbol{x}) \text{ positive definite with probability 1,} \end{aligned} \tag{3.7.40}$$

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

The 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, $W(\boldsymbol{x}|\boldsymbol{x}') = 0$ if $\boldsymbol{x}$ and $\boldsymbol{x}'$ are not both in R_1 or R_2. Then it is possible to have $dK/dt = 0$ if

$$\begin{aligned} p_1(\boldsymbol{x}, t) &= \lambda_1 p_s(\boldsymbol{x}) \qquad x \in R_1 \\ &= \lambda_2 p_s(\boldsymbol{x}) \qquad x \in R_2 \end{aligned} \tag{3.7.41}$$

so that there is no unique stationary distribution. The two regions are disconnected and separate stochastic processes take place 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) $p_s(\boldsymbol{x})$ Vanishes in Some Definite Region

If we have

$$\begin{aligned} p_s(\boldsymbol{x}) &= 0 \qquad \boldsymbol{x} \in R_1 \\ &\neq 0 \qquad \boldsymbol{x} \in R_2 \end{aligned} \tag{3.7.42}$$

and again A_i and B_{ij} vanish, then it follows that, since $p_s(\boldsymbol{x})$ satisfies the stationary equation (3.7.26),

$$W(\boldsymbol{x}|\boldsymbol{y}) = 0 \qquad \boldsymbol{x} \in R_1, \boldsymbol{y} \in R_2. \tag{3.7.43}$$

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

$$\langle \boldsymbol{X}(t) | [\boldsymbol{x}_0, t_0] \rangle = \int d\boldsymbol{x}\, \boldsymbol{x}\, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0)\,, \tag{3.7.44}$$

then the autocorrelation matrix

$$\langle \boldsymbol{X}(t)\boldsymbol{X}(t_0)^{\mathrm{T}} \rangle = \int d\boldsymbol{x}\, d\boldsymbol{x}_0\, \boldsymbol{x}\boldsymbol{x}_0^{\mathrm{T}} p(\boldsymbol{x}, t; \boldsymbol{x}_0, t_0) \tag{3.7.45}$$

$$= \int d\boldsymbol{x}_0 \langle \boldsymbol{X}(t) | [\boldsymbol{x}_0, t_0] \rangle \boldsymbol{x}_0^{\mathrm{T}} p(\boldsymbol{x}_0, t_0)\,. \tag{3.7.46}$$

Thus we see that (3.7.44) defines the mean of $\boldsymbol{X}(t)$ under the condition that $\boldsymbol{X}$ had the value $\boldsymbol{x}_0$ at time t_0, and (3.7.46) tells us that the autocorrelation matrix is obtained by averaging this conditional average (multiplied by $\boldsymbol{x}_0^{\mathrm{T}}$) at time t_0. These results are true by 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, we can state that $\langle \boldsymbol{X}(t) | [\boldsymbol{x}_0, t_0] \rangle$ is a uniquely defined quantity, since the knowledge of $\boldsymbol{x}_0$ at time t_0 completely determines the future of the process. 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

$$p_s(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \ldots \boldsymbol{x}_n, t_n), \tag{3.7.47}$$

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

$$p_a(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \ldots \boldsymbol{x}_n, t_n) = p_s(\boldsymbol{x}_1, t_1, \boldsymbol{x}_2, t_2; \ldots \boldsymbol{x}_n, t_n | \boldsymbol{a}, 0). \tag{3.7.48}$$

Then for this process we note that

$$\langle \boldsymbol{X}(t) | [\boldsymbol{x}_0, t_0] \rangle_a = \int d\boldsymbol{x}\, \boldsymbol{x}\, p_s(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0; \boldsymbol{a}, 0) \tag{3.7.49}$$

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

$$\lim_{\tau \to \infty} p_s(\boldsymbol{x}, t + \tau | \boldsymbol{x}_0, t_0 + \tau; \boldsymbol{a}, 0) = p_s(\boldsymbol{x}, t - t_0 | \boldsymbol{x}_0, 0) \tag{3.7.50}$$

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

$$\langle \boldsymbol{X}(t) | [\boldsymbol{x}_0, t_0] \rangle_s = \lim_{\tau \to \infty} \langle \boldsymbol{X}(t + \tau) | [\boldsymbol{x}_0, t_0 + \tau] \rangle_a \tag{3.7.51}$$

and the stationary autocorrelation matrix is given by

$$\langle \boldsymbol{X}(t)\boldsymbol{X}(t_0)^{\mathrm{T}}\rangle_s = \int d\boldsymbol{x}_0\, \boldsymbol{x}_0^{\mathrm{T}}\langle \boldsymbol{X}(t)|[\boldsymbol{x}_0, t_0]\rangle_s p_s(\boldsymbol{x}_0) \tag{3.7.52}$$

$$= \lim_{\tau\to\infty} \langle \boldsymbol{X}(t+\tau)\boldsymbol{X}(t_0+\tau)^{\mathrm{T}}\rangle_a$$

$$= \lim_{\tau\to\infty} \int d\boldsymbol{x}_0\, \boldsymbol{x}_0^{\mathrm{T}}\langle \boldsymbol{x}(t+\tau)|[\boldsymbol{x}_0, t_0+\tau]\rangle_a p_a(\boldsymbol{x}_0, t_0+\tau)\,. \tag{3.7.53}$$

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

$$\text{Markov} \Longrightarrow \langle \boldsymbol{X}(t)|[\boldsymbol{x}_0, t_0]\rangle_s = \langle \boldsymbol{X}(t)|[\boldsymbol{x}_0, t_0]\rangle_a$$

$$= \langle \boldsymbol{X}(t)|[\boldsymbol{x}_0, t_0]\rangle\,. \tag{3.7.54}$$

Equation (3.7.46) 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 means, i.e.,

$$d\langle \boldsymbol{X}(t)|[\boldsymbol{x}_0, t_0]\rangle/dt = -A\langle \boldsymbol{X}(t)|[\boldsymbol{x}_0, t_0]\rangle \tag{3.7.55}$$

which is very often the case in systems of practical interest, either as an exact result or as an approximation. The initial conditions for (3.7.55) are clearly

$$\langle \boldsymbol{X}(t_0)|[\boldsymbol{x}_0, t_0]\rangle = \boldsymbol{x}_0\,. \tag{3.7.56}$$

Then from (3.7.50, 59)

$$\frac{d}{dt}\langle \boldsymbol{X}(t)\boldsymbol{X}(t_0)^{\mathrm{T}}\rangle = -A\langle \boldsymbol{X}(t)\boldsymbol{X}(t_0)^{\mathrm{T}}\rangle \tag{3.7.57}$$

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

$$\langle \boldsymbol{X}(t)\boldsymbol{X}(t_0)^{\mathrm{T}}\rangle - \langle \boldsymbol{X}(t)\rangle\langle \boldsymbol{X}(t_0)^{\mathrm{T}}\rangle = \langle \boldsymbol{X}(t), \boldsymbol{X}(t_0)^{\mathrm{T}}\rangle \tag{3.7.58}$$

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

$$dG(t)/dt = -A\,G(t) \tag{3.7.59}$$

and

$$G(0) = \sigma \tag{3.7.60}$$

or

$$G(t) = \exp[-At]\sigma \tag{3.7.61}$$

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 (3.7.55).

For non-Markov processes there is no simple procedure. We must carry out the complicated procedure implicit in (3.7.53).

3.8 Examples of Markov Processes

We present here for reference some fundamental solutions of certain cases of the differential Chapman-Kolmogorov equation. These will have a wide application throughout the remainder of this book.

3.8.1 The Wiener Process

This takes its name from N. Wiener who studied it extensively. From the point of view of this chapter, it is the solution of the Fokker-Planck equation as discussed in Sect.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

$$\frac{\partial}{\partial t} p(w, t \mid w_0, t_0) = \frac{1}{2} \frac{\partial^2}{\partial w^2} p(w, t \mid w_0, t_0) . \tag{3.8.1}$$

Utilising the initial condition

$$p(w, t_0 \mid w_0, t_0) = \delta(w - w_0) \tag{3.8.2}$$

on the conditional probability, we solve (3.8.1) by use of the characteristic function

$$\phi(s, t) = \int dw\, p(w, t \mid w_0, t_0) \exp(isw) \tag{3.8.3}$$

which satisfies

$$\frac{\partial \phi}{\partial t} = -\tfrac{1}{2} s^2 \phi \tag{3.8.4}$$

so that

$$\phi(s, t) = \exp\left[-\tfrac{1}{2} s^2 (t - t_0)\right] \phi(s, t_0) . \tag{3.8.5}$$

From (3.8.2), the initial condition is

$$\phi(s, t_0) = \exp(isw_0)$$

so that

$$\phi(s, t) = \exp\left[isw_0 - \tfrac{1}{2} s^2 (t - t_0)\right]. \tag{3.8.6}$$

Performing the Fourier inversion, we have the solution to (3.8.1):

$$p(w, t \mid w_0, t_0) = [2\pi(t - t_0)]^{-1/2} \exp\left[-(w - w_0)^2/2(t - t_0)\right] . \tag{3.8.7}$$

This represents a Gaussian, with

$$\langle W(t) \rangle = w_0 \tag{3.8.8}$$

$$\langle [W(t) - w_0]^2 \rangle = t - t_0 , \tag{3.8.9}$$

so that an initially sharp distribution spreads in time, as graphed in Fig.3.4.

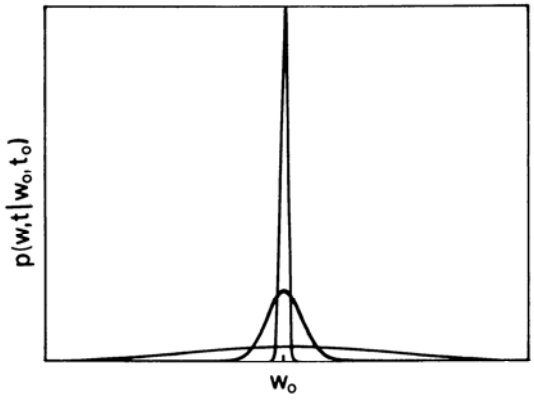

Fig. 3.4. Wiener process: spreading of an initially sharp distribution $p(w, t \mid w_0, t_0)$ with increasing time $t - t_0$

A multivariate Wiener process can be defined as

$$\boldsymbol{W}(t) = [W_1(t), W_2(t), \ldots , W_n(t)] \tag{3.8.10}$$

which satisfies the multivariable Fokker-Planck equation

$$\frac{\partial}{\partial t} p(\boldsymbol{w}, t \mid \boldsymbol{w}_0, t_0) = \frac{1}{2} \sum_i \frac{\partial^2}{\partial w_i^2} p(\boldsymbol{w}, t \mid \boldsymbol{w}_0, t_0) \tag{3.8.11}$$

whose solution is

$$p(\boldsymbol{w}, t \mid \boldsymbol{w}_0 , t_0) = [2\pi(t - t_0)]^{-n/2} \exp\left[-(\boldsymbol{w} - \boldsymbol{w}_0)^2/2(t - t_0)\right] , \tag{3.8.12}$$

a multivariate Gaussian with

$$\langle \boldsymbol{W}(t) \rangle = \boldsymbol{w}_0 \tag{3.8.13}$$

and

$$\langle [W_i(t) - w_{0i}] [W_j(t) - w_{0j}] \rangle = (t - t_0)\delta_{ij} . \tag{3.8.14}$$

The one-variable Wiener process is often simply called Brownian motion, since the Wiener process equation (3.8.1) is exactly the same as the differential equation of diffusion, shown by Einstein to be obeyed by Brownian motion, as we noted in Sect. 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 Fig. 3.5, we have given a few different sample paths with the same initial point to illustrate the extreme non-reproducibility of the paths.

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

$$\text{Prob}\{|[W(t+h) - W(t)]/h| > k\} \,. \tag{3.8.15}$$

From the solution for the conditional probability, this probability is

$$2 \int_{kh}^{\infty} dw (2\pi h)^{-1/2} \exp(-w^2/2h) \tag{3.8.16}$$

and in the limit as $h \to 0$ this is one. This means that no matter what value of k 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 Sect.3.5.2 and the simulated paths given in Fig. 3.5 illustrate in point dramatically. This corresponds, of course, to the well-known experimental fact that the Brownian particles have an exceedingly irregular motion. However, this clearly an idealisation, since if $W(t)$ represents the position

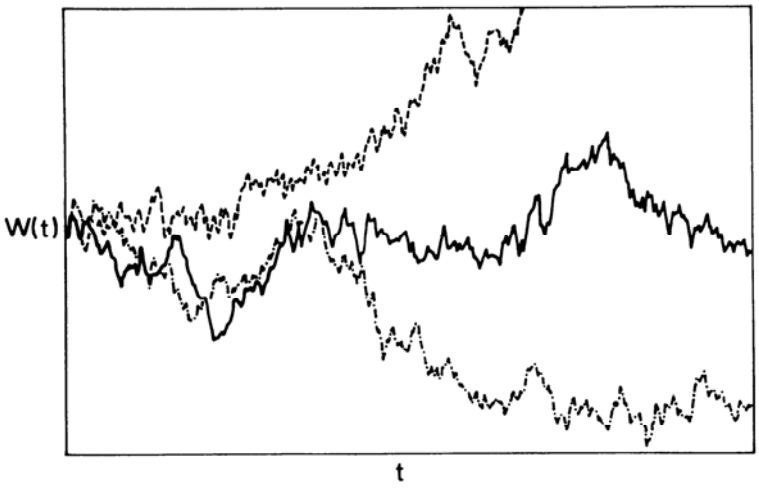

Fig. 3.5. Three simulated sample paths of the Wiener process, illustrating their great variability

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 (Sect. 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

$$p(w_n, t_n; w_{n-1}, t_{n-1}; w_{n-2}, t_{n-2}; \ldots ; w_0, t_0)$$
$$= \prod_{i=0}^{n-1} p(w_{i+1}, t_{i+1} \mid w_i, t_i) p(w_0, t_0) , \tag{3.8.17}$$

and using the explicit form of the conditional probabilities (3.8.7), we see that

$$p(w_n, t_n; w_{n-1}, t_{n-1}; w_{n-2}, t_{n-2}; \ldots; w_0, t_0)$$
$$= \prod_{i=0}^{n-1} \{[2\pi(t_{i+1} - t_i)]^{-1/2} \exp [-(w_{i+1} - w_i)^2/2(t_{i+1} - t_i)]\} p(w_0, t_0) . \tag{3.8.18}$$

If we define the variables

$$\Delta W_i \equiv W(t_i) - W(t_{i-1}) \tag{3.8.19}$$

$$\Delta t_i \equiv t_i - t_{i-1} , \tag{3.8.20}$$

then the joint probability density for the ΔW_i is

$$p(\Delta w_n; \Delta w_{n-1}; \Delta w_{n-2}; \ldots \Delta w_1; w_0)$$
$$= \prod_{i=1}^{n} \{(2\pi\Delta t_i)^{-1/2} \exp (-\Delta w_i^2/2\Delta t_i)\} p(w_0, t_0) \tag{3.8.21}$$

which shows from the definition of statistical independence given in Sect.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 Sect. 4.2.

d) Autocorrelation Functions

A quantity of great interest is the autocorrelation function, already discussed in Sects. 1.4.2 and 3.7.4. The formal definition is

$$\langle W(t)W(s) | [w_0, t_0] \rangle = \int dw_1 dw_2 \, w_1 w_2 p(w_1, t; w_2, s | w_0, t_0), \tag{3.8.22}$$

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

$$\langle W(t)W(s)|[w_0, t_0]\rangle = \langle [W(t) - W(s)]W(s)\rangle + \langle [W(s)]^2\rangle \,. \tag{3.8.23}$$

Using the independence of increments, the first average is zero and the second is given by (3.8.9) so that we have, in general,

$$\langle W(t)W(s)|[w_0, t_0]\rangle = \min(t - t_0, s - t_0) + w_0^2 \tag{3.8.24}$$

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

3.8.2 The Random Walk in One Dimension

This is a very famous problem, which is now considered classical. 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 a 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.

To do a Master equation treatment of the problem, we consider that the transition probability per unit time is given by the form

$$W(n + 1|n,t) = W(n - 1|n, t) = d; \tag{3.8.25}$$

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

$$\begin{aligned}\partial_t P(n, t|n', t') = d[&P(n + 1, t|n', t') + P(n - 1, t|n', t') \\ &- 2P(n, t|n't')].\end{aligned} \tag{3.8.26}$$

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 equaton, 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

$$\begin{aligned}P(n, (N + 1)\tau|n', N'\tau) = \tfrac{1}{2}\,[&P(n + 1, N\tau|n', N'\tau) \\ &+ P(n - 1, N\tau|n', N'\tau)]\,.\end{aligned} \tag{3.8.27}$$

If τ is small, we can view (3.8.26, 27) as approximations to each other by writing

$$P(n, (N + 1)\tau|n', N'\tau) \simeq P(n, N\tau|n', N'\tau) + \tau\partial_t P(n, t|n', t') \tag{3.8.28}$$

with $t = N\tau$, $t' = N'\tau$ and $d = \frac{1}{2}\tau^{-1}$, so that the transition probability per unit time in the Master equation model corresponds to half of the inverse waiting time τ in the discrete time model.

Both systems can be easily solved by introducing the characteristic function

$$G(s, t) = \langle e^{isn} \rangle = \sum_n P(n, t|n', t')e^{isn} \tag{3.8.29}$$

in which case the Master equation gives

$$\partial_t G(s, t) = d(e^{is} + e^{-is} - 2)G(s, t) \tag{3.8.30}$$

and the discrete time equation becomes

$$G(s, (N+1)\tau) = \tfrac{1}{2}(e^{is} + e^{-is})G(s, N\tau) . \tag{3.8.31}$$

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

$$G(s, 0) = 1 \tag{3.8.32}$$

in both cases, so that the solution to (3.8.30) is

$$G_1(s, t) = \exp[(e^{is} + e^{-is} - 2)td] , \tag{3.8.33}$$

and to (3.8.31)

$$G_2(s, N\tau) = [\tfrac{1}{2}(e^{is} + e^{-is})]^N \tag{3.8.34}$$

which can be written

$$G_2(s, t) = \left[1 + \frac{dt}{N}(e^{is} + e^{-is} - 2)\right]^N . \tag{3.8.35}$$

Using the usual exponential limit

$$\lim_{N\to\infty} \left(1 + \frac{a}{N}\right)^N = e^a , \tag{3.8.36}$$

we see that, provided s is sufficently small

$$\lim_{N\to\infty} G_2(s, t) = G_1(s, t) \tag{3.8.37}$$

which, by the properties of the characteristic function (v) in Sect.2.6, means the probability distributions approach each other.

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

$$P_1(n, t|0, 0) = e^{-2td} I_n(4td) \tag{3.8.38}$$

$$P_2(n, N\tau|0, 0) = (\tfrac{1}{2})^N N!\left[\left(\frac{N-n}{2}\right)!\left(\frac{N+n}{2}\right)!\right]^{-1}. \tag{3.8.39}$$

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.

The limit of continuous space is also of interest. If we set the distance travelled as

$$x = nl \tag{3.8.40}$$

so that the characteristic function of the distribution of x is

$$\phi_1(s, t) = \langle e^{isx}\rangle = G_1(ls, t)\exp[(e^{ils} + e^{-ils} - 2)td]\,. \tag{3.8.41}$$

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

$$\phi_1(s, t) \to \exp(-s^2tD)\,, \tag{3.8.42}$$

$$\text{where} \quad D = \lim_{l\to 0}(l^2d). \tag{3.8.43}$$

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

$$p(x, t|0, 0) = (4\pi Dt)^{-1/2}\exp(-x^2/4Dt) \tag{3.8.44}$$

and is of course the distribution for the Wiener process (Sect.3.8.1) or Brownian motion, as mentioned in Sect.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

$$l \to 0,\ \tau \to 0, \quad \text{with} \quad D = \lim_{l\to 0}(l^2/\tau) \tag{3.8.45}$$

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 (3.8.26) as a function of x up to second order in l gives

$$\partial_t p(x, t|0, 0) = (l^2d)\partial_x^2 p(x, t|0, 0)\,. \tag{3.8.46}$$

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 Sect.1.2 shows that he modelled 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 generalisations of these results to less specialised 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 (Sect.3.5.3) rather than the full Fokker-Planck equation. These results are presented in Sect.7.2.

3.8.3 Poisson Process

We have already noted the Poisson process in Sect.1.4.1. The process in which electrons arrive at an anode or customers arrive at a shop with a probability per unit time d of arriving, is governed by the Master equation for which

$$W(n+1\,|\,n,t) = d\,; \tag{3.8.47}$$

otherwise,

$$W(n\,|\,m,t) = 0\,. \tag{3.8.48}$$

This Master equation becomes

$$\partial_t P(n,t\,|\,n',t') = d[P(n-1,t\,|\,n',t') - P(n,t\,|\,n',t')] \tag{3.8.49}$$

and by comparison with (3.8.26) also represents a "one-sided" random walk, in which the walker steps to the right only with probability per unit time equal to d.

The characteristic function equation is similar to (3.8.30):

$$\partial_t G(s,t) = d[\exp(\mathrm{i}s) - 1]G(s,t) \tag{3.8.50}$$

with the solution

$$G(s,t) = \exp\{td[\exp(\mathrm{i}s) - 1] \tag{3.8.51}$$

for the initial condition that there are initially no customers (or electrons) at time $t = 0$, yielding

$$P(n,t\,|\,0,0) = \exp(-td)(td)^n/n!\,, \tag{3.8.52}$$

a Poisson distribution with mean given by

$$\langle N(t)\rangle = td\,. \tag{3.8.53}$$

In contrast to the random walk, the only limit that exists is $l \to 0$, with

$$dl \equiv v \tag{3.8.54}$$

held fixed and the limiting characteristic function is

$$\lim_{l\to 0}\{\exp[td(e^{ils}-1)]\} = \exp(itvs) \tag{3.8.55}$$

with the solution

$$p(x,t|0,0) = \delta(x-vt)\,. \tag{3.8.56}$$

We also see that in this limit, we would obtain from the Master equation (3.8.49) Liouville's equation, whose solution would be the deterministic motion we have derived.

We can do a slightly more refined analysis. We expand the characteristic function up to second order in s in the exponent and find

$$\phi(s,t) = G(ls,t) \simeq \exp[t(ivs - s^2D/2)] \tag{3.8.57}$$

where, as in the previous section,

$$D = l^2d\,.$$

This is the characteristic function of a Gaussian with variance Dt and mean vt, so that we now have

$$p(x,t|0,0) \simeq (2\pi Dt)^{-1/2}\exp[-(x-vt)^2/2Dt]\,. \tag{3.8.58}$$

It is also clear that this solution is the solution of

$$\partial_t p(x,t|0,0) = -v\,\partial_x p(x,t|0,0) + \tfrac{1}{2}D\,\partial_x^2 p(x,t|0,0 \tag{3.8.59}$$

which is obtained by expanding the Master equation (3.8.49) to order l^2, by writing

$$\begin{aligned} P(n-1,t|0,0) &= d\,p(x-l,t|0,0) \\ &\simeq d\,p(x,t|0,0) - ld\,\partial_x p(x,t|0,0) + \tfrac{1}{2}l^2 d\,\partial_x^2 p(x,t|0,0). \end{aligned} \tag{3.8.60}$$

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 (3.8.56). Effectively, the limit $l \to 0$ with well-defined v corresponds to $D = 0$. The kind of approximation just mentioned is a special case of van Kampen's system size expansion which we treat fully in Sect.7.2.3.

3.8.4 The Ornstein-Uhlenbeck Process

All the examples so far have 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

$$\partial_t p = \partial_x(kxp) + \tfrac{1}{2}D\,\partial_x^2 p\,, \tag{3.8.61}$$

where by p we mean $p(x, t \mid x_0, 0)$. This is the *Ornstein-Uhlenbeck process* [3.5]. The equation for the characteristic function

$$\phi(s) = \int_{-\infty}^{\infty} e^{isx} p(x, t \mid x_0, 0) dx \qquad \text{is} \tag{3.8.62}$$

$$\partial_t \phi + ks\partial_s \phi = -\tfrac{1}{2} Ds^2 \phi \,. \tag{3.8.63}$$

The method of characteristics can be used to solve this equation, namely, if

$$u(s, t, \phi) = a \qquad \text{and} \qquad v(s, t, \phi) = b \tag{3.8.64}$$

are two integrals of the subsidary equation (with a and b arbitrary const)

$$\frac{dt}{1} = \frac{ds}{ks} = -\frac{d\phi}{\frac{1}{2}Ds^2\phi}, \tag{3.8.65}$$

then a general solution of (3.8.63) is given by

$$f(u, v) = 0.$$

The particular integrals are readily found by integrating the equation involving dt and ds and that involving ds and $d\phi$; they are

$$u(s, t, \phi) = s \exp(-kt) \qquad \text{and} \tag{3.8.66}$$

$$v(s, t, \phi) = \phi \exp(Ds^2/4k), \tag{3.8.67}$$

and 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

$$\phi(s, t) = \exp(-Ds^2/4k) g[s \exp(-kt)] \tag{3.8.68}$$

The boundary condition

$$p(x, 0 \mid x_0, 0) = \delta(x - x_0) \tag{3.8.69}$$

clearly requires

$$\phi(s, 0) = \exp(ix_0 s) \tag{3.8.70}$$

and gives

$$g(s) = \exp(Ds^2/4k + ix_0 s),$$

and hence,

$$\phi(s, t) = \exp\left[\frac{-Ds^2}{4k}(1 - e^{-2kt}) + isx_0 e^{-kt}\right] \tag{3.8.71}$$

which, from Sect.2.8.1, corresponds to a Gaussian with

$$\langle X(t)\rangle = x_0 \exp(-kt) \tag{3.8.72}$$

$$\text{var}\{X(t)\} = \frac{D}{2k}[1 - \exp(-2kt)]\,. \tag{3.8.73}$$

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

$$\partial_x\left[kxp + \tfrac{1}{2}D\partial_x p\right] = 0 \tag{3.8.74}$$

and integrating once, we find

$$\left[kxp + \tfrac{1}{2}D\partial_x p\right]_{-\infty}^{x} = 0. \tag{3.8.75}$$

The requirement that p vanish at $-\infty$ together with its derivative, is necessary for normalisation. Hence, we have

$$\frac{1}{p}\partial_x p = -\frac{2kx}{D} \tag{3.8.76}$$

$$\text{so that} \quad p_s(x) = (\pi D/k)^{-1/2} \exp(-kx^2/D) \tag{3.8.77}$$

which 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 unnormalisable solution.

Time Correlation Functions for the Ornstein-Uhlenbeck Process. 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 definition

$$\langle X(t)X(s)|[x_0, t_0]\rangle = \iint dx_1 dx_2\, x_1 x_2\, p(x_1, t; x_2, s|x_0, t_0) \tag{3.8.78}$$

and using the Markov property

$$= \iint dx_1 dx_2\, x_1 x_2\, p(x_1, t|x_2, s)p(x_2, s|x_0, t_0) \tag{3.8.79}$$

on the assumption that

$$t \geqslant s \geqslant t_0\,. \tag{3.8.80}$$

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 Sect. 3.7.2. Letting $t_0 \to -\infty$, we find

$$\lim_{t_0 \to -\infty} p(x_2, s | x_0, t_0) = p_s(x_2) = (\pi D/k)^{-1/2} \exp(-kx_2^2/D) . \tag{3.8.81}$$

and by straightforward substitution and integration and noting that the stationary mean is zero, we get

$$\langle X(t)X(s)\rangle_s = \langle \overset{*}{X}(t), X(s)\rangle_s = \frac{D}{2k} \exp(-k|t-s|) . \tag{3.8.82}$$

This result demonstrates the general property of stationary processes: that the correlation functions depend only on time differences. It is also a general result [3.6] that the process we have described in this section is the only stationary Gaussian Markov process in one real variable.

The results of this subsection are very easily obtained by the stochastic differential equation methods which will be developed in Chap.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

$$|t-s| \sim 1/k \equiv \tau . \tag{3.8.83}$$

(More precisely, τ, known as the *correlation time* can be defined for arbitrary processes $X(s)$ by

$$\tau = \int_0^\infty dt \, \langle X(t), X(0)\rangle_s / \mathrm{var}\{X\}_s , \tag{3.8.84}$$

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

$$\left.\begin{aligned} \partial_t P(a, t | x, t_0) &= -\lambda P(a, t | x, t_0) + \mu P(b, t | x, t_0) \\ \partial_t P(b, t | x, t_0) &= \lambda P(a, t | x, t_0) - \mu P(b, t | x, t_0) \end{aligned}\right\} \tag{3.8.85}$$

for which the solution can simply be found by noting that

$$P(a, t | x, t_0) + P(b, t | x, t_0) = 1$$

and that a simple equation can be derived for $\lambda P(a, t|x, t_0) - \mu P(b, t|x, t_0)$, whose solution is, because of the initial condition

$$P(x', t_0|x, t_0) = \delta_{x,x'}, \tag{3.8.86}$$

$$\lambda P(a, t|x, t_0) - \mu P(b, t|x, t_0) = \exp[-(\lambda + \mu)(t - t_0)](\lambda\delta_{a,x} - \mu\delta_{b,x}) \tag{3.8.87}$$

so that

$$\left.\begin{aligned} P(a, t|x, t_0) &= \frac{\mu}{\lambda+\mu} + \exp[-(\lambda+\mu)(t - t_0)]\left(\frac{\lambda}{\lambda+\mu}\delta_{a,x} - \frac{\mu}{\lambda+\mu}\delta_{b,x}\right) \\ P(b, t|x, t_0) &= \frac{\lambda}{\lambda+\mu} - \exp[-(\lambda+\mu)(t - t_0)]\left(\frac{\lambda}{\lambda+\mu}\delta_{a,x} - \frac{\mu}{\lambda+\mu}\delta_{b,x}\right) \end{aligned}\right\} \tag{3.8.88}$$

This process clearly has the stationary solution obtained by letting $t_0 \to -\infty$:

$$\left.\begin{aligned} P_s(a) &= \frac{\mu}{\lambda+\mu} \\ P_s(b) &= \frac{\lambda}{\lambda+\mu} \end{aligned}\right\} \tag{3.8.89}$$

which is, of course, obvious from the Master equation.

The mean of $X(t)$ and its variance are straightforwardly computed:

$$\begin{aligned} \langle X(t)|[x_0, t_0]\rangle &= \sum xP(x, t|x_0, t_0) \\ &= \frac{a\mu + b\lambda}{\mu + \lambda} + \exp[-(\lambda+\mu)(t - t_0)]\left(x_0 - \frac{a\mu + b\lambda}{\mu + \lambda}\right) \end{aligned} \tag{3.8.90}$$

so that

$$\langle X\rangle_s = \frac{a\mu + b\lambda}{\mu + \lambda}. \tag{3.8.91}$$

The variance can also be computed but is a very messy expression. The stationary variance is easily computed to be

$$\mathrm{var}\{X\}_s = \frac{(a - b)^2\mu\lambda}{(\lambda + \mu)^2}. \tag{3.8.92}$$

To compute the stationary time correlation function, we write (assuming $t \geqslant s$)

$$\langle X(t)X(s)\rangle_s = \sum_{xx'} xx'P(x, t|x', s)P_s(x') \tag{3.8.93}$$

$$= \sum_{x'} x'\langle X(t)|[x', s]\rangle P_s(x'). \tag{3.8.94}$$

we now use (3.8.90–92) to obtain

$$\langle X(t)X(s)\rangle_s = \langle X\rangle_s^2 + \exp[-(\lambda+\mu)(t-s)](\langle X^2\rangle_s - \langle X\rangle_s^2) \tag{3.8.95}$$

$$= \left(\frac{a\mu + b\lambda}{\mu+\lambda}\right)^2 + \exp[-(\lambda+\mu)(t-s)]\frac{(a-b)^2\mu\lambda}{(\lambda+\mu)^2}. \tag{3.8.96}$$

Hence,

$$\langle X(t), X(s)\rangle_s = \langle X(t)X(s)\rangle_s - \langle X\rangle_s^2 = \exp[-(\lambda+\mu)|t-s|]\frac{(a-b)^2\mu\lambda}{(\lambda+\mu)^2}. \tag{3.8.97}$$

Notice that this time correlation function is of 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.