MARKOV PROCESSES

AN INTRODUCTION FOR PHYSICAL SCIENTISTS

Daniel T. Gillespie

Research Department
Naval Weapons Center
China Lake, California

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Boston San Diego New York
London Sydney Tokyo Toronto

So the corresponding F_i functions in Eqs. (1.8-14) are

$$F_1(s) \equiv \int_0^s Q_1(s')\,ds' = 1 - \exp(-s^2/2a^2),$$

$$F_2(\theta;s) \equiv \int_0^\theta Q_2^{(1)}(\theta'|s)\,d\theta' = \frac{\theta}{2\pi}.$$

Now we are in a position to make use of Eqs. (1.8-15). Specifically, letting r_1 and r_2 be two independent unit uniform random numbers, we first set $F_1(s) = r_1$ and solve for s, and we then set $F_2(\theta;s) = r_2$ and solve for θ. The results, after replacing r_1 by the equally serviceable unit uniform random number $1 - r_1$, are

$$s = a[2\ln(1/r_1)]^{1/2}, \quad \theta = 2\pi r_2. \tag{1.8-16a}$$

Then taking

$$x_1 = m + s\cos\theta, \quad x_2 = m + s\sin\theta, \tag{1.8-16b}$$

in accordance with the formulas relating X_1 and X_2 to S and Θ, we finally obtain the desired simultaneous sample values x_1 and x_2 of the statistically independent normal random variables X_1 and X_2. So, although we cannot analytically transform a *single* unit uniform random number r into a sample value x of $N(m,a^2)$, Eqs. (1.8-16) provide a simple recipe for analytically transforming a statistically independent *pair* of unit uniform random numbers r_1 and r_2 into a statistically independent *pair* of sample values x_1 and x_2 of $N(m,a^2)$. We shall find frequent use for the generating algorithm (1.8-16) in later chapters of this book.

- 2 -

GENERAL FEATURES OF A MARKOV PROCESS

In this chapter we shall use the concepts of random variable theory set forth in Chapter 1 to define a *Markov process* and broadly frame its fundamental properties. This will involve introducing the key functions that are used to describe Markov processes, as well as deriving some general equations which those functions must obey. We shall quickly discover that any substantive characterization of a Markov process requires that we specify the form of what is called the *propagator density function*. In Chapters 3 and 4 we shall consider two different ways of specifying that critical function, which specifications lead to the two principal classes of Markov processes called "continuous" and "jump." In the present chapter we shall see how the propagator density function comes to play its pivotal role in Markov process theory, and we shall develop that theory as fully as we can without committing ourselves to a specific form for the propagator density function.

2.1 THE MARKOV STATE DENSITY FUNCTION

We consider a time-evolving or "dynamical" system whose possible states can be represented by points on the real axis, and we let

$$X(t) \equiv \text{the state point, or state, of the system at time } t. \tag{2.1-1}$$

We shall assume that the value of X at some initial time t_0 is fixed,

$$X(t_0) = x_0, \tag{2.1-2}$$

but that $X(t)$ for any $t > t_0$ can be predicted only probabilistically; more specifically, we assume that $X(t)$ for any given $t > t_0$ is a *random variable*, as defined in Section 1.2. Since it makes sense to inquire about the state

of the system at successive instants $t_1, t_2, \ldots, t_n$, where $t_0 < t_1 < t_2 < \ldots < t_n$, then we can ascribe to the corresponding n random variables $X(t_1), X(t_2), \ldots, X(t_n)$ a *joint density function* $P_n^{(1)}$, which is defined as follows:

$$P_n^{(1)}(x_n,t_n; x_{n-1},t_{n-1}; \ldots; x_1,t_1 \mid x_0,t_0)\, dx_n\, dx_{n-1} \cdots dx_1$$
$$\equiv \text{Prob}\{ X(t_i) \in [x_i, x_i + dx_i) \text{ for } i = 1,2,\ldots,n,$$
$$\text{given that } X(t_0) = x_0, \text{ with } t_0 \le t_1 \le \ldots \le t_n \}. \quad (2.1\text{-}3)$$

If all these assumptions are satisfied, then we say that $X(t)$ is a **stochastic process**.

It is evident that a stochastic process $X(t)$ has infinitely many joint density functions $P_n^{(1)}$, corresponding to $n = 1, 2, \ldots$. And associated with each of these joint density functions is a plethora of subordinate density functions; for example,

$$P_{n-j}^{(j+1)}(x_n,t_n; \ldots; x_{j+1},t_{j+1} \mid x_j,t_j; \ldots; x_1,t_1; x_0,t_0),$$

is defined to be the joint density function of the $n-j$ random variables $X(t_{j+1}), \ldots, X(t_n)$ given the $j+1$ conditions $X(t_0) = x_0$, $X(t_1) = x_1$, $\ldots$, $X(t_j) = x_j$. Notice that the *subscript* on the density function $P_k^{(j)}$ refers to the number of (x,t) pairs to the *left* of the "given" bar, while the *superscript* refers to the number of (x,t) pairs to the *right* of the "given" bar; thus, $P_k^{(j)}$ is a k-variate joint density function with j conditionings.

It is always possible to calculate the function $P_{n-1}^{(1)}$ from the function $P_n^{(1)}$ by simply integrating the latter over any one of the variables $x_1, \ldots, x_n$. However, it is *not* in general possible to deduce the function $P_{n+1}^{(1)}$ from the function $P_n^{(1)}$. This "open-ended" nature of the density functions for a general stochastic process usually makes any substantive analysis extremely difficult. But we shall be concerned here with only a very restricted subclass of stochastic processes, namely those that have the "past-forgetting" property that, for all $j \ge 2$ and $t_{i-1} \le t_i$,

$$P_1^{(j)}(x_j,t_j \mid x_{j-1},t_{j-1}; \ldots; x_1,t_1; x_0,t_0)$$
$$= P_1^{(1)}(x_j,t_j \mid x_{j-1},t_{j-1}) \equiv P(x_j,t_j \mid x_{j-1},t_{j-1}). \quad (2.1\text{-}4)$$

This is called the **Markov property**, and it says that *only the most recent conditioning matters*: Given that $X(t') = x'$, then our ability to predict $X(t)$ for any $t > t'$ will not be enhanced by a knowledge of any values of the process earlier than t'. Any stochastic process $X(t)$ that has this past-forgetting property is called a *Markovian* stochastic process, or more simply, a **Markov process**. In what follows it may always be assumed, unless explicitly stated otherwise, that the stochastic process $X(t)$ under consideration is a Markov process.

The Markov property (2.1-4) breaks the open-endedness of the hierarchy of joint state density functions in a dramatic way. For the joint density function $P_2^{(1)}$ we have

$$P_2^{(1)}(x_2,t_2; x_1,t_1 \mid x_0,t_0)$$
$$= P_1^{(1)}(x_1,t_1 \mid x_0,t_0)\, P_1^{(2)}(x_2,t_2 \mid x_1,t_1; x_0,t_0) \quad \text{[by (1.5-9d)]}$$
$$= P_1^{(1)}(x_1,t_1 \mid x_0,t_0)\, P_1^{(1)}(x_2,t_2 \mid x_1,t_1). \quad \text{[by (2.1-4)]}$$

Hence, writing $P_1^{(1)} \equiv P$ in accordance with the notation suggested in Eq. (2.1-4), we have

$$P_2^{(1)}(x_2,t_2; x_1,t_1 \mid x_0,t_0) = P(x_2,t_2 \mid x_1,t_1)\, P(x_1,t_1 \mid x_0,t_0). \quad (2.1\text{-}5)$$

The same kind of reasoning shows that

$$P_3^{(1)}(x_3,t_3; x_2,t_2; x_1,t_1 \mid x_0,t_0) = P(x_3,t_3 \mid x_2,t_2)\, P(x_2,t_2 \mid x_1,t_1)\, P(x_1,t_1 \mid x_0,t_0),$$

and more generally, for any set of times $t_n \ge t_{n-1} \ge \cdots \ge t_0$,

$$P_n^{(1)}(x_n,t_n; \ldots; x_1,t_1 \mid x_0,t_0) = \prod_{i=1}^{n} P(x_i,t_i \mid x_{i-1},t_{i-1}). \quad (2.1\text{-}6)$$

So for a *Markov* process, every conditioned state density function $P_n^{(1)}$ can be written solely in terms of the particular conditioned state density function $P_1^{(1)} \equiv P$. The function $P_1^{(1)} \equiv P$ thus becomes the principle focus of our study, and we shall henceforth refer to it as the **Markov state density function**. For future reference, the formal definition of the Markov state density function is [cf. Eq. (2.1-3)]

$$P(x_2,t_2 \mid x_1,t_1)\, dx_2$$
$$\equiv \text{Prob}\{ X(t_2) \in [x_2, x_2 + dx_2), \text{ given } X(t_1) = x_1, \text{ with } t_2 \ge t_1 \}. \quad (2.1\text{-}7)$$

2.2 THE CHAPMAN-KOLMOGOROV EQUATION

Since $P(x_2,t_2 \mid x_1,t_1)$ is a density function with respect to its argument x_2, it must satisfy conditions analogous to Eqs. (1.2-3) and (1.2-4), namely,

$$P(x_2,t_2 \mid x_1,t_1) \geq 0, \tag{2.2-1}$$

$$\int_{-\infty}^{\infty} dx_2\, P(x_2,t_2 \mid x_1,t_1) = 1. \tag{2.2-2}$$

Also, if we let t_2 *equal* its minimum value t_1 then the condition $X(t_1)=x_1$ obviously implies that $X(t_2)=x_1$, or that $X(t_2)$ is the *sure* variable x_1; hence, by (1.2-7), we have the requirement

$$P(x_2,\, t_2=t_1 \mid x_1,t_1) = \delta(x_2-x_1). \tag{2.2-3}$$

Of course, the function P would have to satisfy the preceding three equations even if the process $X(t)$ were not Markovian. A condition upon P that arises specifically because of the Markov property may be deduced as follows: For any three times $t_1 \leq t_2 \leq t_3$, we have

$$\begin{aligned} P_1^{(1)}(x_3,t_3 \mid x_1,t_1) &= \int_{-\infty}^{\infty} dx_2\, P_2^{(1)}(x_3,t_3;\, x_2,t_2 \mid x_1,t_1) && \text{[by (1.5-12a)]} \\ &= \int_{-\infty}^{\infty} dx_2\, P_1^{(1)}(x_2,t_2 \mid x_1,t_1)\, P_1^{(2)}(x_3,t_3 \mid x_2,t_2;\, x_1,t_1) && \text{[by (1.5-9d)]} \\ &= \int_{-\infty}^{\infty} dx_2\, P_1^{(1)}(x_2,t_2 \mid x_1,t_1)\, P_1^{(1)}(x_3,t_3 \mid x_2,t_2), && \text{[by (2.1-4)]} \end{aligned}$$

where the last step explicitly invokes the assumed Markovian nature of $X(t)$. Interchanging the two factors in the last integrand and then abbreviating $P_1^{(1)}$ everywhere by P, we obtain what is known as the **Chapman-Kolmogorov equation**:

$$P(x_3,t_3 \mid x_1,t_1) = \int_{-\infty}^{\infty} P(x_3,t_3 \mid x_2,t_2)\, P(x_2,t_2 \mid x_1,t_1)\, dx_2 \qquad (t_1 \leq t_2 \leq t_3). \tag{2.2-4}$$

This integral equation, a graphical interpretation of which is given in Fig. 2-1, is essentially a *consistency condition* on the Markov state density function P for any Markov process $X(t)$. As we shall see later, it severely limits the range of acceptable functional forms of P for any Markov process.

Because of the initial condition (2.1-2), the quantity $P(x,t \mid x_0,t_0)$ is of special importance. As we shall see later, time-evolution equations for $P(x,t \mid x_0,t_0)$ can be derived from two specially phrased versions of the Chapman-Kolmogorov equation, obtained by relabeling the variables x_i

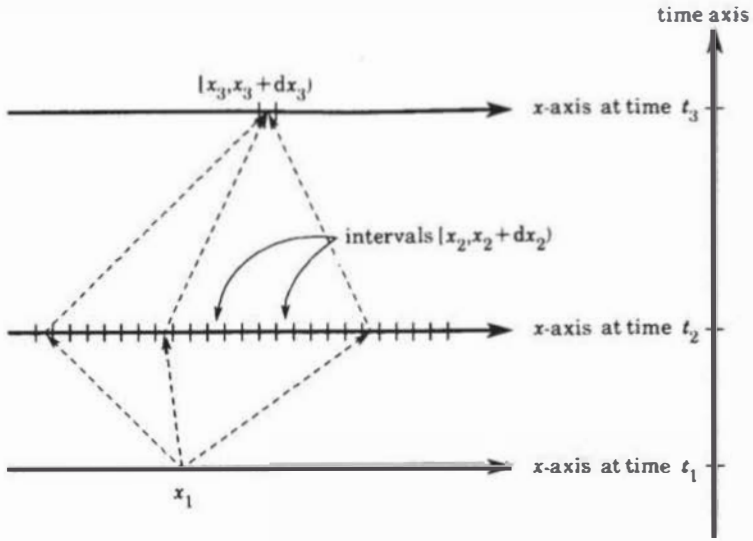

Figure 2-1. Graphical interpretation of the Chapman-Kolmogorov equation (2.2-4). The probability $P(x_3,t_3 \mid x_1,t_1)dx_3$ of going from x_1 at time t_1 to the interval $[x_3,x_3+dx_3)$ at time t_3 can always be written as the sum of the probabilities of this occurring via all possible intervals $[x_2,x_2+dx_2)$ at any fixed intermediate time t_2. When the Markov property (2.1-4) holds, the summand takes the form $[P(x_2,t_2 \mid x_1,t_1)dx_2]\times[P(x_3,t_3 \mid x_2,t_2)dx_3]$, whence the integral equation (2.2-4) for the function P.

and t_i according to the two diagrams in Fig. 2-2. These two "utility" versions of the Chapman-Kolmogorov equation are as follows:

$$P(x,t+\Delta t \mid x_0,t_0) = \int_{-\infty}^{\infty} d\xi\, P(x,t+\Delta t \mid x-\xi,t)\, P(x-\xi,t \mid x_0,t_0) \qquad (t_0<t<t+\Delta t); \tag{2.2-5a}$$

$$P(x,t \mid x_0,t_0) = \int_{-\infty}^{\infty} d\xi\, P(x,t \mid x_0+\xi,t_0+\Delta t_0)\, P(x_0+\xi,t_0+\Delta t_0 \mid x_0,t_0) \qquad (t_0<t_0+\Delta t_0<t). \tag{2.2-5b}$$

Equation (2.2-5a) will give rise to equations that govern the x and t behavior of $P(x,t \mid x_0,t_0)$ for fixed x_0 and t_0, the so-called "forward" time-evolution equations. And Eq. (2.2-5b) will give rise to equations that govern the x_0 and t_0 behavior of $P(x,t \mid x_0,t_0)$ for fixed x and t, the so-called "backward" time-evolution equations. Notice that Δt and Δt_0 here need *not* be infinitesimally small.

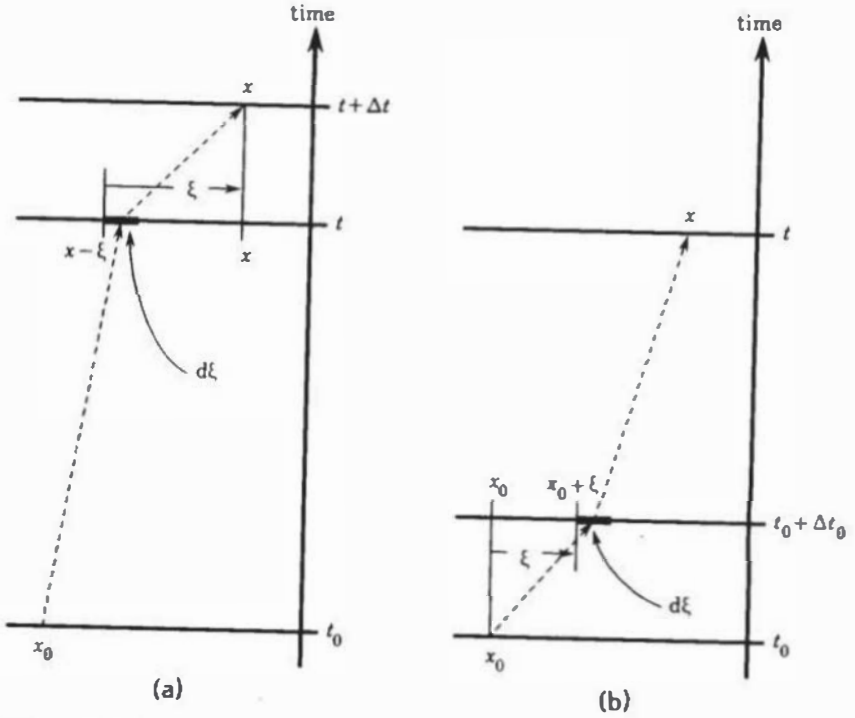

Figure 2-2. Showing the initial, intermediate and final state variables for the two special Chapman-Kolmogorov equations (2.2-5a) and (2.2-5b).

Another version of the Chapman-Kolmogorov equation that occasionally proves useful for determining $P(x,t\,|\,x_0,t_0)$ is obtained as follows: By an obvious extension of Eq. (1.5-12a), we have for any set of times $t_0<t_1<\cdots<t_n$,

$$P(x_n,t_n\,|\,x_0,t_0)=\int_{-\infty}^{\infty}dx_{n-1}\cdots\int_{-\infty}^{\infty}dx_1\;P_n^{(1)}(x_n,t_n;\ldots;x_1,t_1\,|\,x_0,t_0).$$

Substituting from Eq. (2.1-6) on the right then gives the result

$$P(x_n,t_n\,|\,x_0,t_0)=\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_{n-1}\prod_{i=1}^{n}P(x_i,t_i\,|\,x_{i-1},t_{i-1})$$
$$(t_0<t_1<\ldots<t_n).\quad(2.2\text{-}6)$$

This equation, which evidently reduces to Eq. (2.2-4) for $n=2$, will be referred to as the **compounded Chapman-Kolmogorov equation.**

2.3 FUNCTIONS OF STATE AND THEIR AVERAGES

Any univariate function g that takes a value $g(x)$ for any possible system state x can be regarded as a **function of state.** We can also think of a function of state as a *random variable* $g(X(t))$, which is *defined* to be "the function g of the random variable $X(t)$" [see the discussion of Eqs. (1.6-1)]. **The conditional state average,**

$$\langle g(X(t))\,|\,X(t')=x'\rangle=\int_{-\infty}^{\infty}dx\,g(x)\,P(x,t\,|\,x',t')\quad(t'\le t),\quad(2.3\text{-}1)$$

may accordingly be viewed *either* as (i) the average of the function g with respect to the random variable $X(t)$ given that $X(t')=x'$, *or* as (ii) the mean of the random variable $g(X(t))$ given that $X(t')=x'$ [see the discussion following Eq. (1.6-12)].

Similarly, any bivariate function g can be used to define a **two-time function of state** $g(X(t_1),X(t_2))$ for $t_1\le t_2$, and its average or mean, given $X(t')=x'$ for some $t'\le t_1$, is

$$\langle g(X(t_1),X(t_2))\,|\,X(t')=x'\rangle=\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dx_2\,g(x_1,x_2)\,P_2^{(1)}(x_2,t_2;x_1,t_1\,|\,x',t').$$

With the Markov simplification (2.1-5), this is

$$\langle g(X(t_1),X(t_2))\,|\,X(t')=x'\rangle$$
$$=\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dx_2\;g(x_1,x_2)\,P(x_2,t_2\,|\,x_1,t_1)\,P(x_1,t_1\,|\,x',t')$$
$$(t'\le t_1\le t_2)\quad(2.3\text{-}2)$$

If t' in Eqs. (2.3-1) and (2.3-2) is taken to be the initial time t_0, then we shall simply omit the conditioning notation; thus, the initially conditioned averages will be denoted by

$$\langle g(X(t))\,|\,X(t_0)=x_0\rangle\equiv\langle g(X(t))\rangle$$
$$=\int_{-\infty}^{\infty}dx\,g(x)\,P(x,t\,|\,x_0,t_0)\quad(t_0\le t),\quad(2.3\text{-}3)$$

and

$$\langle g(X(t_1),X(t_2))\,|\,X(t_0)=x_0\rangle\equiv\langle g(X(t_1),X(t_2))\rangle$$
$$=\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dx_2\;g(x_1,x_2)\,P(x_2,t_2\,|\,x_1,t_1)\,P(x_1,t_1\,|\,x_0,t_0)$$
$$(t_0\le t_1\le t_2).\quad(2.3\text{-}4)$$

The most important applications of the univariate average (2.3-3) are to the two functions $g(x)=x$ and $g(x)=x^2$. This is because [see Eqs. (1.3-6) and (1.3-8)] the initially conditioned mean and variance of the process $X(t)$ are given respectively by

$$\text{mean}\{X(t)\} = \langle X(t)\rangle \quad (t_0 \le t) \tag{2.3-5}$$

and

$$\text{var}\{X(t)\} = \langle X^2(t)\rangle - \langle X(t)\rangle^2 \quad (t_0 \le t). \tag{2.3-6}$$

And the initially conditioned standard deviation of $X(t)$ is given by

$$\text{sdev}\{X(t)\} = [\text{var}\{X(t)\}]^{1/2} = [\langle X^2(t)\rangle - \langle X(t)\rangle^2]^{1/2} \quad (t_0 \le t). \tag{2.3-7}$$

Given that $X(t_0)=x_0$, then at any time $t>t_0$ we can "usually expect" to find $X(t)$ to be within roughly $\text{sdev}\{X(t)\}$ of the value $\langle X(t)\rangle$.

The most important application of the bivariate average (2.3-4) is to the function $g(x_1,x_2)=x_1x_2$. This is because [see Eq. (1.5-17)] the covariance of $X(t_1)$ and $X(t_2)$ for $t_0 \le t_1 \le t_2$, given that $X(t_0)=x_0$, is defined by

$$\text{cov}\{X(t_1),X(t_2)\} \equiv \langle X(t_1)X(t_2)\rangle - \langle X(t_1)\rangle\langle X(t_2)\rangle \quad (t_0 \le t_1 \le t_2). \tag{2.3-8}$$

As shown in Eq. (1.6-18), $\text{cov}\{X(t_1),X(t_2)\}$ is always bounded in absolute value by $\text{sdev}\{X(t_1)\}\cdot\text{sdev}\{X(t_2)\}$. If $\text{cov}\{X(t_1),X(t_2)\}$ assumes its positive bound then $X(t_1)$ and $X(t_2)$ are said to be *maximally correlated*, while if $\text{cov}\{X(t_1),X(t_2)\}$ assumes its negative bound then $X(t_1)$ and $X(t_2)$ are said to be *maximally anti-correlated*. If $\text{cov}\{X(t_1),X(t_2)\}=0$, then $X(t_1)$ and $X(t_2)$ are said to be *uncorrelated*. One way in which $X(t_1)$ and $X(t_2)$ can be uncorrelated is for them to be *statistically independent*, in the sense that $P(x_2,t_2\,|\,x_1,t_1)$ is independent of x_1 (and hence also independent of t_1); because in that case we have from Eq. (2.3-4),

$$\begin{aligned}\langle X(t_1)X(t_2)\rangle &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, x_1x_2 P(x_2,t_2\,|\,x_1,t_1)\,P(x_1,t_1\,|\,x_0,t_0)\\ &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, x_1x_2 P(x_2,t_2)\,P(x_1,t_1)\\ &= \int_{-\infty}^{\infty} dx_1\, x_1 P(x_1,t_1) \int_{-\infty}^{\infty} dx_2\, x_2 P(x_2,t_2)\\ &= \langle X(t_1)\rangle\langle X(t_2)\rangle.\end{aligned}$$

So if $X(t_1)$ and $X(t_2)$ are statistically independent, then $\text{cov}\{X(t_1),X(t_2)\}$ vanishes, implying by definition that $X(t_1)$ and $X(t_2)$ are uncorrelated. But notice that $X(t_1)$ and $X(t_2)$ could be uncorrelated (i.e., have a vanishing covariance) *without* being statistically independent.

It is an instructive exercise to show directly for a Markov process that

$$\text{cov}\{X(t_1),X(t_2=t_1)\} = \text{var}\{X(t_1)\}, \tag{2.3-9}$$

as we should expect on the basis of Eq. (1.5-18). We have from Eq. (2.3-4),

$$\begin{aligned}\langle X(t_1)X(t_2=t_1)\rangle &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, x_1x_2 P(x_2,t_2=t_1\,|\,x_1,t_1)\,P(x_1,t_1\,|\,x_0,t_0)\\ &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, x_1x_2\, \delta(x_2-x_1)\,P(x_1,t_1\,|\,x_0,t_0)\\ &= \int_{-\infty}^{\infty} dx_1\, x_1x_1\,P(x_1,t_1\,|\,x_0,t_0) = \langle X^2(t_1)\rangle.\end{aligned}$$

where the second equality follows from Eq. (2.2-3). Substituting this result into Eq. (2.3-8) with $t_2=t_1$, and then recalling the definition of the variance in Eq. (2.3-6), we obtain Eq. (2.3-9).

2.4 THE MARKOV PROPAGATOR

Suppose that our system — our Markov process — is in the state x at time t; i.e., suppose that $X(t)=x$. Then by the infinitesimally later time $t+dt$, the system will have evolved to some new state that is *displaced* from x by the amount

$$\Xi(dt;x,t) \equiv X(t+dt) - X(t), \text{ given } X(t)=x. \tag{2.4-1}$$

Notice that the condition "$X(t)=x$" in this definition affects the *first* term on the right side as well as the second. This state displacement $\Xi(dt;x,t)$ from state x during time $[t,t+dt)$ is clearly a *random variable*; we shall call it the propagator of the process $X(t)$. Like any random variable, the propagator is completely specified by its density function. We shall denote the density function of $\Xi(dt;x,t)$ by $\Pi(\xi\,|\,dt;x,t)$, and refer to it as the propagator density function; thus we have, by definition,

$$\Pi(\xi\,|\,dt;x,t)\,d\xi = \text{Prob}\{\Xi(dt;x,t)\in[\xi,\xi+d\xi)\}. \tag{2.4-2}$$

Evidently, the propagator $\Xi(dt;x,t)$ tells us where the process, in state x at time t, will be at the infinitesimally later time $t+dt$; specifically, the process will be in the state $x+\Xi(dt;x,t)$. Here it is perhaps appropriate to remark that we are regarding dt as a real variable whose allowed range is the open interval $(0,\varepsilon)$, where ε is positive but "arbitrarily close to zero." Although we shall always take $\varepsilon \ll 1$, so that $(dt)^2$ is negligibly small

compared to dt, the precise value of ε will depend upon the situation. For example, if we are considering a specific differentiable function h of the variable t, then we can always find a positive number ε (which may depend upon t) such that $h(t+dt)$ can for most purposes be equated to $h(t)+h'(t)dt$ for all dt in $(0,\varepsilon)$. In our work here we shall never need to know the actual value of ε, but merely that such a value exists.

So the random variable $\Xi(dt; x,t)$, and its associated density function $\Pi(\xi \mid dt; x,t)$, depend parametrically on the three real variables dt, x and t. We have chosen to notationally separate the parameter dt from the parameters x and t because, as we shall see later, it is possible for $\Pi(\xi \mid dt; x,t)$ to be explicitly independent of x or t, but *not* of dt. In fact, it usually turns out that the *moments* of $\Xi(dt; x,t)$ have the analytical structure

$$\langle \Xi^n(dt; x,t) \rangle \equiv \int_{-\infty}^{\infty} d\xi\, \xi^n\, \Pi(\xi \mid dt; x,t) = B_n(x,t)\, dt + o(dt)$$

$$(n = 1,2,\ldots), \quad (2.4\text{-}3a)$$

where $B_1(x,t)$, $B_2(x,t)$, ... are all well-behaved functions of x and t, and $o(dt)$ denotes terms that go to zero with dt faster than dt:

$$o(dt)/dt \to 0 \quad \text{as} \quad dt \to 0.$$

Equation (2.4-3a) essentially *defines* the functions $B_n(x,t)$, as can be seen a little more clearly by dividing through by dt and then noting that $o(dt)/dt$ can be made arbitrarily small by taking dt small enough; in other words, for vanishingly small dt we have

$$B_n(x,t) = \frac{1}{dt} \langle \Xi^n(dt; x,t) \rangle = \frac{1}{dt} \int_{-\infty}^{\infty} d\xi\, \xi^n\, \Pi(\xi \mid dt; x,t)$$

$$(n = 1,2,\ldots). \quad (2.4\text{-}3b)$$

We shall call $B_n(x,t)$ the **n^{th} propagator moment function** of the Markov process $X(t)$. Much of what follows in this chapter will be devoted to deriving some fundamental equations involving these propagator moment functions. Obviously, those equations will have meaning *only* if the propagator density function $\Pi(\xi \mid dt; x,t)$ is such that the above definition of $B_n(x,t)$ truly makes sense. As we shall discover in the following chapters, the propagator moment functions $B_n(x,t)$ are nearly always well defined for Markov processes of practical interest.

Notice that the propagator $\Xi(dt; x,t)$, its density function $\Pi(\xi \mid dt; x,t)$ and its moments $\langle \Xi^n(dt; x,t) \rangle$ all need be specified *only to lowest order in* dt. This is because we can always choose the allowed range $(0,\varepsilon)$ of the variable dt to be arbitrarily small.

There is a very important connection between the propagator density function Π and the Markov state density function P. To expose that connection, we observe that the definition (2.1-7) of P implies that, for dt a positive infinitesimal,

$$\begin{aligned} &P(x+\xi, t+dt \mid x,t)\,d\xi \\ &\quad = \text{Prob}\{\, X(t+dt) \in [x+\xi,\, x+\xi+d\xi) \text{ given that } X(t) = x \,\} \\ &\quad = \text{Prob}\{\, [X(t+dt)-x] \in [\xi, \xi+d\xi) \text{ given that } X(t) = x \,\} \\ &\quad = \text{Prob}\{\, [X(t+dt)-X(t)] \in [\xi, \xi+d\xi) \text{ given that } X(t) = x \,\} \\ &\quad = \text{Prob}\{\, \Xi(dt; x,t) \in [\xi, \xi+d\xi) \,\}, \end{aligned}$$

where the last equality follows from the definition (2.4-1) of the propagator $\Xi(dt; x,t)$. Hence, by the definition (2.4-2) of the propagator density function $\Pi(\xi \mid dt; x,t)$, we conclude that [see Fig. 2-3]

$$\Pi(\xi \mid dt; x,t) = P(x+\xi, t+dt \mid x,t). \quad (2.4\text{-}4)$$

So we see that the propagator density function $\Pi(\xi \mid dt; x,t)$ is just the Markov state density function $P(x',t' \mid x,t)$ in which x' exceeds x by the amount ξ, while t' exceeds t by the *infinitesimal* amount dt. Notice that ξ, unlike dt, is *not* assumed to be an infinitesimal; ξ is an unrestricted real variable.

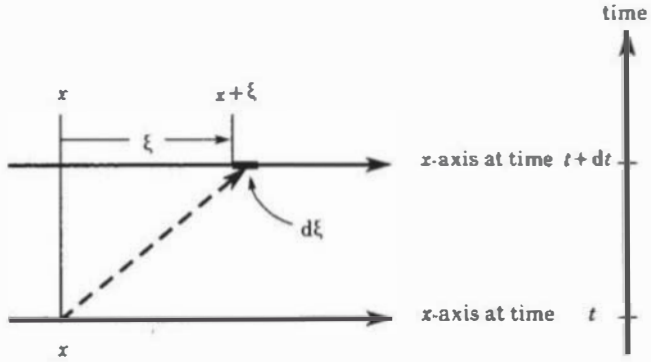

Figure 2-3. Illustrating the transition of the process from state x at time t to the interval $[x+\xi, x+\xi+d\xi)$ at time $t+dt$. The infinitesimal nature of $d\xi$ allows us to write the probability for this transition to occur as $P(x+\xi, t+dt \mid x,t)d\xi$. And because dt is likewise an infinitesimal, we can also express this transition probability as $\Pi(\xi \mid dt; x,t)d\xi$.

Because $\Pi(\xi \mid \mathrm{d}t; x,t)$ is a *density* function with respect to its argument ξ, then it must satisfy the two conditions

$$\Pi(\xi \mid \mathrm{d}t; x,t) \geq 0 \tag{2.4-5}$$

and

$$\int_{-\infty}^{\infty} \mathrm{d}\xi \; \Pi(\xi \mid \mathrm{d}t; x,t) = 1. \tag{2.4-6}$$

Furthermore, since it is clear from the definition (2.4-1) that $\Xi(\mathrm{d}t=0; x,t)$ must be the *sure* number zero, then we must have [see (1.2-7)]

$$\Pi(\xi \mid \mathrm{d}t=0; x,t) = \delta(\xi). \tag{2.4-7}$$

In addition to these three basic conditions on the propagator density function Π, there is one more important requirement that follows from the fundamental relation (2.4-4) and the Chapman-Kolmogorov equation (2.2-4). According to the latter equation, P must satisfy for any $\Delta t > 0$ and any value of α between 0 and 1,

$$\begin{aligned} &P(x+\xi, t+\Delta t \mid x,t) \\ &= \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \; P(x+\xi, t+\Delta t \mid x+\xi_1, t+\alpha\Delta t)\; P(x+\xi_1, t+\alpha\Delta t \mid x,t) \\ &= \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \; P(x+\xi_1+\xi-\xi_1,\, t+\alpha\Delta t+(1-\alpha)\Delta t \mid x+\xi_1, t+\alpha\Delta t) \\ &\qquad \times P(x+\xi_1, t+\alpha\Delta t \mid x,t). \end{aligned}$$

Then taking Δt to be the *infinitesimal* $\mathrm{d}t$, it follows on substituting from Eq. (2.4-4) that

$$\Pi(\xi \mid \mathrm{d}t; x,t) = \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \; \Pi(\xi-\xi_1 \mid (1-\alpha)\mathrm{d}t; x+\xi_1, t+\alpha\mathrm{d}t)\, \Pi(\xi_1 \mid \alpha\mathrm{d}t; x,t). \tag{2.4-8}$$

This condition, which must hold to lowest order in $\mathrm{d}t$ for all α between 0 and 1, will be called the **Chapman-Kolmogorov condition** on the propagator density function. It and conditions (2.4-5) – (2.4-7) serve to restrict the possible functional forms that can be ascribed to the propagator density function $\Pi(\xi \mid \mathrm{d}t; x,t)$.

We can gain a little more insight into the Chapman-Kolmogorov condition (2.4-8) by writing it in the equivalent form

$$\begin{aligned} \Pi(\xi \mid \mathrm{d}t; x,t) = \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \int_{-\infty}^{\infty} \mathrm{d}\xi_2 \; &\Pi(\xi_1 \mid \alpha\mathrm{d}t; x,t) \\ &\times \Pi(\xi_2 \mid (1-\alpha)\mathrm{d}t; x+\xi_1, t+\alpha\mathrm{d}t)\, \delta(\xi-\xi_1-\xi_2), \end{aligned}$$

from which Eq. (2.4-8) readily follows by integrating over ξ_2 with the help of the delta function. Viewing this last equation in light of the RVT theorem (1.6-4), we see that it implies the following functional relation among three random variables corresponding to ξ, ξ_1 and ξ_2:

$$\Xi(\mathrm{d}t; X(t),t) = \Xi(\alpha\mathrm{d}t; X(t),t) + \Xi((1-\alpha)\mathrm{d}t; X(t)+\Xi(\alpha\mathrm{d}t; X(t),t),\, t+\alpha\mathrm{d}t). \tag{2.4-9}$$

Although this relation may at first glance appear to be very complicated, it is merely the statement that {the change in the process over $(t, t+\mathrm{d}t)$} must be equal to {the change in the process over $(t, t+\alpha\mathrm{d}t)$} plus {the *subsequent* change in the process over $(t+\alpha\mathrm{d}t, t+\mathrm{d}t)$}. The *random variable* relation (2.4-9), which again must hold to lowest order in $\mathrm{d}t$ for all α between 0 and 1, is entirely equivalent to the *density function* relation (2.4-8). *Either* may be referred to as the "Chapman-Kolmogorov condition."

The importance of the propagator density function Π lies in the fact that it completely determines the Markov state density function P, from which of course everything knowable about the (Markov) process $X(t)$ can be computed. This fundamental fact is not readily apparent from Eq. (2.4-4), which shows rather that a knowledge of P implies a knowledge of Π; however, it can be proved by reasoning as follows.

In the compounded Chapman-Kolmogorov equation (2.2-6), let $t_n = t$ be any value greater that t_0, and let the points $t_1, t_2, \ldots, t_{n-1}$ divide the interval $[t_0, t]$ into n subintervals of equal length $(t-t_0)/n$. Further, change the integration variables in that equation according to

$$x_i \to \xi_i \equiv x_i - x_{i-1} \quad (i = 1, \ldots, n-1).$$

Notice that the Jacobian determinant of this transformation has zeros everywhere on one side of its main diagonal, and ones everywhere on its main diagonal, so $\mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} = \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_{n-1}$. Notice also that this transformation implies that

$$x_i = x_{i-1} + \xi_i = [x_{i-2} + \xi_{i-1}] + \xi_i = \ldots = x_0 + \xi_1 + \ldots + \xi_i.$$

Finally, relabel $x_n = x$, and *define* $\xi_n \equiv x - x_{n-1}$. With all these changes, the compounded Chapman-Kolmogorov equation (2.2-6) becomes

$$\begin{aligned} P(x,t \mid x_0,t_0) = \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \cdots \int_{-\infty}^{\infty} \mathrm{d}\xi_{n-1} \\ \times \prod_{i=1}^{n} P(x_{i-1}+\xi_i,\, t_{i-1}+(t-t_0)/n \mid x_{i-1}, t_{i-1}), \end{aligned} \tag{2.4-10}$$

wherein

$$t_i = t_{i-1} + (t-t_0)/n \qquad (i=1,\ldots,n-1), \qquad (2.4\text{-}11a)$$

$$x_i = x_0 + \xi_1 + \ldots + \xi_i \qquad (i=1,\ldots,n-1), \qquad (2.4\text{-}11b)$$

$$\xi_n \equiv x - x_0 - \xi_1 - \ldots - \xi_{n-1}. \qquad (2.4\text{-}11c)$$

We emphasize that Eq. (2.4-10) is merely the compounded Chapman-Kolmogorov equation (2.2-6) in which the intermediate times t_1, t_2, ..., t_{n-1} divide the interval $(t-t_0)$ *evenly*, and the integrations are taken with respect to the state *displacement* variables ξ_i instead of the state variables x_i. Figure 2-4 schematizes the relations among the variables in the above equations for the particular case $n=4$. Now suppose we choose *n so large* that

$$(t-t_0)/n = \mathrm{d}t, \text{ an infinitesimal.} \qquad (2.4\text{-}11d)$$

In that case, the P-factors in the integrand of Eq. (2.4-10) all become Π-factors by virtue of Eq. (2.4-4), and we obtain

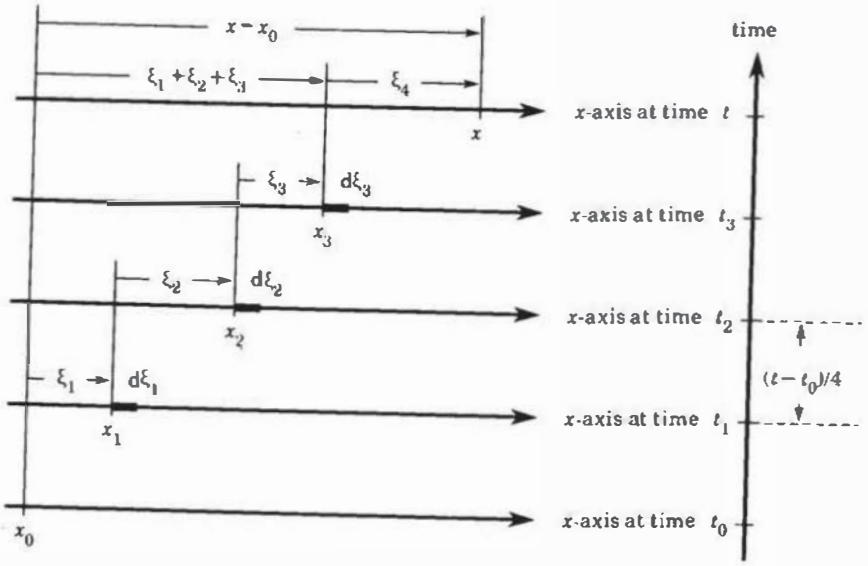

Figure 2.4. Illustrating the relations among the variables in Eq. (2.4-10) for the case $n=4$. Notice that ξ_1, ξ_2 and ξ_3 are integration variables, while ξ_4 is simply defined to be the difference between $(x-x_0)$ and $(\xi_1+\xi_2+\xi_3)$. This figure is a graphical encapsulation of relations (2.4-11a) – (2.4-11c) in which, for the sake of clarity, all ξ_i's have been taken positive.

$$P(x,t|x_0,t_0) = \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \cdots \int_{-\infty}^{\infty} \mathrm{d}\xi_{n-1} \prod_{i=1}^{n} \Pi(\xi_i | \mathrm{d}t; x_{i-1}, t_{i-1}), \qquad (2.4\text{-}12)$$

where now *all four* of Eqs. (2.4-11) apply. This result shows that if we specify $\Pi(\xi|\mathrm{d}t; x',t')$ as a function of ξ for all x', all $t' \in [t_0,t)$, and all infinitesimally small $\mathrm{d}t$, then $P(x,t|x_0,t_0)$ is uniquely determined for all x.

Although Eq. (2.4-12) gives proof to the assertion that the propagator density function Π completely determines the Markov state density function P, that formula would not seem to afford a universally simple means of actually computing P from Π. Notice in particular the potentially complicated dependence of the integrand in Eq. (2.4-12) on the integration variables through the functions x_i defined in Eq. (2.4-11b). In the next section we shall consider a different way of using the Chapman-Kolmogorov equation and the fundamental relation (2.4-4) to compute P from a knowledge of Π. But first it might be appropriate to make some general observations about the rationale for *any* method that seeks to calculate P from Π.

To say that we must specify Π in order to determine P is, because of Eq. (2.4-4), to say that we must specify $P(x',t'|x,t)$ for t' *infinitesimally* larger than t in order to calculate $P(x',t'|x,t)$ for *any* t' larger than t. This may seem like circular reasoning, but it does have an obvious precedent in mathematical physics in the elementary problem of a particle of mass m moving along the x-axis. The "state" of that simple system at time t can be represented by the ordered pair $[x_t,p_t]$, the position and momentum of the particle at time t. Now, it is well known that in order to calculate the particle's state at any time $t>t_0$, given the state at time t_0, it is necessary to specify the force function, $F(x,p;t)$. But the ultimate purpose of this force function is to specify the "state at time $t+\mathrm{d}t$ given the state at time t," namely

$$[x_{t+\mathrm{d}t}, p_{t+\mathrm{d}t} | x_t, p_t] = [x_t + (p_t/m)\mathrm{d}t,\ p_t + F(x_t,p_t;t)\mathrm{d}t]. \qquad (2.4\text{-}13)$$

Once this "conditional state" has been specified, the mathematical machinery of integral calculus may then be used to deduce an explicit formula for $[x_t,p_t|x_0,p_0]$ for *any* $t>t_0$. Of course, in the case of a Markov process, it is not generally possible to specify with certainty the value of $X(t+\mathrm{d}t)$ given the value of $X(t)$. The most we can do toward that end is to specify the probability distribution for $X(t+\mathrm{d}t)$ given the value of $X(t)$. And that is precisely what the propagator density function Π does [recall the definitions (2.4-1) and (2.4-2)].

Equation (2.4-13) also makes it clear that our freedom to specify the form of a particle's state at time $t+dt$ given its state at time t is *not* unlimited; because, even though we have considerable latitude in specifying the form of the force function $F(x,p;t)$, that function can be used in the conditional state formula (2.4-13) in only a very particular way. Similarly, for a Markov process we do not have unlimited freedom to invent propagator density functions $\Pi(\xi\,|\,d\xi;x,t)$. We must make sure that any such function satisfies conditions (2.4-5) – (2.4-8), of which the last is especially restrictive. And when those four conditions are supplemented by even rather modest additional requirements, the form of Π can become surprisingly rigid. In Chapters 3 and 4, we shall present two different sets of "reasonable additional requirements" on Π that lead to two particularly interesting and useful classes of Markov processes. More specifically, in Chapter 3 we shall derive the functional form that $\Pi(\xi\,|\,d\xi;x,t)$ must have if the random variable $\Xi(d\xi;x,t)$ is to have the character of a *well behaved infinitesimal*; this will give rise to the class of so-called *continuous* Markov processes. And then in Chapter 4 we shall derive the functional form that $\Pi(\xi\,|\,d\xi;x,t)$ must have if the random variable $\Xi(d\xi;x,t)$ is to be *usually zero but occasionally finite*; this will give rise to the class of so-called *jump* Markov processes. In the present chapter we shall develop some general results, involving either the propagator density function $\Pi(\xi\,|\,d\xi;x,t)$ or the propagator moment functions $\{B_n(x,t)\}$, that will be applicable to *both* of those classes of Markov processes.

2.5 THE KRAMERS-MOYAL EQUATIONS

Equation (2.4-12) is essentially an infinite order *integral* equation for the Markov state density function P — infinite because Eq. (2.4-11d) evidently requires n to be infinitely large. That equation is evidently completely determined if the propagator density function Π is specified. Now we shall derive two infinite order *differential* equations for P, which are likewise completely determined if Π is specified. These two infinite order differential equations are called the *Kramers-Moyal equations*, and they, like Eq. (2.4-12), are consequences of the Chapman-Kolmogorov equation; in particular, the "forward" Kramers-Moyal equation follows from Eq. (2.2-5a), and the "backward" Kramers-Moyal equation follows from Eq. (2.2-5b).

The derivation of the forward Kramers-Moyal equation starts by observing that if we define the function f by

$$f(x) \equiv P(x+\xi,t+\Delta t\,|\,x,t)\,P(x,t\,|\,x_0,t_0),$$

then the integrand of Eq. (2.2-5a) can be written $f(x-\xi)$. Assuming f to be infinitely differentiable, we have from Taylor's theorem that

$$f(x-\xi) = f(x) + \sum_{n=1}^{\infty} \frac{(-\xi)^n}{n!}\,\frac{\partial^n}{\partial x^n} f(x).$$

Substituting this for the integrand in Eq. (2.2-5a), we get

$$\begin{aligned} &P(x,t+\Delta t\,|\,x_0,t_0) \\ &\quad = \int_{-\infty}^{\infty} d\xi\; P(x+\xi,t+\Delta t\,|\,x,t)\,P(x,t\,|\,x_0,t_0) \\ &\qquad + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!}\,\frac{\partial^n}{\partial x^n}\left|\int_{-\infty}^{\infty} d\xi\,\xi^n\,P(x+\xi,t+\Delta t\,|\,x,t)\,P(x,t\,|\,x_0,t_0)\right|, \end{aligned}$$

where we have assumed that the order in which the n-summation and ξ-integration are performed is not important. The first term on the right, because of Eq. (2.2-2), integrates to $P(x,t\,|\,x_0,t_0)$. Subtracting that quantity from both sides and dividing by Δt gives

$$\begin{aligned} &[P(x,t+\Delta t\,|\,x_0,t_0) - P(x,t\,|\,x_0,t_0)]\,/\,\Delta t \\ &\quad = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!}\,\frac{\partial^n}{\partial x^n}\left|\left\{\frac{1}{\Delta t}\int_{-\infty}^{\infty} d\xi\,\xi^n\,P(x+\xi,t+\Delta t\,|\,x,t)\right\} P(x,t\,|\,x_0,t_0)\right|. \end{aligned}$$

Now taking the limit $\Delta t \to 0$, and noting that

$$\begin{aligned} \lim_{\Delta t\to 0}\ &\frac{1}{\Delta t}\int_{-\infty}^{\infty} d\xi\,\xi^n\,P(x+\xi,t+\Delta t\,|\,x,t) \\ &= \frac{1}{dt}\int_{-\infty}^{\infty} d\xi\,\xi^n\,P(x+\xi,t+dt\,|\,x,t) \\ &= \frac{1}{dt}\int_{-\infty}^{\infty} d\xi\,\xi^n\,\Pi(\xi\,|\,d\xi;x,t) \qquad \text{[by (2.4-4)]} \\ &\equiv B_n(x,t), \qquad \text{[by (2.4-3b)]} \end{aligned}$$

we obtain the equation

$$\frac{\partial}{\partial t} P(x,t\,|\,x_0,t_0) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[B_n(x,t)\, P(x,t\,|\,x_0,t_0) \right]. \qquad (2.5\text{-}1)$$

This is the **forward Kramers-Moyal equation**. If all the functions $B_n(x,t)$ are known — i.e., if all the propagator moments $\langle \Xi^n(dt; x,t) \rangle$ are known *and* have the analytical form assumed in Eqs. (2.4-3) — then the forward Kramers-Moyal equation (2.5-1) constitutes a self-contained t-evolution equation for $P(x,t\,|\,x_0,t_0)$ for fixed values of x_0 and t_0.

The derivation of the backward Kramers-Moyal equation starts by observing that if we define the function h by

$$h(x_0) \equiv P(x,t\,|\,x_0,t_0+\Delta t_0),$$

then the *first factor* in the integrand of Eq. (2.2-5b) can be written as $h(x_0+\xi)$. Assuming h to be infinitely differentiable, we have from Taylor's theorem that

$$h(x_0+\xi) = h(x_0) + \sum_{n=1}^{\infty} \frac{\xi^n}{n!} \frac{\partial^n}{\partial x_0^{\,n}} h(x_0).$$

Subs tituting this for the first factor in the integrand of Eq. (2.2-5b), we get

$$\begin{aligned} &P(x,t\,|\,x_0,t_0) \\ &\quad = \int_{-\infty}^{\infty} d\xi\; P(x,t\,|\,x_0,t_0+\Delta t_0)\, P(x_0+\xi,t_0+\Delta t_0\,|\,x_0,t_0) \\ &\quad + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{\infty} d\xi\; \xi^n \left| \frac{\partial^n}{\partial x_0^{\,n}} P(x,t\,|\,x_0,t_0+\Delta t_0) \right| P(x_0+\xi,t_0+\Delta t_0\,|\,x_0,t_0), \end{aligned}$$

where again we have assumed that the order of n-summation and ξ-integration can be changed. The first term on the right, because of Eq. (2.2-2), integrates to $P(x,t\,|\,x_0,t_0+\Delta t_0)$. Subtracting that quantity from both sides and dividing through by Δt_0 gives

$$\begin{aligned} &[P(x,t\,|\,x_0,t_0) - P(x,t\,|\,x_0,t_0+\Delta t_0)]/\Delta t_0 \\ &\quad = \sum_{n=1}^{\infty} \frac{1}{n!} \left| \frac{1}{\Delta t_0} \int_{-\infty}^{\infty} d\xi\, \xi^n\, P(x_0+\xi,t_0+\Delta t_0\,|\,x_0,t_0) \right| \frac{\partial^n}{\partial x_0^{\,n}} P(x,t\,|\,x_0,t_0+\Delta t_0). \end{aligned}$$

Now taking the limit $\Delta t_0 \to 0$, and noting that

$$\begin{aligned} \lim_{\Delta t_0 \to 0} \frac{1}{\Delta t_0} &\int_{-\infty}^{\infty} d\xi\, \xi^n\, P(x_0+\xi,t_0+\Delta t_0\,|\,x_0,t_0) \\ &= \frac{1}{dt_0} \int_{-\infty}^{\infty} d\xi\, \xi^n\, P(x_0+\xi,t_0+dt_0\,|\,x_0,t_0) \\ &= \frac{1}{dt_0} \int_{-\infty}^{\infty} d\xi\, \xi^n\, \Pi(\xi\,|\,dt_0; x_0,t_0) \qquad \text{[by (2.4-4)]} \\ &\equiv B_n(x_0,t_0), \qquad \text{[by (2.4-3b)]} \end{aligned}$$

we obtain the equation

$$-\frac{\partial}{\partial t_0} P(x,t\,|\,x_0,t_0) = \sum_{n=1}^{\infty} \frac{1}{n!} B_n(x_0,t_0) \frac{\partial^n}{\partial x_0^{\,n}} \left[P(x,t\,|\,x_0,t_0) \right]. \qquad (2.5\text{-}2)$$

This is the **backward Kramers-Moyal equation**. If all the functions $B_n(x,t)$ are known — i.e., if all the propagator moments $\langle \Xi^n(dt; x,t) \rangle$ are known *and* have the analytical form assumed in Eqs. (2.4-3) — then the backward Kramers-Moyal equation (2.5-2) constitutes a self-contained t_0-evolution equation for $P(x,t\,|\,x_0,t_0)$ for fixed values of x and t.

It follows from Eq. (2.2-3) that the *forward* Kramers-Moyal equation (2.5-1), being a t-evolution equation equation for fixed x_0 and t_0, is to be solved subject to the *initial condition*

$$P(x,t=t_0\,|\,x_0,t_0) = \delta(x-x_0), \qquad (2.5\text{-}3)$$

whereas the *backward* Kramers-Moyal equation (2.5-2), being a t_0-evolution equation for fixed x and t, is to be solved subject to the *final condition*

$$P(x,t\,|\,x_0,t_0=t) = \delta(x-x_0). \qquad (2.5\text{-}4)$$

The fact that the Kramers-Moyal equations are *infinite* order partial differential equations obviously makes their solution rather problematic. In any event, like the infinite order *integral* equation (2.4-12), the Kramers-Moyal equations show clearly that a determination of the Markov state density function $P(x,t\,|\,x_0,t_0)$ requires that the process propagator $\Xi(dt; x,t)$ be specified. In the case of the integral equation (2.4-12), $\Xi(dt; x,t)$ is specified by giving its density function $\Pi(\xi\,|\,dt; x,t)$; in the case of the Kramers-Moyal equations (2.5-1) and (2.5-2), $\Xi(dt; x,t)$ is specified by giving all of its moments $\langle \Xi^n(dt; x,t) \rangle = B_n(x,t)dt$.

It often happens that *all* these equations for the density function $P(x,t|x_0,t_0)$ of $X(t)$ are just too difficult for one to solve, and one would gladly settle for a knowledge of the behavior of just a few low order moments of $X(t)$, such as its mean, variance and covariance. In Section 2.7 we shall derive time-evolution equations for these moments in terms of the propagator moment functions $B_n(x,t)$. But before doing that, we want to explore the possibility of *differentiating* and *integrating* a Markov process $X(t)$ with respect to it argument t.

2.6 THE TIME-INTEGRAL OF A MARKOV PROCESS

In this section we shall consider whether it is possible to sensibly define the *derivative* and *integral* of a Markov process $X(t)$ with respect to t. We shall discover that the time-integral of $X(t)$ can indeed be well defined, but that the time-derivative of $X(t)$ does not exist unless $X(t)$ happens to be a completely deterministic process.

Let us consider first the matter of the *time-derivative* of $X(t)$. By the definition (2.4-1) of the process propagator, if $X(t)=x$ then in the next infinitesimal time interval dt the process will change by the amount

$$X(t+dt) - X(t) = \Xi(dt; x,t).$$

It follows that a *typical value* of this process increment in time dt will be

$$X(t+dt) - X(t) \sim \text{mean}\{\Xi(dt; x,t)\} \pm \text{sdev}\{\Xi(dt; x,t)\}. \qquad (2.6\text{-}1)$$

To estimate this typical value, we are going to *assume* that the first two propagator moment functions B_1 and B_2, as defined in Eq. (2.4-3b), both exist as well-behaved functions. We shall find in Chapters 3 and 4 that this is not a very restrictive assumption, as it is valid for nearly all Markov processes of practical interest; certainly it is a benign assumption for any Markov process for which the Kramers-Moyal equations (2.5-1) and (2.5-2) have meaning. Now, by the definition (2.4-3a), we have

$$\text{mean}\{\Xi(dt; x,t)\} = \langle \Xi(dt; x,t)\rangle = B_1(x,t)\,dt + o(dt),$$

and

$$\begin{aligned}\text{sdev}\{\Xi(dt; x,t)\} &= \{\langle \Xi^2(dt; x,t)\rangle - \langle \Xi(dt; x,t)\rangle^2\}^{1/2} \\ &= \{[B_2(x,t)\,dt + o(dt)] - [B_1(x,t)\,dt + o(dt)]^2\}^{1/2} \\ &= B_2^{1/2}(x,t)(dt)^{1/2}\{1 + o(dt)/dt\}^{1/2} + o(dt).\end{aligned}$$

Substituting these typical values into Eq. (2.6-1), we thus see that if $X(t)=x$ then we will typically have

$$\begin{aligned}&X(t+dt) - X(t) \\ &\quad \sim B_1(x,t)\,dt + o(dt) \pm \left[B_2^{1/2}(x,t)(dt)^{1/2}\left\{1 + o(dt)/dt\right\}^{1/2} + o(dt)\right].\end{aligned}$$

Dividing through by dt, we get

$$\begin{aligned}&[X(t+dt) - X(t)]/dt \\ &\quad \sim B_1(x,t) + \frac{o(dt)}{dt} \pm \frac{B_2^{1/2}(x,t)}{(dt)^{1/2}}\left\{1 + \frac{o(dt)}{dt}\right\}^{1/2} \pm \frac{o(dt)}{dt}. \qquad (2.6\text{-}2)\end{aligned}$$

Now taking the limit $dt\to 0$, keeping in mind that $o(dt)/dt\to 0$ in that limit, we get

$$\lim_{dt\to 0} \frac{X(t+dt) - X(t)}{dt} \sim \begin{cases} B_1(x,t), & \text{if } B_2(x,t)=0; \\ \pm\infty, & \text{if } B_2(x,t)\neq 0. \end{cases} \qquad (2.6\text{-}3)$$

So we see that the limit on the left, which obviously defines the derivative of the process at time t in state x, does *not* exist *except* in the special circumstance that $B_2(x,t)=0$. We shall discover in Chapters 3 and 4 [see especially Section 3.3A] that the only circumstance in which $B_2(x,t)$ vanishes identically is when the process $X(t)$ is a completely deterministic process. We thus conclude that *a genuinely stochastic Markov process does not have a derivative with respect to time.*

Now let us consider the possibility of defining a *time-integral* of the Markov process $X(t)$. We begin by noting that if the stochastic process

$$S(t) \equiv \int_{t_0}^{t} X(t')\,dt' \qquad (2.6\text{-}4)$$

exists at all, it *cannot* be a *Markov* process! This is because $S(t)$ would by definition have a derivative with respect to t, namely

$$dS(t)/dt = X(t), \qquad (2.6\text{-}5)$$

and this is something that, as we have just demonstrated, a genuinely stochastic Markov process does *not* have. Nevertheless, it may still be possible to define $S(t)$ as a viable *non-Markovian* process.

So, how might we define the process $S(t)$ which we have denoted *symbolically* by Eqs. (2.6-4) and (2.6-5)? If $X(t)$ in those representations were a *sure* function of t, then $S(t)$ would of course be that sure function for which