Stochastic Processes

ALAN J. MCKANE

Theory Group, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK

Article Outline

Glossary

Fokker-Planck equation

A partial differential equation of the second order for the time evolution of the probability density function of a stochastic process. It resembles a diffusion equation, but has an extra term which represents the deterministic aspects of the process.

Langevin equation

A stochastic differential equation of the simplest kind: linear and with an additive Gaussian white noise. Introduced by Langevin in 1908 to describe Brownian motion; many stochastic differential equations in physics go by this name.

Markov process

A stochastic process in which the current state of the system is only determined from its state in the immediate past, and not by its entire history.

Markov chain

A Markov process where both the states and the time are discrete and where the process is stationary.

Master equation

The equation describing a continuous-time Markov chain.

Stochastic process

A sequence of stochastic variables. This sequence is usually a time-sequence, but could also be spatial.

Stochastic variable

A random variable. This is a function which maps outcomes to numbers (real or integer).

I. Definition of the Subject and It's Importance

The most common type of stochastic process comprises of a set of random variables $\{x(t)\}$, where t represents the time which may be real or integer valued. Other types of stochastic process are possible, for instance when the stochastic variable depends on the spatial position $\mathbf{r}$, as well as, or instead of, t. Since in the study of complex systems we will predominantly be interested in applications relating to stochastic dynamics, we will suppose that it depends only on t. One of the earliest investigations of a stochastic process was carried out by Bachelier (1900), who used the idea of a random walk to analyse stock market fluctuations. The problem of a random walk was more generally discussed by Pearson (1905), and applied to the investigation of Brownian motion by Einstein (1905, 1906), Smoluchowski (1906) and Langevin (1908). The example of a random walk illustrates the fact that in addition to time being discrete or continuous, the stochastic variable itself can be discrete (for instance, the walker moves with fixed step size in one dimension) or continuous (for instance, the velocity of a Brownian particle). The modelling of the process may lead to an equation for the stochastic variable, such as a stochastic differential equation, or for an equation which predicts how the probability density function (pdf) for the stochastic variable changes in time. Stochastic processes are ubiquitous in the physical, biological and social sciences; they may come about through the perception of very complicated processes being essentially random (the toss of a coin, roll of a die, birth or death of individuals in populations), the inclusion of diverse and poorly characterised effects external to the system under consideration, or thermal fluctuations, among others.

II. Introduction

Deterministic dynamical processes are typically formulated as a set of rules which allow for the state of the system at time $t + 1$ (or $t + \delta t$) to be found from the state of the system at time t. By contrast, for stochastic systems, we can only specify the probability of finding the system in a given state. If this only depends on the state of the system at the previous time step, but not those before this, the stochastic process is said to be Markov. Fortunately many stochastic processes are Markovian to a very good approximation, since the theory of non-Markov processes is considerably more complicated than Markov processes and much less well developed. In this article we will deal almost exclusively with Markov processes.

The mathematical definition of a Markov process follows from the definition of the hierarchy of pdfs for a given process. This involves the joint pdfs $P(x_1, t_1; x_2, t_2; \ldots; x_n, t_n)$, which are the probability that the system is in state x_1 at time t_1, state x_2 at time t_2,..., and state x_n at time t_n, and also the conditional pdfs $P(x_1, t_1; \ldots; x_m, t_m | x_{m+1}, t_{m+1}; \ldots; x_n, t_n)$,

which are the probability that the system is in state x_1 at time t_1,...,x_m at time t_m,*given* that it was in state x_{m+1} at time t_{m+1},...,x_n at time t_n. These pdfs are all non-negative and normalisable, and relations exist between them due to symmetry and reduction (integration over some of the state variables). Nevertheless, for a general non-Markov process, a whole family of these pdfs will be required to specify the process. On the other hand, for a Markov process the history of the system, apart from the immediate past, is forgotten, and so $P(x_1, t_1; \ldots; x_m, t_m | x_{m+1}, t_{m+1}; \ldots; x_n, t_n) = P(x_1, t_1; \ldots, x_m; t_m | x_{m+1}, t_{m+1})$. A direct consequence of this is that the whole hierarchy of pdfs can be determined from only two of them: $P(x,t)$ and $P(x,t|x',t')$. The hierarchy of defining equations then collapses to only two:

$$P(x_2, t_2) = \int dx_1 P(x_2, t_2 | x_1, t_1) P(x_1, t_1) \tag{1}$$

and

$$P(x_3, t_3 | x_1, t_1) = \int dx_2 P(x_3, t_3 | x_2, t_2) P(x_2, t_2 | x_1, t_1)\,, \quad t_1 < t_2 < t_3\,. \tag{2}$$

The pdf $P(x,t|x',t')$ is referred to as the transition probability and Eq. (2) as the Chapman-Kolmogorov equation. While the pdfs for a Markov process must obey Eqs. (1) and (2), the converse also holds: any two non-negative functions $P(x,t)$ and $P(x,t|x',t')$ which satisfy Eqs. (1) and (2), uniquely define a Markov process.

We will begin our discussion in Section III with what is probably the simplest class of Markov processes: the case when both the state space and time are discrete. These are called Markov chains and were first investigated, for a finite number of states, by Markov in 1906. The extension to an infinite number of states was carried out by Kolmogorov in 1936. If time is continuous, an analogous formalism may be developed, which will be discussed in Section IV. In physics the equation describing the time evolution of the pdf in this case is called the master equation and was introduced by Pauli in 1928, in connection with the approach to equilibrium for quantum systems, and also by Nordsieck, Lamb and Uhlenbeck in 1940, in connection with fluctuations in cosmic ray physics. The term "master equation" refers to that fact that many of the quantities of interest can be derived from this equation. The connection with previous work on Markov processes was clarified by Siegert in 1949.

In many instances when the master equation cannot be solved exactly, it is useful to approximate it by a rather coarser description of the system, known as the Fokker-Planck equation. This approach will be discussed in Section V. This equation was used in its linear form by Rayleigh (1891), Einstein (1905, 1906), Smoluchowski (1906, 1916), and Fokker (1914), but it was Planck who derived the general nonlinear form from a master equation in 1917, and Kolmogorov who made the procedure rigorous in 1931. All the descriptions which we have mentioned so far have been based on the time evolution of the pdfs. An alternative specification is to give the time evolution of the stochastic variables themselves. This will necessarily involve random variables appearing in the equations describing this evolution, and they will therefore be *stochastic differential equations*. The classic example is the Langevin equation (1908) used to describe Brownian motion. This equation is linear and can therefore be solved exactly. The Langevin approach, and its relation to the Fokker-Planck equation is described in Section VI.

A good summary of the understanding of stochastic processes that had been gained by the mid-1950s is given in the book edited by Wax. This covers the basics of the subject, and what is discussed in the first six sections of this article. The article by Chandrasekhar, first published in 1943, and reprinted in Wax, gives an extensive bibliography of stochastic problems in physics before 1943. Since then the applications of the subject have grown

enormously, and the equations modelling these systems have correspondingly become more complex. We illustrate some of the procedures which have been developed to deal with these equations in the next two sections. In Section VII we discuss how the path-integral formalism may be applied to stochastic processes and in Section VIII we describe how master equations can be analysed when the size of the system is large. We end with a look forward to the future in Section IX.

III. Markov Chains

The simplest version of the Markov process is when both the states and the time are discrete, and when the stochastic process is *stationary*. When the states are discrete we will denote them by n or m, rather than x, which we reserve for continuous state variables. In this notation the two equations (1) and (2) governing Markov processes read

$$P(n_2,t_2) = \sum_{n_1} P(n_2,t_2|n_1,t_1)P(n_1,t_1) \tag{3}$$

$$P(n_3,t_3|n_1,t_1) = \sum_{n_2} P(n_3,t_3|n_2,t_2)P(n_2,t_2|n_1,t_1)\,, \quad t_1<t_2<t_3\,. \tag{4}$$

A stationary process is one in which the conditional pdf $P(n,t|n',t')$ only depends on the time difference $(t-t')$. For such processes, when time is discrete so that $t=t'+1,t'+2,\ldots$, we may write $P(n,t'+k|n',t')$ as $p^{(k)}_{n\,n'}$. The most elementary form of the Chapman-Kolmogorov equation (4) may then be expressed as

$$p^{(2)}_{n\,m} = \sum_{n'} p^{(1)}_{n\,n'} p^{(1)}_{n'\,m}\,. \tag{5}$$

This corresponds to the matrix multiplication of $p^{(1)}$ with itself, and therefore $p^{(2)}$ is simply $(p^{(1)})^2$. In the same way $p^{(k)}=(p^{(1)})^k$, and from now on we drop the superscript on $p^{(1)}$ and denote the matrix by $\mathcal{P}$. The entries of $\mathcal{P}$ are non-negative, with the sum of entries in each column being equal to unity, since

$$\sum_n p_{n\,n'} = \sum_n P(n,t+1|n',t) = 1\,. \tag{6}$$

Such matrices are called *stochastic matrices*. From Eq. (5) it is clear that $\mathcal{P}^2$ is also a stochastic matrix, and by induction it follows that $\mathcal{P}^k$ is a stochastic matrix if $\mathcal{P}$ is.

The other defining relation for a Markov process, Eq. (3), now becomes

$$P(n,t+1) = \sum_{n'} p_{n\,n'} P(n',t)\,. \tag{7}$$

This relation defines a *Markov chain*. It has two ingredients: the probability that the system is in state n at time t, $P(n,t)$ — which is usually what we are trying to determine, and the stochastic matrix with entries $p_{n\,n'}$ which gives the probabilities of transitions from the state n' to the state n. The transition probabilities are typically given; they define the model. Note that in many texts the probability of making a transition from n' to n is written as $p_{n'\,n}$, not $p_{n\,n'}$. If we write $P(n,t)$ as a vector $\mathbf{P}(t)$, then we may write Eq. (7) as $\mathbf{P}(t+1)=\mathcal{P}\mathbf{P}(t)$. Therefore,

$$\mathbf{P}(t) = \mathcal{P}\mathbf{P}(t-1) = \mathcal{P}\mathcal{P}\mathbf{P}(t-2) = \ldots = \mathcal{P}^t\mathbf{P}(0)\,, \tag{8}$$

and so if the initial state of the system $\mathbf{P}(0)$ is given, then we can find the state of the system at time t ($\mathbf{P}(t)$) by matrix multiplication by the tth power of the transition matrix.

Examples of Markov chains

1. *A one-dimensional random walk.*
   The most widely known example of a Markov chain is a random walk on the real axis, where the walker takes single steps between integers on the line. The simplest version is where the walker has to move during every time interval:

$$p_{n\,n'} = \begin{cases} p, & \text{if } n = n' + 1 \\ q, & \text{if } n = n' - 1 \\ 0, & \text{otherwise}\,, \end{cases} \tag{9}$$

   where $p + q = 1$. There are many variants. For instance, the walker could have a non-zero probability of staying put, in which case $p_{n\,n} = r$, with $p + q + r = 1$. The walk could be heterogeneous, in which case p and q (and r), could depend on n'. If there are boundaries, the boundary conditions have to be given. The most common two are *absorbing boundaries* defined by

$$\begin{aligned} p_{n+1\,n} &= p,\ p_{n-1\,n} = q,\ (n = 2, \ldots, N-1) \\ p_{11} &= 1, \quad p_{NN} = 1, \\ p_{n\,n'} &= 0, \quad \text{otherwise}\,, \end{aligned} \tag{10}$$

   and *reflecting boundaries* defined by

$$\begin{aligned} p_{n+1\,n} &= p, \quad p_{n-1\,n} = q,\ (n = 2, \ldots, N-1) \\ p_{21} &= p,\ p_{N\,N-1} = q, \\ p_{11} &= q, \quad p_{NN} = p, \\ p_{n\,n'} &= 0, \quad \text{otherwise}\,. \end{aligned} \tag{11}$$

   With absorbing boundaries (10), if we reach the state 1 or N, we can never leave it, since we stay there with probability 1. When the boundary is reflecting, we can never move beyond it; the only options are to move back towards the other boundary, or stay put. Well known examples of absorbing boundaries include the gambler's ruin problem, where a gambler bets a given amount against the house at each time step and can win with a probability p. Here the absorbing boundary is situated at $n = 0$. Eventually he will arrive at the state $n = 0$, where is has no money left, and so cannot re-enter the game. Birth/death processes will also have absorbing states at $n = 0$: if n is the number of individuals at a given time, and p is the probability of a birth and q of a death, then if there are no individuals left ($n = 0$), none can be born. This condition will be automatically applied if the transition probabilities are proportional to the number of individuals in the population.

2. *The Ehrenfest urn*
   This Markov chain was introduced by the Ehrenfests in 1907, to illustrate the approach to equilibrium in a gas. Two containers, A and B, contain molecules of the same gas, the sum of the number of molecules in A and B being fixed to be N. At time t a molecule is removed at random from the containers and put into the other. If n' is the number of molecules in container A at a certain time, then at the next time step the transition probabilities will be:

$$p_{n\,n'} = \begin{cases} \frac{n'}{N}, & \text{if } n' = n + 1 \\ (1 - \frac{n'}{N}), & \text{if } n' = n - 1 \\ 0, & \text{otherwise}\,. \end{cases} \tag{12}$$

   This is clearly a heterogeneous random walk of the type (9), and another interpretation of this model is as a random walk, but with a central force.

The most frequently asked question concerning Markov chains is: what is their eventual fate; how does the system behave at large time? Clearly if it tends towards a non-trivial stationary state, $P_{\rm st}(n)$, then from Eq. (7):

$$P_{\rm st}(n) = \sum_{n'} p_{n\,n'} P_{\rm st}(n')\,, \tag{13}$$

and so $P_{\rm st}(n)$ is a right eigenvector of $\mathcal{P}$ with unit eigenvalue. It follows from the properties of a general stochastic matrix that the eigenvalues of a stochastic matrix are such that $|\lambda| \leq 1$ (Gantmacher, 1959). Furthermore every stochastic matrix has an eigenvalue equal to 1, however it may not be simple — there may be a multiplicity of unit eigenvalues. The classification of Markov chains can be used to decide which of these possibilities is the case. For example, Markov chains may be reducible or irreducible, and states recurrent or transient. We shall not discuss this in detail; Feller (1968) gives a clear account of this classification and Cox and Miller (1965) explore the consequences for the nature of the eigenvalues. Instead we will examine a specific example, that of the Ehrenfest urn introduced above, and focus on the explicit calculation of the eigenvalues and eigenvectors in that case.

Suppose that $\Psi^{(k)}$ and $\Theta^{(k)}$ are the right- and left-eigenvectors of $\mathcal{P}$, respectively, corresponding to the eigenvalue $\lambda^{(k)}$, so that $\Theta^{(k)}.\Psi^{(\ell)} = \delta_{k\,\ell}$. Then, in general, and for the Ehrenfest urn in particular,

$$\mathcal{P}^t_{n\,n'} = \sum_{k=0}^{N} \psi^{(k)}_n \left(\lambda^{(k)}\right)^t \theta^{(k)}_{n'}\,, \tag{14}$$

where $\psi^{(k)}_n$ is the nth component of the vector $\Psi^{(k)}$ and similarly for the left-eigenvector. The eigenvalues and eigenvectors for the Ehrenfest urn can be found exactly (Kac, 1947; see also Krafft and Schaefer (1993)). The eigenvalues are $\lambda^{(k)} = 1-(2k/N)$, $k = 0, \ldots, N$. Thus in this case there is a single eigenvalue $\lambda = 1$. The corresponding right-eigenvector, which gives the stationary state, is a binomial distribution:

$$P_{\rm st}(n) = \psi^{(0)}_n = \frac{N!}{n!(N-n)!}\,\frac{1}{2^N}\,. \tag{15}$$

The left-eigenvector is $\theta^{(0)}_n = 1$ for all n. The other eigenvectors have the form $\theta^{(k)}_n = a_{kn}$ and $\psi^{(k)}_n = a_{kn} P_{\rm st}(n)$, where the a_{kn} are the Krawtchouk polynomials (Abramowitz and Stegun, 1965). Clearly, $a_{0n} = 1$, and the first non-trivial polynomial is $a_{1n} = \sqrt{N}[1 - (2n/N)]$. Therefore, for a suitable choice of initial conditions and using Eq. (8), the large t behaviour of the Ehrenfest urn can be found from

$$\mathcal{P}^t_{n\,n'} \approx P_{\rm st}(n)\left\{1 + c_{n\,n'}\left(1 - \frac{2}{N}\right)^t\right\}\,, \tag{16}$$

where $c_{n\,n'} = N[1 - (2n/N)][1 - (2n'/N)]$.

IV. The Master Equation

The master equation is a Markov chain in the limit where time is continuous. To derive it we will assume that the states are discrete (the derivation is essentially identical if they are continuous) and write down the Chapman-Kolmogorov equation (4) in the form:

$$P(n, t + \Delta t|n_0, t_0) = \sum_{n'} P(n, t + \Delta t|n', t) P(n', t|n_0, t_0)\,. \tag{17}$$

We consider only stationary processes, so that we may take $t_0 = 0$ without loss of generality and $P(n, t + \Delta t|n', t)$ is independent of t. We now assume that

$$P(n, t + \Delta t|n', t) = \begin{cases} 1 - \kappa(n)\Delta t + o(\Delta t), & \text{if } n = n' \\ \\ T(n|n')\Delta t + o(\Delta t), & \text{if } n \neq n' \,, \end{cases} \tag{18}$$

where $o(\Delta t)$ means that $o(\Delta t)/\Delta t$ tends to zero as $\Delta t \to 0$. This reasonable: after a very short times the transition probability to stay put is unity minus a term of order Δt and the transition probabilities to move to any other state is of order Δt, but this is still an additional assumption on the process. The quantity $T(n|n')$ is the transition rate, and is only defined for $n \neq n'$. Since $\sum_n P(n, t + \Delta t|n', t) = 1$ for all n', we have that

$$\kappa(n') = \sum_{n \neq n'} T(n|n') \,. \tag{19}$$

Substituting Eq. (18) into Eq. (17), and making use of Eq. (19) we find that

$$\begin{aligned} \frac{P(n, t + \Delta t|n_0, 0) - P(n, t|n_0, 0)}{\Delta t} &= \sum_{n' \neq n} \left[T(n|n') P(n', t|n_0, 0) \right] \\ &- P(n, t|n_0, 0) \sum_{n' \neq n} \left[T(n'|n) \right] + \frac{o(\Delta t)}{\Delta t} \,. \end{aligned} \tag{20}$$

Taking the limit $\Delta t \to 0$ gives *the master equation* for how the probability of finding the system in state n at time t changes with time:

$$\frac{dP(n, t)}{dt} = \sum_{n' \neq n} T(n|n') \, P(n', t) - \sum_{n' \neq n} T(n'|n) \, P(n, t) \,. \tag{21}$$

We have dropped the initial conditions, assuming that they are understood. It should be noticed that an analogous analysis starting from Eq. (3), rather than Eq. (4), may be carried out, leading to identical equations for $P(n, t|n_0, 0)$ and $P(n, t)$. If the state space is continuous the master equation reads

$$\frac{\partial P(x, t)}{\partial t} = \int dx' \, \left[T(x|x') \, P(x', t) - T(x'|x) \, P(x, t) \right] \,. \tag{22}$$

In most applications transitions only take place between states whose label differs by one. That is, $T(n|n')$ and $T(n'|n)$ are zero unless $n' = n + 1$ and $n' = n - 1$. These are called *one-step processes*. For such processes the master equation takes the simpler form

$$\begin{aligned} \frac{dP(n, t)}{dt} &= T(n|n + 1)P(n + 1, t) + T(n|n - 1)P(n - 1, t) \\ &- \left[T(n - 1|n) + T(n + 1|n) \right] \, P(n, t) \,. \end{aligned} \tag{23}$$

For simplicity let us write

$$g_n = T(n + 1|n) \quad \text{and} \quad r_n = T(n - 1|n) \,, \tag{24}$$

then the master equation may be written as

$$\frac{dP(n, t)}{dt} = r_{n+1} P(n + 1, t) + g_{n-1} P(n - 1, t) - \left[r_n + g_n \right] \, P(n, t) \,. \tag{25}$$

Examples of master equations

1. *The simple birth-death process.*

   For a population of simple organisms, for example a colony of bacteria, it might be reasonable to assume that the rate of birth of new bacteria is proportional to the number present at that time, and similarly for the rate of death. This is clearly a Markov process with $g_n = bn$ and $r_n = dn$, where b and d are rate constants. A variant is to include "immigrants" coming into the population from the outside at a constant rate c, so that $g_n = bn + c$.

   In this example g_n and r_n are linear in n. Such *linear one-step processes* can be solved by the introduction of the generating function $F(z,t) = \sum_n P(n,t)z^n$. This converts the master equation (which is a differential-difference equation) into a partial differential equation for $F(z,t)$ which can be solved if the process is linear. The simplest case of a pure death process ($b = c = 0$ in the above) can illustrate the general procedure. By rescaling the time ($t = \tau/d$), we may write the master equation in the very simple form

$$\frac{dP(n,\tau)}{d\tau} = (n+1)\,P(n+1,\tau) - n\,P(n,\tau)\,.$$

   Multiplying this equation by z^n and summing over all $n \geq 0$ gives

$$\begin{aligned}\frac{\partial}{\partial\tau}\left\{\sum_{n=0}^{\infty} z^n P(n,\tau)\right\} &= \sum_{n=0}^{\infty}(n+1)z^n P(n+1,\tau) - \sum_{n=1}^{\infty} n z^n P(n,\tau) \\ &= \sum_{m=1}^{\infty} m z^{m-1} P(m,\tau) - z\sum_{n=1}^{\infty} n z^{n-1} P(n,\tau)\,,\end{aligned}$$

   that is,

$$\frac{\partial F}{\partial\tau} = (1-z)\,\frac{\partial F}{\partial z}\,.$$

   A change of variable to $\xi = (1-z)e^{-\tau}$ and $\eta = \tau$ shows F to be a function of ξ only: $F(z,\tau) = \phi([1-z]e^{-\tau})$. The function ϕ may be determined from the initial condition. For instance, if $P(n,0) = \delta_{n\,N}$, then $F(z,0) = z^N$ and so $\phi(\xi) = (1-\xi)^N$. This gives the solution for F to be

$$F(z,\tau) = \left[\left(1 - e^{-\tau}\right) + ze^{-\tau}\right]^N\,.$$

   In this case F can easily be expanded as a power series in z, and the $P(n,\tau)$ read off, but even if this is difficult, the moments of the distribution can be readily found by differentiation with respect to z and then setting $z = 1$. It should now be clear why F is called a generating function. In the general case the partial differential equation for F may be solved by standard methods (Sneddon, 1957). The solution for a birth-death process without immigration is given in the book by Reichl (1998). The solution with immigration was first given by Kendall (1948), who also introduced the technique of the generating function as a method of solution of the master equation.

2. *The Moran model of genetic drift.*

   Stochastic processes occur extensively in population genetics. The simplest, and most widely known, is a model of genetic drift introduced by Fisher (1930) and

Wright (1931), in which a population of individuals in generation t mate randomly to produce the new generation $t+1$. We assume, for simplicity, that each individual has only one gene of a particular type, and that this may exist in one of two forms (alleles) denoted by A and B. The Wright-Fisher model is based on the sampling of the gene pool at generation t, which consists of n genes of type A and $(N-n)$ genes of type B, to produce the next generation of N genes. Although this may be formulated as a Markov chain, neither Fisher nor Wright did so; this was first carried out by Malécot in 1944. Here we will describe a variant of the model introduced by Moran (1958, 1962), which is a one-step process and can be formulated as a master equation.

The Moran model does not have non-overlapping generations, as in the Wright-Fisher model, and is more akin to a birth-death process where birth and death are coupled. At a given time, two individuals are sampled with replacement: one is designated the parent which is copied to create an offspring and the other is sacrificed to make way for the new offspring. Clearly if a B (chosen with probability $(N-n)/N$) is sacrificed and an A (chosen with probability n/N) is copied, this gives a contribution to $T(n+1|n)$. If the choice is that with A and B interchanged, this gives a contribution to $T(n-1|n)$. The transition rates for the Moran model are thus

$$T(n+1|n) = \beta\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right)\,, \quad T(n-1|n) = \beta\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)\,, \tag{26}$$

where β is a rate constant, which may be absorbed into the time t.

This may be extended in various ways. For example, mutations may be included: $A \xrightarrow{u} B$ and $B \xrightarrow{v} A$. With probability $(1-u-v)$ the offspring is taken to be a copy of the parent without mutation, as previously described. For the rest of the time (that is, with probability $u+v$), a mutation occurs. If the parent is an A, the offspring becomes a B with probability $u/(u+v)$, and if the parent is a B, the offspring becomes an A with probability $v/(u+v)$. This leads to the transition rates

$$\begin{aligned} T(n+1|n) &= (1-u-v)\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + v\left(1-\frac{n}{N}\right)\,, \\ T(n-1|n) &= (1-u-v)\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + u\left(\frac{n}{N}\right)\,. \end{aligned} \tag{27}$$

The master equation for the Moran model will be discussed again in the next section.

3. *Competition in a single species model.*
The birth-death process described in Example 1 can be generalised to more complex ecological situations. As it stands it consists of the two processes $A \xrightarrow{d} E$ and $A \xrightarrow{b} A+A$ representing death and birth respectively. Here A represents an individual and E is a null state. To model the finite resources available in a given patch, we put a limit on the number of allowed individuals: $n = 0, 1, \ldots, N$. We can also only allow a birth if enough space and/or other resources are available: $A + E \xrightarrow{b} A + A$ and include competition for these resources: $A + A \xrightarrow{c} A + E$. Since the probability of obtaining an A when sampling the patch is n/N and of obtaining an E is $(N-n)/N$, the birth term is now proportional to $n(N-n)/[N(N-1)]$ and the competition term proportional to $n(n-1)/[N(N-1)]$. This gives the transition rates to be

$$T(n+1|n) = \frac{2bn(N-n)}{N(N-1)}\,, \quad T(n-1|n) = \frac{cn(n-1)}{N(N-1)} + \frac{dn}{N}\,. \tag{28}$$

This approach can be extended to more than one species, for instance competition between and within two species (McKane and Newman, 2004) or predator-prey interactions (McKane and Newman, 2005). In these cases the state space is multi-dimensional: $\mathbf{n} = (n_1, n_2, \ldots)$. The master equation still has the form (21), but with n replaced everywhere by the vector $\mathbf{n}$.

Whether or not a stationary state of the master equation exists depends on the nature of the boundary conditions. There are many types of boundary conditions, but two are particularly important. If the boundaries are reflecting, then the probability current vanishes there. If they are absorbing, then the probability of being at the boundary is zero. In the former case probability is conserved, in the latter case it is not, and leaks out of the system.

So to find a non-zero pdf as $t \to \infty$ (a stationary distribution) we therefore assume that the system lies within two reflecting boundaries. For a one-step process, the net flow of probability from the state n to the state $n+1$, is $J(n,t) = g_n P(n,t) - r_{n+1} P(n+1,t)$, where $J(n,t)$ is the probability current. The master equation (25) may be written as

$$\frac{dP(n,t)}{dt} = J(n-1,t) - J(n,t)\,.$$

For a stationary state, the left-hand side of this equation is zero, and the currents will be time-independent. Therefore $J(n-1) = J(n)$ for all n, that is, all the currents are equal. Since the current vanishes at the boundaries, this constant must be zero. Therefore, for reflecting boundary conditions, $r_{n+1} P_{\rm st}(n+1) = g_n P_{\rm st}(n)$ for all n. If we suppose that one boundary is at $n = 0$ and the other at $n = N$, then we have that $P_{\rm st}(1) = (g_0/r_1) P_{\rm st}(0)$, $P_{\rm st}(2) = (g_1/r_2) P_{\rm st}(1), \ldots$, which implies $P_{\rm st}(2) = (g_1 g_0)/(r_2 r_1) P_{\rm st}(0), \ldots$. Iterating, the stationary state can be expressed as a simple product:

$$P_{\rm st}(n) = \frac{g_{n-1} g_{n-2} \cdots g_0}{r_n r_{n-1} \ldots r_1} P_{\rm st}(0)\,, \quad n = 1, \ldots, N\,. \tag{29}$$

The constant $P_{\rm st}(0)$ is determined by normalisation:

$$\begin{aligned} \sum_{n=0}^{N} P_{\rm st}(n) &= P_{\rm st}(0) + \sum_{n>0} P_{\rm st}(n) = 1 \\ \Rightarrow (P_{\rm st}(0))^{-1} &= 1 + \sum_{n=1}^{N} \frac{g_{n-1} g_{n-2} \cdots g_0}{r_n r_{n-1} \ldots r_1}\,. \end{aligned} \tag{30}$$

As an example, we return to the Ehrenfest urn (12), which in the language of the master equation is defined by $g_n = (N-n)/N$ and $r_n = n/N$ (any overall rate may be absorbed into the time, and this is irrelevant as far as the stationary state is concerned). Here $n = 0, 1, \ldots, N$ and the molecules never go outside this range, so the boundaries are reflecting. Applying Eqs. (29) and (30) shows that the stationary state is the binomial distribution given by Eq. (15).

V. The Fokker-Planck Equation

The Fokker-Planck equation describes stochastic processes at a more coarse grained level than those that we have discussed so far. It only involves continuous stochastic variables; these could be for instance the fraction of individuals or genes of a certain kind

in a population, whereas the master equation recognised the individuals or genes as discrete entities. To obtain the Fokker-Planck equation we first derive the Kramers-Moyal expansion (Kramers, 1940; Moyal, 1949).

We begin by defining the *jump moments* for the system:

$$M_\ell(x,t,\Delta t) = \int d\xi\ (\xi - x)^\ell\, P(\xi, t+\Delta t|x,t)\,. \tag{31}$$

We will assume that these are known, that is, they can be obtained by some other means. They will, however, only be required in the limit of small Δt.

The starting point for the derivation is the Chapman-Kolmogorov equation (4), with the choice of variables analogous to that used in Eq. (17) for the discrete case:

$$P(x,t+\Delta t) = \int dx' P(x,t+\Delta t|x',t)\, P(x',t)\,, \tag{32}$$

again dropping the dependence on the initial conditions. The integrand may be written as

$$\begin{aligned} P(x,t+\Delta t|x',t)\, P(x',t) &= P([x-\Delta x]+\Delta x, t+\Delta t|[x-\Delta x],t)P([x-\Delta x],t) \\ &= \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!}\,(\Delta x)^\ell\, \frac{\partial^\ell}{\partial x^\ell}\left\{P(x+\Delta x, t+\Delta t|x,t)P(x,t)\right\}\,, \end{aligned} \tag{33}$$

where $\Delta x = x - x'$. Integrating over x' gives for Eq. (32):

$$P(x,t+\Delta t) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \frac{\partial^\ell}{\partial x^\ell}\left\{M_\ell(x,t,\Delta t)P(x,t)\right\}\,. \tag{34}$$

Since $P(\xi,t|x,t) = \delta(\xi - x)$, it follows from Eq. (31), that $\lim_{\Delta t\to 0} M_\ell(x,t,\Delta t) = 0$ for $\ell \geq 1$. Also $M_0(x,t,\Delta t) = 1$. Bearing these results in mind, we will now assume that the jump moments for $\ell \geq 1$ take the form

$$M_\ell(x,t,\Delta t) = D^{(\ell)}(x,t)\Delta t + o\left(\Delta t\right)\,. \tag{35}$$

Substituting this into Eq. (34), dividing by Δt and taking the limit $\Delta t \to 0$ gives

$$\frac{\partial P}{\partial t} = \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!} \frac{\partial^\ell}{\partial x^\ell}\left\{D^{(\ell)}(x,t)P(x,t)\right\}\,. \tag{36}$$

Equation (36) is the Kramers-Moyal expansion. So far nothing has been assumed other than the Markov property and the existence of Taylor series expansions. However, in many situations, examination of the jump moments reveal that in a suitable approximation they may be neglected for $\ell > 2$. In this case, we may truncate the Kramers-Moyal expansion (36) at second order and obtain the *Fokker-Planck equation*:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left[A(x,t)P(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[B(x,t)P(x,t)\right]\,, \tag{37}$$

where $A = D^{(1)}$ and $B = D^{(2)}$ are independent of t if the process is stationary.

To calculate the jump moments (31), it is convenient to write them in terms of the underlying stochastic process $x(t)$. We use the notation

$$\langle x(t)\rangle_{x(t_0)=x_0} = \int dx\, xP(x,t|x_0,t_0)\,, \tag{38}$$

for the mean of the stochastic variable at time t, conditional on the value of $x(t)$ being given to be x_0 at time t_0. With this notation $x(t)$ denotes the process and the angle brackets are averages over realisations of this process. More generally, we may define $\langle f(x(t))\rangle$ in a similar way, and in particular the jump moments are given by

$$M_\ell(x,t,\Delta t) = \langle (x(t+\Delta t)-x)^\ell\rangle_{x(t)=x}\,. \tag{39}$$

Examples of Fokker-Planck equations

1. *Simple diffusion.*
   For the simple symmetric random walk, $g_n = 1$ and $r_n = 1$ when expressed in the language of the master equation (after a rescaling of the time so that the rates may taken to be equal to unity). From Eqs. (18) and (19) we find that for a one-step stationary process,

$$\langle (n(t+\Delta t)-n)^\ell\rangle_{n(t)=n} = \begin{cases} (g_n - r_n)\,\Delta t + o\,(\Delta t)\,, & \text{if } \ell \text{ is odd} \\ (g_n + r_n)\,\Delta t + o\,(\Delta t)\,, & \text{if } \ell \text{ is even}\,, \end{cases} \tag{40}$$

   and so for the symmetric random walk the odd moments all vanish, and the even moments are equal to $2\Delta t + o(\Delta t)$. We now make the approximation which will yield the Fokker-Planck equation: we let $x = nL$, where L is the step size, and let $L \to 0$. Since, for ℓ even, $\langle (x(t+\Delta t)-x)^\ell\rangle = (2L^\ell)\Delta t$, if we rescale the time by introducing $\tau = L^2 t$, then all jump moments higher than the second disappear in the limit $L \to 0$, and Eq. (36) becomes

$$\frac{\partial P}{\partial t} = \frac{\partial^2 P}{\partial x^2}\,. \tag{41}$$

   This is the familiar diffusion equation obtained from a continuum approximation to the discrete random walk.

2. *The diffusion limit of the Moran model.*
   For the Moran model with no mutation, we have from Eqs. (26) and (40) that the odd moments again vanish. If we describe the process by $x(t) = n(t)/N$, the fraction of the genes that are of type A at time t, then the even jump moments are given by

$$\langle (x(t+\Delta t)-x)^\ell\rangle_{x(t)=x} = \frac{1}{N^\ell}\,2x\,(1-x)\,\Delta t + o\,(\Delta t)\,,$$

   and so introducing a rescaled time $\tau = 2t/N^2$, and letting $N \to \infty$ we obtain the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P\right]\,. \tag{42}$$

   The factor of 2 in the rescaling of the time is included so that the diffusive form of the Moran model agrees with that found in the Wright-Fisher model (Crow and Kimura, 1970).

Now suppose mutations are included. The transition rates are given by Eq. (27) and lead to jump moments $\langle (x(t+\Delta t)-x)^\ell \rangle \sim N^{-\ell}$. So the first and second jump moments are not of the same order, and the introduction of a rescaled time $\tau = t/N$, and subsequently letting $N \to \infty$, gives a Fokker-Planck equation of the form (37), but with $B = 0$ and $A(x) = v - (u+v)x$. This corresponds to a *deterministic* process with $dx/dt = v - (u+v)x$ (Haken, 1983). The fact that the system tends to a macroscopic equation when $N \to \infty$ has to be taken into account when determining the nature of the fluctuations for large N. We will discuss this further in Section VIII.

On the other hand, suppose that the mutation rates scale with N according to $u = 2\tilde{u}/N$ and $v = 2\tilde{v}/N$, where $\tilde{u}$ and $\tilde{v}$ have a finite limit as $N \to \infty$, and where the 2 has again been chosen to agree with the Wright-Fisher model. Now both the first and second jump moments are of order N^{-2}, with the higher moments falling off faster with N. Therefore once again introducing the rescaled time $\tau = 2t/N^2$ and letting $N \to \infty$, we obtain the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left[\{\tilde{v} - (\tilde{u}+\tilde{v})\,x\}\,P\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P\right]\,. \tag{43}$$

So depending on the precise scaling with N, the Moran model gives different limits as $N \to \infty$ (Karlin and McGregor, 1964). In the first case, the mutational effects are strong enough that a macroscopic description exists, with fluctuations about the macroscopic state, as discussed in Section VIII. In the second case, the mutational effects are weaker, and there is no macroscopic equation describing the system; the large N limit is a non-linear Fokker-Planck equation of the diffusive type.

The Fokker-Planck equation (37) for stationary processes, where A and B are functions of x only, can be solved by separation of variables, with solutions of the form $p(x)e^{-\lambda t}$, where λ is a constant. The equation for $p(x)$ is then a second order differential equation which, when boundary conditions are given, is of the Sturm-Liouville type. To specify the boundary conditions we once again introduce the probability current, this time through the continuity equation,

$$\frac{\partial P(x,t)}{\partial t} + \frac{\partial J(x,t)}{\partial x} = 0\,, \tag{44}$$

where the probability current $J(x,t)$ is given by

$$J(x,t) = A(x,t)P(x,t) - \frac{1}{2}\frac{\partial}{\partial x}\left[B(x,t)P(x,t)\right]\,. \tag{45}$$

Let us suppose that the system is defined on the interval $[a,b]$. Then if the boundaries are reflecting, there is no net flow of probability across $x = a$ and $x = b$. This implies that $J(a,t) = 0$ and $J(b,t) = 0$. If we integrate the equation of continuity (44) from $x = a$ to $x = b$, and apply these boundary conditions, we see that $\int_a^b P(x,t)dx$ is independent of time. Therefore, if the pdf is initially normalised, it remains normalised. This is in contrast with the case of absorbing boundary conditions, defined by $P(a,t) = 0$ and $P(b,t) = 0$. If the boundary conditions are at infinity we require that $\lim_{x\to\pm\infty} P(x,t) = 0$, so that if P is well-behaved it is normalisable, and also that $\partial P/\partial x$ is well-behaved in this limit: $\lim_{x\to\pm\infty} \partial P/\partial x = 0$. If A or B do not diverge as $x \to \pm\infty$ this implies that $\lim_{x\to\pm\infty} J(x,t) = 0$. Other types of boundary conditions are possible, and we do not attempt a complete classification here (Risken, 1989; Gardiner, 2004).

If A and B are independent of time, then from Eq. (44), the stationary state of the system must be given by $dJ(x)/dx = 0$, that is, J is a constant. For reflecting boundary conditions this constant is zero, and so from Eq. (45) the stationary pdf, $P_{\rm st}(x)$ must satisfy

$$0 = A(x)P_{\rm st}(x) - \frac{1}{2}\frac{\partial}{\partial x}\left[B(x)P_{\rm st}(x)\right] . \tag{46}$$

This may be integrated to give

$$P_{\rm st}(x) = \frac{C}{B(x)} \exp\left\{2\int^x dx' \, \frac{A(x')}{B(x')}\right\} , \tag{47}$$

where C is a constant which has to be chosen so that $P_{\rm st}(x)$ is normalised.

The Fokker-Planck equation, with A independent of t and B constant, can be transformed into the Schrödinger-like problem

$$-B\frac{\partial \psi}{\partial t} = -\frac{B^2}{2}\frac{\partial^2 \psi}{\partial x^2} + U(x)\psi \, , \tag{48}$$

by the transformation

$$P(x,t) = \left[P_{\rm st}(x)\right]^{1/2} \psi(x,t) \, , \tag{49}$$

where

$$U(x) = \frac{1}{2}\left[A(x)\right]^2 + \frac{B}{2}\frac{dA}{dx} \, . \tag{50}$$

So a one-dimensional stationary stochastic process, under certain conditions (such as constant second jump moment) is equivalent to quantum mechanics in imaginary time, with B taking over the role of Planck's constant.

As an example we consider the *Ornstein-Uhlenbeck process* defined by

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left[axP\right] + D\frac{\partial^2 P}{\partial x^2} \, , \quad x \in (-\infty, \infty) \, , \; a > 0 \, . \tag{51}$$

In this case, the potential (50) is given by $U(x) = [a^2x^2 - 2aD]/2$, and so the problem is equivalent to the one-dimensional simple harmonic oscillator in quantum mechanics, but with an energy shift. As in that problem (Schiff, 1968), the eigenfunctions are Hermite polynomials. Specifically, the right-eigenfunctions are

$$p_m(x) = P_{\rm st}(x)\frac{1}{\left[2^m m!\right]^{1/2}} H_m(\alpha x) \, , \tag{52}$$

where H_m are the Hermite polynomials (Abramowitz and Stegun, 1965) and $\alpha = (a/2D)^{1/2}$. The eigenvalue corresponding to the eigenfunction (52) is $\lambda_m = am$, $m = 0, 1, \ldots$ and the left-eigenfunctions are $q_m(x) = \left[P_{\rm st}(x)\right]^{-1} p_m(x)$. From these explicit solutions we may calculate other quantities of interest, such as correlation functions. We note that the eigenvalues are all non-negative and that the stationary state corresponds to $\lambda = 0$. The left-eigenfunction for the stationary state is equal to 1. These latter results hold true for a wide-class of such problems.

VI. Stochastic Differential Equations

So far we have described stochastic processes in terms of equations which give the time evolution of pdfs. In this section, we will describe equations for the stochastic variables themselves. The most well known instance of such an equation is the Langevin equation for the velocity of a Brownian particle, and so we begin with this particular example.

Suppose that a small macroscopic particle of mass m (such as a pollen grain) is immersed in a liquid at a temperature T. In addition to any macroscopic motion that the particle may have, its velocity fluctuates due to the random collisions of the particle with the molecules of the liquid. For simplicity, we confine ourselves to one-dimensional motion — along the $x-$axis. Then the equation of motion of the particle may be written in the form

$$m\frac{d^2x}{dt^2} = -\alpha\frac{dx}{dt} - \frac{dV}{dx} + \mathcal{F}(t)\,. \tag{53}$$

The first term on the right-hand side is due to the viscosity of the fluid and α is the friction constant. The second term, where $V(x)$ is a potential, represents the interaction of the particle with any external forces, such as gravity. The final term is the random force due to collisions with the molecules of the liquid. Clearly to complete the specification of the dynamics of the particle we need to give (i) the initial position and velocity of the particle, and (ii) the statistics of the random force $\mathcal{F}(t)$.

To make progress with these points, we imagine a large number of realisations of the dynamics, in which the particle starts with the same initial position, x_0, and velocity, v_0, but where the initial positions and velocities of the molecules in the liquid will be different. Taking the average over a large number of such realisations will give the average position $\langle x(t)\rangle$ and velocity $\langle v(t)\rangle$ at time t, conditional on $x(0) = x_0$ and $v(0) = v_0$. The statistics of the fluctuating force $\mathcal{F}(t)$ are assumed to be such that

(a) $\langle\mathcal{F}(t)\rangle = 0$, since we do not expect one direction to be favoured over the other.

(b) $\langle\mathcal{F}(t)\mathcal{F}(t')\rangle = 2D\delta(t-t')$, since we expect that after a few molecular collisions the value that $\mathcal{F}$ takes on will be independent of its former value. That is, the force $\mathcal{F}$ becomes uncorrelated over times of the order of a few collision times between molecules. This is tiny on observational time scales, and so taking the correlation function to be a delta-function is an excellent approximation. The weight of the delta-function is denoted by $2D$, where at this stage D is undetermined.

(c) $\mathcal{F}(t)$ is taken to be Gaussianly distributed on grounds of simplicity, but also because by the central limit theorem it is assumed that the net effect of the large number of molecules which collide with the pollen grain will lead to a distribution which is Gaussian.

Since a Gaussian distribution is specified by its first two moments, conditions (a), (b) and (c) completely define the statistics of $\mathcal{F}(t)$.

Finally, Eq. (53) as it stands does not define a Markov process. This is most easily seen if we write down a discrete time version of the equation. The second derivative means that $x(t+\delta t)$ not only depends on $x(t)$, but also on $x(t-\delta t)$. Therefore only first order derivatives should be included in such equations if the process is to be Markov. This is easily achieved by promoting $v(t)$ to be a second stochastic variable in addition to $x(t)$.

Then Eq. (53) may be equivalent written as

$$\begin{aligned} \frac{dx}{dt} &= v\,, \\ m\frac{dv}{dt} &= -\alpha v - \frac{dV}{dx} + \mathcal{F}(t)\,, \end{aligned} \tag{54}$$

which does define a Markov process. Although, as we remarked in the Introduction, we deal almost exclusively with Markov processes in this article, the situation we have just discussed is a good illustration of one way of dealing with processes which are presented as being non-Markovian. The method simply consists of adding a sufficient number of supplementary variables to the definition of the state variables of the process until it becomes Markovian. There is no guarantee that this will be possible or require only a small number of additional variables to be promoted in this way, but it is the most straightforward and direct way of rendering non-Markovian processes tractable.

To begin the analysis of Eq. (54) we assume that there are no external forces and so the term dV/dx is equal to zero. We may then write Eq. (54) as the Langevin equation

$$\frac{dv}{dt} = -\gamma v + F(t)\,; \quad v(0) = v_0\,, \tag{55}$$

where $\gamma = \alpha/m$ and $F(t) = \mathcal{F}(t)/m$. This implies that

$$\langle F(t)\rangle = 0 \quad \text{and} \quad \langle F(t)F(t')\rangle = \frac{2D}{m^2}\delta(t-t')\,. \tag{56}$$

Note that since $F(t)$ is a random variable, solving the Langevin equation will give $v(t)$ as a random variable (having a known distribution). It is therefore a *stochastic differential equation*. The function F is frequently called "the noise term" or simply "the noise". It is *white noise* since the Fourier transform of a delta-function is a constant — all frequencies are present in equal amounts.

Multiplying the Langevin equation (55) by the integrating factor $e^{\gamma t}$ gives

$$\frac{d}{dt}\left[v(t)e^{\gamma t}\right] = F(t)e^{\gamma t} \quad \Rightarrow \quad v(t) = v_0 e^{-\gamma t} + e^{-\gamma t}\int_0^t dt'\, F(t')e^{\gamma t'}\,. \tag{57}$$

By taking the average of the expression for $v(t)$ we find $\langle v(t)\rangle = v_0 e^{-\gamma t}$. More interestingly, if we square the expression for $v(t)$ and take the average, then we find

$$\langle v^2(t)\rangle = v_0^2 e^{-2\gamma t} + \frac{D}{\alpha m}\left[1 - e^{-2\gamma t}\right]\,, \tag{58}$$

which implies that

$$\lim_{t\to\infty}\langle v^2(t)\rangle = \frac{D}{\alpha m}\,. \tag{59}$$

On the other hand, as $t \to \infty$, the Brownian particle will be in thermal equilibrium:

$$\lim_{t\to\infty}\langle v^2(t)\rangle = v_{\text{eq}}^2 \quad \text{and} \quad \frac{1}{2}mv_{\text{eq}}^2 = \frac{1}{2}kT\,,$$

where T is the temperature of the liquid and k is Boltzmann's constant. This implies that

$$\frac{1}{2}m\left(\frac{D}{\alpha m}\right) = \frac{1}{2}kT \quad \Rightarrow \quad D = \alpha kT\,. \tag{60}$$

The molecules of the liquid are acting as a heat bath for the system — which in this case is a single Brownian particle. The equation $D = \alpha kT$ is a simple example of a fluctuation-dissipation theorem, and determines D in terms of the friction constant, α, and of the temperature of the liquid, T.

Although we have presented a somewhat heuristic rationale for Eq. (54), it may be derived in a more controlled way. A particularly clear derivation has been given by Zwanzig (1973), where the starting point is a Hamiltonian which contains three terms: for the system, the heat bath and the interaction between the system and the heat bath. Taking the heat bath to be made up of coupled harmonic oscillators and the interaction term between the system and heat bath to be linear, it is possible to integrate out the bath degrees of freedom exactly, and be left only with the equations of motion of the system degrees of freedom plus the initial conditions of the bath degrees of freedom. Assuming that the bath is initially in thermal equilibrium, so that these initial values are distributed according to a Boltzmann distribution, adds extra "noise" terms to the equations of motion which, with a few more plausible assumptions, make them of the Langevin type.

Examples of Langevin-like equations

1. *Overdamped Brownian motion.*

   Frequently the viscous damping force $-\alpha v$ is much larger than the inertial term md^2x/dt^2 in Eq. (53), and so to a good approximation the left-hand side of Eq. (53) can be neglected. Scaling time by α, we arrive at the Langevin equation for the motion of an overdamped Brownian particle:

$$\frac{dx}{dt} = -V'(x) + \mathcal{F}(t)\,; \quad x(0) = x_0\,, \tag{61}$$

   where $'$ denotes differentiation with respect to x and where, due to the rescaling of time by α,

$$\langle \mathcal{F}(t) \rangle = 0\,; \quad \langle \mathcal{F}(t)\mathcal{F}(t') \rangle = 2\tilde{D}\delta(t-t')\,; \quad \tilde{D} = \frac{D}{\alpha}\,. \tag{62}$$

   A particularly well-known case is when the Brownian particle is moving in the harmonic potential $V(x) = ax^2/2$. Then

$$\frac{dx}{dt} = -ax + \mathcal{F}(t)\,; \quad x(0) = x_0\,. \tag{63}$$

   Since Eq. (63) relating $x(t)$ to $\mathcal{F}(t)$ is linear, and since the distribution of $\mathcal{F}(t)$ is Gaussian, then $x(t)$ is also distributed according to a Gaussian distribution. Comparing with Eqs. (55) and (56), which also define a linear system, we find that $\langle x(t) \rangle = x_0 e^{-at}$ and, from Eq. (58), $\langle x^2(t) \rangle = x_0^2 e^{-2at} + (\tilde{D}/a)\,[1 - e^{-2at}]$. This gives

$$P(x,t|x_0,0) = \sqrt{\frac{a}{2\pi\tilde{D}\,[1-e^{-2at}]}}\,\exp\left\{-\frac{a\,(x-x_0e^{-at})^2}{2\tilde{D}\,[1-e^{-2at}]}\right\}\,. \tag{64}$$

   It is straightforward to check that this conditional pdf satisfies the Fokker-Planck equation (51) for the Ornstein-Uhlenbeck process. Below we will show this more directly, by starting from the Langevin equation (63) and deriving Eq. (51).

   Another case of interest is when $V(x)$ is a double-well potential, $V(x) = -ax^2/2 + bx^4/4$. If the particle is initially located near the bottom of one of the potential

wells, it will take on average a time of the order of $e^{\Delta V/D}$ to hop over the barrier and into the well on the other side. Here ΔV is the height of the barrier that it has to hop over (Kramers, 1940).

2. *Environmental noise in population biology.*
One of the simplest models of two species with population sizes N_1 and N_2 which are competing for a common resource, is the two coupled deterministic ordinary differential equations $\dot{N}_i = r_i N_i$, $i = 1, 2$. The growth rates, r_i, depend on the population sizes in such a way that as the population sizes increase, the r_i decrease to reflect the increased competition for resources. This could be modelled, for instance, by taking $r_i = a_i - b_{ii}N_i - b_{ij}N_j$ with $i, j = 1, 2$ and $i \neq j$. In reality, external factors such as climate, terrain, the presence of other species, and indeed any factor which has an uncertain influence on these two species, will also affect the growth rate. This can be modelled by adding an external random term to the r_i which represents this *environmental stochasticity* (May, 1973). Then the equations become

$$\begin{aligned} \frac{dN_1}{dt} &= a_1 N_1 - b_{11}N_1^2 - b_{12}N_1 N_2 + N_1\zeta_1(t) \\ \frac{dN_2}{dt} &= a_2 N_2 - b_{22}N_2^2 - b_{21}N_2 N_1 + N_2\zeta_2(t)\,. \end{aligned} \tag{65}$$

Since the noise terms, $\zeta_i(t)$ are designed to reflect the large number of coupled variables omitted from the description of the model, it is natural, by virtue of the central limit theorem, to assume that they are Gaussianly distributed. It also seems reasonable to assume that any temporal correlation between these external influences is on scales very much shorter than those of interest to us here, and that the noises have zero mean. We therefore assume that

$$\langle \zeta_i(t) \rangle = 0\,; \quad \langle \zeta_i(t)\zeta_j(t') \rangle = 2D_i\delta_{ij}\delta(t - t')\,, \tag{66}$$

where the D_i describe the strength of the stochastic effects. The deterministic equations (that is, Eq. (65) without the noise terms) have a fixed point at the origin, one on each of the N_1 and N_2 axes, and may have another at non-zero N_1 and N_2. For some values of the parameters this latter fixed point may be a saddle, with those on the axes being stable and the origin unstable. In this situation the eventual fate of the species depends significantly on the noise: if the combination of the nonlinear dynamics and the noise drives the system to the vicinity of the fixed point on the N_1 axis, then species 2 will become extinct, and vice-versa.

Langevin equations with Gaussian white noise are equivalent to Fokker-Planck equations. This can be most easily seen by calculating the jump moments (39) from the Langevin equation. For instance, if we begin from the Langevin equation for an overdamped Brownian particle (61),

$$\begin{aligned} \Delta x(t) \equiv x(t + \Delta t) - x(t) &= \int_t^{t+\Delta t} dt'\, \dot{x}(t') \\ &= -\int_t^{t+\Delta t} dt'\, V'(x(t')) + \eta(t)\,, \end{aligned} \tag{67}$$

where $\eta(t) = \int_t^{t+\Delta t} dt'\, \mathcal{F}(t')$. From Eq. (62) it is straightforward to calculate the moments of $\eta(t)$: $\langle \eta(t) \rangle = 0$,

$$\langle \eta^2(t) \rangle = \int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt'' \langle \mathcal{F}(t')\mathcal{F}(t'') \rangle = 2\tilde{D}\Delta t \tag{68}$$

and, since $\eta(t)$ is Gaussian, $\langle \eta^n(t) \rangle$ is zero if n is odd, and at least of order $(\Delta t)^2$ for $n \geq 4$. This implies that

$$M_1(x, \Delta t) = -V'(x)\Delta t + \mathcal{O}\left(\Delta t\right)^2 \,, \quad M_2(x, \Delta t) = 2\tilde{D}\Delta t + \mathcal{O}\left(\Delta t\right)^2 \,, \tag{69}$$

with all moments of order $(\Delta t)^2$ or higher for $\ell > 2$. The notation $\mathcal{O}\left(\Delta t\right)^2$ means that the magnitude of this quantity is less than a constant times $(\Delta t)^2$, for sufficiently small nonzero $(\Delta t)^2$. This is a weaker, but more specific, statement than saying it is $o\left(\Delta t\right)$. Using Eqs. (35) and (36), the Fokker-Planck equation which is equivalent to the Langevin equation (61) is found to be

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left[V'(x)P\right] + \tilde{D}\frac{\partial^2 P}{\partial x^2} \,. \tag{70}$$

From Eq. (47), the stationary pdf is $P_{\rm st}(x) = C\exp\{-V(x)/\tilde{D}\} = C\exp\{-V(x)/kT\}$, as expected.

Although in this article we have largely restricted our attention to stochastic processes involving one variable, the construction of a Fokker-Planck equation from the Langevin equation goes through in a similar way for an $n-$dimensional process $\mathbf{x} = (x_1, \ldots, x_n)$. In this case the jump moments are

$$\langle \Delta x_{i_1}(t)\Delta x_{i_2}(t)\ldots\Delta x_{i_\ell}(t)\rangle_{\mathbf{x}(t)=\mathbf{x}} = D_{i_1\ldots i_\ell}(\mathbf{x}, t)\Delta t + o(\Delta t) \,, \tag{71}$$

where $\Delta x_{i_\alpha} = x_{i_\alpha}(t + \Delta t) - x_{i_\alpha}$. The Fokker-Planck equation is then

$$\frac{\partial P}{\partial t} = \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!} \frac{\partial^\ell}{\partial x_{i_1}\ldots\partial x_{i_\ell}} \left\{D_{i_1\ldots i_\ell}(\mathbf{x}, t)P\right\} \,. \tag{72}$$

The Langevin equation for Brownian motion (54), without going to the overdamped limit, serves as a simple illustration of this generalisation. Here $\Delta x(t) = v\Delta t$ and $\Delta v(t) = -\gamma v\,\Delta t - m^{-1}V'(x)\,\Delta t + m^{-1}\eta(t)$. This results in the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left[vP\right] + \frac{\partial}{\partial v}\left[\left\{\gamma v + m^{-1}V'(x)\right\}P\right] + \frac{\gamma kT}{m}\frac{\partial^2 P}{\partial v^2} \,. \tag{73}$$

This is Kramer's equation. It has a stationary pdf $P_{\rm st}(x, v) = C\exp\{-E/kT\}$, where $E = mv^2/2 + V(x)$.

We end this section by finding the Fokker-Planck equation which is equivalent to the general set of Langevin equations of the form

$$\dot{x}_i = A_i(\mathbf{x}, t) + \sum_{\alpha=1}^{m} g_{i\alpha}(\mathbf{x}, t)\,\zeta_\alpha(t)\,; \quad i = 1, \ldots, n \,, \tag{74}$$

where $\zeta_\alpha(t)$, $\alpha = 1, \ldots, m$, is a Gaussian white noise with zero mean and with

$$\langle \zeta_\alpha(t)\zeta_\beta(t')\rangle = \delta_{\alpha\beta}\delta(t - t') \,. \tag{75}$$

Proceeding as in Eq. (67), but noting the dependence of the function $g_{i\alpha}$ on the stochastic variable, yields

$$\begin{aligned} M_i(\mathbf{x}, t, \Delta t) &= \left[A_i(\mathbf{x}, t) + \theta(0)\sum_{j=1}^{n}\sum_{\alpha=1}^{m} g_{j\alpha}(\mathbf{x}, t)\frac{\partial}{\partial x_j}g_{i\alpha}(\mathbf{x}, t)\right]\Delta t + \mathcal{O}\left(\Delta t\right)^2 \,, \\ M_{ij}(\mathbf{x}, t, \Delta t) &= \sum_{\alpha=1}^{m}\left[g_{i\alpha}(\mathbf{x}, t)g_{j\alpha}(\mathbf{x}, t)\right]\Delta t + \mathcal{O}\left(\Delta t\right)^2 \,, \end{aligned} \tag{76}$$

with all jump moments higher than the second being of order $(\Delta t)^2$ or higher. The quantity $\theta(0)$ is the value of the Heaviside theta function, $\theta(x)$, at $x = 0$ and is indeterminate. This indicates that the Langevin description does not correspond to a unique Fokker-Planck equation. This situation occurs whenever the white noise in a Langevin equation is multiplied by a function which depends on the state variable, as in Eq. (74). For systems such as this acted upon by *multiplicative noise* the Langevin description has to be supplemented by a rule which says whether the state variable in the multiplying function ($g_{i\alpha}$ in Eq. (74)) is that before or after the noise pulse acts (van Kampen, 1981). If it is taken to be the value immediately after the noise pulse acts then $\theta(0) = 0$ (Itô rule), whereas if it is taken to be the average of the values before and after, then $\theta(0) = 1/2$ (Stratonovich rule). The Fokker-Planck equation is now found from Eq. (72) to be

$$\frac{\partial P}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left[A_i(\mathbf{x}, t) P(\mathbf{x}, t) \right] + \frac{1}{2} \sum_{i,j=1}^{n} \sum_{\alpha=1}^{m} \frac{\partial^2}{\partial x_i \partial x_j} \left[g_{i\alpha}(\mathbf{x}, t) g_{j\alpha}(\mathbf{x}, t) P(\mathbf{x}, t) \right] , \qquad (77)$$

in the Itô case and

$$\frac{\partial P}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left[A_i(\mathbf{x}, t) P(\mathbf{x}, t) \right] + \frac{1}{2} \sum_{i,j=1}^{n} \sum_{\alpha=1}^{m} \frac{\partial}{\partial x_i} \left[g_{i\alpha}(\mathbf{x}, t) \frac{\partial}{\partial x_j} \left\{ g_{j\alpha}(\mathbf{x}, t) P(\mathbf{x}, t) \right\} \right] , \qquad (78)$$

in the Stratonovich case.

VII. Path Integrals

While most early work on stochastic processes was concerned with linear systems, naturally attention soon moved on to the many interesting systems which could be modelled as nonlinear stochastic processes. These systems are much more difficult to analyse. For example, a nonlinear Langevin equation cannot be solved directly, and so the averaging procedure cannot be carried out in the same explicit way as described in Section VI. There is however one method which is applicable to many nonlinear stochastic differential equations of interest: the solution of these equations can be formally written down as a path-integral, and from this correlation functions and other quantities of physical interest can be obtained. This also has the advantage that all the formalism and approximation schemes developed to study functional integrals over the years can be called into play.

Path-integrals are intimately related to Brownian motion and the earliest work on the subject by Wiener (1921a, 1921b), emphasised this. If the problem of interest is formulated as a set of Langevin equations, the derivation of the path-integral representation is particularly straightforward, if rather heuristic. For clarity we begin with the simplest case: an overdamped system with a single degree of freedom, x, acted upon by white noise. The Langevin equation is given by Eq. (61) and the noise is defined by Eq. (62). Since the noise is assumed to be Gaussian, Eq. (62) is a complete specification. An equivalent way of giving it is through the pdf (Feynman and Hibbs, 1965):

$$P[\mathcal{F}]\, \mathcal{D}\mathcal{F} \propto \exp\left(-\frac{1}{4\tilde{D}} \int dt \mathcal{F}^2(t) \right) \mathcal{D}\mathcal{F} , \qquad (79)$$

where $\mathcal{D}\mathcal{F}$ is the functional measure. The idea is now to regard the Langevin equation (61) as defining a mapping $\mathcal{F} \mapsto x$. The pdf for the x variable is then given by

$$\begin{aligned} P[x] &= P[\mathcal{F}]|_{\mathcal{F}=\dot{x}+V'(x)}\, J[x] \\ &\propto \exp\left(-\frac{1}{4\tilde{D}} \int dt \, [\dot{x} + V'(x)]^2 \right) J[x] , \end{aligned} \qquad (80)$$

where

$$J[x] = \det\left[\frac{\delta\mathcal{F}}{\delta x}\right], \tag{81}$$

is the Jacobian of the transformation. An explicit expression for the Jacobian may be obtained either by direct calculation of a discretised form of the Langevin equation (Graham, 1975) or through use of the identity relating the determinant of a matrix to the exponential of the trace of the logarithm of that matrix (Zinn-Justin, 2002). One finds that $J[x] \propto \exp\{\theta(0)\int dt\, V''(x)\}$. The quantity $\theta(0)$ is once again the indeterminate value of the Heaviside theta function $\theta(x)$ at $x = 0$. Its appearance is a reflection of the fact that, due to the Brownian-like nature of the paths in the functional integral, the nature of the discretisation appears explicitly through this factor (Schulman, 1981). If we consistently use the mid-point rule throughout, then we may take $\theta(0) = 1/2$, which gives

$$P[x] \propto \exp\left(-\frac{1}{4\tilde{D}}\int dt\,[\dot{x} + V'(x)]^2 + \frac{1}{2}\int dt\, V''(x)\right) = \exp\left(-S[x]/\tilde{D}\right). \tag{82}$$

All quantities of interest can now be found from expression (82). For example, the conditional probability distribution, $P(x, t|x_0, t_0)$ is given by

$$\langle\delta(x - x(t))\rangle_{x(t_0)=x_0} = \int_{x(t_0)=x_0}\mathcal{D}x\,\delta(x - x(t))\,P[x]. \tag{83}$$

The expression (82) has much in common with Feynman's formulation of quantum mechanics as a path-integral (Feynman, 1948). In fact another way to obtain the result is to exploit the transformation (49) to write the Fokker-Planck equation (70) as a Schrödinger equation in imaginary time $\tau = it$, with a potential $U(x) = (1/2)[V'(x)]^2 - \tilde{D}V''(x)$, following Eq. (50). The action in the quantum-mechanical path-integral is

$$\frac{i}{\hbar}\int dt\,\left[\frac{1}{2}\dot{x}^2 - U(x)\right] \longrightarrow \frac{1}{2\tilde{D}}\int d\tau\,\left[-\frac{1}{2}\dot{x}^2 - \frac{1}{2}\left[V'(x)\right]^2 + \tilde{D}V''(x)\right], \tag{84}$$

which is Eq. (82) since $\int_{t_0}^{t} dt\,\dot{x}V'(x) = \int_{x_0}^{x} dx\,V'(x) = V(x) - V(x_0)$ does not depend on the path, only on the end-points. The functional $S[x]$ is analogous to the action in classical mechanics, and is frequently referred to as such. It is also sometimes referred to as the generalised Onsager-Machlup functional, in recognition of the original work carried out by Onsager and Machlup, in the case of a linear Langevin equation, in 1953.

The above discussion can be generalised in many ways. For example, if the Langevin equation for an $n-$dimensional process takes the form

$$\dot{x}_i = A_i(\mathbf{x}) + \zeta_i(t), \quad \langle\zeta_i(t)\zeta_j(t')\rangle = 2D_{ij}\delta(t - t'), \tag{85}$$

where $\zeta_i(t)$ is a Gaussian noise with zero mean and D_{ij} is independent of $\mathbf{x}$, then the general Onsager-Machlup functional is (Graham, 1975)

$$S[x] = \int dt\,\left[\frac{1}{4}\sum_{i,j}\left\{\dot{x}_i - A_i(\mathbf{x})\right\}D^{-1}_{ij}\left\{\dot{x}_j - A_j(\mathbf{x})\right\} + \frac{1}{2}\sum_i\frac{\partial A_i}{\partial x_i}\right], \tag{86}$$

if the matrix D_{ij} is non-singular. The generalisation to the situation where the noise is multiplicative is more complicated, and is analogous to the path-integral formulation of quantum mechanics in curved space (Graham, 1977).

VIII. System Size Expansion

In Example 2 of Section V we explicitly showed how the master equation may have different limits when the size of the system, N, becomes large. In one case both the first and second jump moments were of the same order (and much larger than the higher jump moments) and so a nonlinear Fokker-Planck equation of the diffusion type was obtained in the limit $N \to \infty$. In another case, the first jump moment scaled in a different way to the second moment, and so the $N \to \infty$ limit gave a deterministic macroscopic equation of the form $\dot{x} = f(x)$, with finite N effects presumably consisting of $1/\sqrt{N}$ fluctuations about the macroscopic state, $x(t)$. It is the second scenario that we will explore in this section. It can be formalised by writing

$$\frac{n}{N} = x(t) + \frac{\xi}{\sqrt{N}}\,, \tag{87}$$

and substituting this into the master equation, then equating terms of the same order in $1/\sqrt{N}$. The leading order equation obtained in this way will be the macroscopic equation, and the function $f(x)$ will emerge from the analysis. The next-to-leading order equation turns out to be a *linear* Fokker-Planck equation in the variable ξ. Higher order terms may also be included. This formalism was first developed by van Kampen (1961) and is usually referred to as van Kampen's system-size expansion. We will describe it in the specific case of a one-step process for a single stochastic variable in order to bring out the essential features of the method.

When using this formalism it is useful to rewrite the master equation (23) using step operators which act on an arbitrary function of n according to $\mathcal{E}f(n) = f(n+1)$ and $\mathcal{E}^{-1}f(n) = f(n-1)$. This gives

$$\frac{dP(n,t)}{dt} = (\mathcal{E}-1)\left[T(n-1|n)P(n,t)\right] + \left(\mathcal{E}^{-1}-1\right)\left[T(n+1|n)P(n,t)\right]\,. \tag{88}$$

We begin by using Eq. (87) to write the pdf which appears in the master equation (88) as

$$P(Nx(t)+\sqrt{N}\xi,t) = \Pi(\xi,t) \quad \Rightarrow \dot{P} = \frac{\partial \Pi}{\partial t} - N^{1/2}\frac{dx}{dt}\frac{\partial \Pi}{\partial \xi}\,. \tag{89}$$

This gives an expression for the left-hand side of the master equation, $\dot{P}$. To get an expression for the right-hand side,

(a) the step operators are expanded in powers of $1/\sqrt{N}$ (van Kampen, 1992):

$$\mathcal{E}^{\pm 1} = 1 \pm \frac{1}{\sqrt{N}}\frac{\partial}{\partial \xi} + \frac{1}{2!}\frac{1}{N}\frac{\partial^2}{\partial \xi^2} + \mathcal{O}\left(\frac{1}{N^{3/2}}\right), \tag{90}$$

(b) $T(n \pm 1|n)$ is expressed in terms of ξ and N,

(c) $P(n,t)$ is replaced by $\Pi(\xi,t)$.

Steps (a), (b) and (c) gives the right-hand side of the master equation as a power-series in $1/\sqrt{N}$. Equating the left-hand and right-hand sides order by order in $1/\sqrt{N}$ (this may require a rescaling of the time, t, by a power of $\sqrt{N}$), gives to leading order (the $\partial\Pi/\partial\xi$ cancels) an equation of the form $dx/dt = f(x)$. This may be solved subject to the

condition $x(0) = x_0 = n_0/N$, if we take the initial condition on the master equation to be $P(n,0) = \delta_{n,n_0}$. We denote the solution of this macroscopic equation by $x_M(t)$.

To next order in $1/\sqrt{N}$, the Fokker-Planck equation

$$\frac{\partial \Pi}{\partial t} = -f'(x)\frac{\partial}{\partial \xi}\,[\xi \Pi] + \frac{1}{2}g(x)\frac{\partial^2 \Pi}{\partial \xi^2}\,, \tag{91}$$

describing a linear stochastic process is found. The functions $f'(x)$ and $g(x)$ are to be evaluated when $x = x_M(t)$, and so are simply functions of time. If the macroscopic system tends to a fixed point: $x_M(t) \to x^*$, as $t \to \infty$, then $f'(x)$ and $g(x)$ may be replaced by constants in order to study the fluctuations about this stationary state.

To illustrate the method we use Example 3, Section IV. Equating both sides of the master equation in this case one finds

$$\begin{aligned} -N^{1/2}\frac{dx}{dt}\frac{\partial \Pi}{\partial \xi} \quad &+ \quad \frac{\partial \Pi}{\partial t} \;=\; \frac{1}{\sqrt{N}}\,[f_-(x) - f_+(x)]\,\frac{\partial \Pi}{\partial \xi} \\ +\frac{1}{2}\frac{1}{N}\,[f_-(x) + f_+(x)]\,\frac{\partial^2 \Pi}{\partial \xi^2} \quad &+ \quad \frac{1}{N}\left[f'_-(x) - f'_+(x)\right]\frac{\partial}{\partial \xi}\,[\xi P] + \ldots\,, \end{aligned} \tag{92}$$

where the functions $f_-(x)$ and $f_+(x)$ are given by:

$$f_-(x) = cx^2 + dx\,, \quad f_+(x) = 2bx(1-x)\,. \tag{93}$$

For the left- and right-hand sides to balance in Eq. (92), a rescaled time $\tau = t/N$ needs to be introduced. Then to leading order one finds $dx/d\tau = f(x)$, where $f(x) = f_+(x) - f_-(x)$. At next to leading order Eq. (91) is found with $g(x) = f_+(x) + f_-(x)$. The explicit form of the macroscopic equation is

$$\frac{dx}{d\tau} = x\,(r - ax)\,, \tag{94}$$

where $r = 2b - d$ and $a = 2b + c$. Eq. (94) is the logistic equation, which is the usual phenomenological way to model intraspecies competition.

Since the Fokker-Planck equation (91) describes a linear process, its solution is a Gaussian. This means that the probability distribution $\Pi(\xi, t)$ is completely specified by the first two moments $\langle \xi(t) \rangle$ and $\langle \xi^2(t) \rangle$. Multiplying Eq. (91) by ξ and ξ^2 and integrating over all ξ one finds

$$\frac{d}{dt}\langle \xi(t) \rangle = f'(x_M(t))\langle \xi(t) \rangle\,, \quad \frac{d}{dt}\langle \xi^2(t) \rangle = 2f'(x_M(t))\langle \xi^2(t) \rangle + g(x_M(t))\,. \tag{95}$$

We have chosen the initial condition to be $x_0 = n_0/N$, which implies that $\xi(0) = 0$. The first equation in (95) then implies that $\langle \xi(t) \rangle = 0$ for all t. Multiplying the second equation by $f^{-2}(x_M(t))$ one finds that

$$\langle \xi^2(t) \rangle = f^2(x_M(t)) \int_0^t dt'\, \frac{g(x_M(t'))}{f^2(x_M(t'))}\,, \tag{96}$$

and so the determination of $\langle \xi^2(t) \rangle$ is reduced to quadrature. The method can be applied to systems with more than one stochastic variable and those which are not one-step processes. Details are given in van Kampen's book (1992).

IX. Future Directions

This article has focused largely on classical topics in the theory of stochastic processes, since these form the foundations on which the subject is built. Much of the current work, and one would expect future work, will be numerical in character. Some of this will begin from a basic Markovian description in the form of reactions among chemical species — even if the system is not chemical in nature (Example 3 of Section IV is an example). A straightforward algorithm developed by Gillespie (1976, 1977), and since then extended and improved (Gillespie 2001, Cao et. al. 2004), provides an efficient way of simulating such systems. It thus provides a valuable method of investigating systems which may be formulated as complicated multivariable master equations, which complements the methods we have discussed here. However, many current studies do not begin from a system which can be described in this way, and there is every indication that this will be more true in the future. For instance, in agent based models the stochastic element may be due to mutations in characteristics, traits or behaviour, which may be difficult or impossible to formulate mathematically. Such agent based models are certainly individually based, but each individual may have different attributes and generally behave in such a complex way that only numerical simulations can be used to explore the behaviour of the system as a whole. Although these complex systems may be used to model more realistic situations, the well-known problems associated with the large number of parameters typically required to describe such systems, will mean that simplified versions will need to be analysed in order to understand them at a deeper level. These simpler models are likely to include those where the agents of a particular species are essentially identical. In this article we have discussed how the classical equations of the theory of stochastic processes, such as the Fokker-Planck equation, can be obtained from such models. They will therefore form a bridge between the agent-based approaches which are expected to become more prevalent in the future, and the analytic approaches which lie at the heart of the theory of stochastic processes.

X. Bibliography

Books and Reviews

1. Abramowitz, M. and Stegun, I. (eds) (1965) Handbook of Mathematical Functions. Dover, New York.
2. Chandrasekhar, S. (1943) Stochastic Problems in Physics and Astronomy. Rev. Mod. Phys. 15: 1-89. Reprinted in Wax (1954).
3. Cox, D. R. and Miller, H. D. (1968) The Theory of Stochastic Processes. Chapman and Hall, London. Chapter 3.
4. Crow, J. F. and Kimura, M. (1970) An Introduction to Population Genetics Theory. Harper and Row, New York.
5. Feller, W. (1968) An Introduction to Probability Theory and its Applications. Third Edition. Wiley, New York. Chapter XV.
6. Feynman, R. P. and Hibbs, A. R. (1965) Quantum Mechanics and Path Integrals. McGraw-Hill, New York. Chapter 12.
7. Fisher, R. A. (1930) The Genetical Theory of Natural Selection. Clarendon Press, Oxford.
8. Gantmacher, F. R. (1959) The Theory of Matrices Vol. 2. Chelsea Publishing Co., New York. Chapter 13, Section 6.

9. Gardiner, C. W. (2004) Handbook of Stochastic Methods. Third Edition. Springer, Berlin.
10. Graham, R. (1975) Macroscopic theory of fluctuations and instabilities, in "Fluctuations, Instabilities, and Phase Transitions" (ed T. Riste). Plenum, New York. pp. 215–293.
11. Haken, H. (1983) Synergetics. Springer-Verlag, Berlin. Section 6.3.
12. May, R. M. (1973) Model Ecosystems. Princeton University Press, Princeton. Chapter 5.
13. Moran, P. A. P. (1962) The Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford. Chapter 4.
14. Reichl, L. E. (1998) A Modern Course in Statistical Physics. Second Edition. Wiley, New York. Chapter 5.
15. Risken, H. (1989) The Fokker-Planck Equation. Second Edition. Springer, Berlin.
16. Schiff, L. I. (1968) Quantum Mechanics. Third Edition. McGraw-Hill, Tokyo. Chapter 4.
17. Schulman, L. S. (1981) Techniques and Applications of Path-Integration. Wiley, New York. Chapter 5.
18. Sneddon, I. N. (1957) Elements of Partial Differential Equations. McGraw-Hill, New York. Chapter 2.
19. van Kampen, N. G. (1992) Stochastic Processes in Physics and Chemistry. Second Edition. North-Holland, Amsterdam.
20. Wax, N. (1954) Selected Papers on Noise and Stochastic Processes. Dover, New York.
21. Zinn-Justin, J. (2002) Quantum Field Theory and Critical Phenomena. Fourth Edition. Clarendon Press, Oxford. Section 4.8.2.

Primary Literature

1. Bachelier, L. (1900) Théorie de la spéculation. Annales Scientifiques de L'Ecole Normale Supérieure III-17: 21–86.
2. Cao, Y., Li, H. and Petzold, L. (2004) Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. J. Chem. Phys. 121: 4059–4067.
3. Ehrenfest, P. and Ehrenfest, T. (1907) Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. Physick Zeits. 8: 311–314.
4. Einstein, A. (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Ann. Physik 17: 549–560. For a translation see: A. Einstein, "Investigations on the Theory of the Brownian Movement" R. Fürth (ed), A. D. Cowper (tr) (Dover, New York, 1956). Chapter I.
5. Einstein, A. (1906) Zur Theorie der Brownschen Bewegung. Ann. Physik 19: 371–381. For a translation see: A. Einstein, "Investigations on the Theory of the Brownian Movement" R. Fürth (ed), A. D. Cowper (tr) (Dover, New York, 1956). Chapter II.
6. Feynman, R. P. (1948) Space-time approach to non-relativistic quantum mechanics. Rev. Mod. Phys. 20: 367–387.

7. Fokker, A. D. (1914) Die mittlere Energie rotierende elektrischer Dipole im Strahlungsfeld. Ann. Physik 43: 810–820.
8. Gillespie, D. T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22: 403–434.
9. Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81: 2340–2361.
10. Gillespie, D. T. (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115: 1716–1733.
11. Graham, R. (1977) Path integral formulation of general diffusion processes. Z. Physik B26: 281–290.
12. Kac, M. (1947) Random walk and the theory of Brownian motion. Amer. Math. Monthly 54: 369–391.
13. Karlin, S. and McGregor, J. (1964) On some stochastic models in genetics, in Stochastic Problems in Medicine and Biology. J. Gurland (ed). University of Wisconsin Press, Madison. pp. 245–279.
14. Kendall, D. G. (1948) On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. Biometrika 35: 6–15.
15. Kolmogorov, A. N. (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechung. Math. Ann. 104: 415–458.
16. Kolmogorov, A. N. (1936) Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen. Mat. Sbornik (N.S.) 1: 607–610.
17. Krafft, O. and Schaefer, M. (1993) Mean passage times for tridiagonal transition matrices and a two-parameter Ehrenfest urn model. J. Appl. Prob. 30: 964–970.
18. Kramers, H. A. (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7: 284–304.
19. Langevin, P. (1908) Sur la théorie du mouvement brownien. C. R. Acad. Sci. Paris 146: 530–533. For a translation see: D. S. Lemons and A. Gythiel, Am. J. Phys. 65: 1079-1081 (1997).
20. Malécot, G. (1944) Sur un problème de probabilités en chaine que pose la génétique. C. R. Acad. Sci. Paris 219: 379–381.
21. Markov, A. A. (1906) Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. Izv. Fiz.-Matem. Obsch. Kazan Univ. (Series 2) 15: 135-156. See also: "Extension of the limit theorems of probability theory to a sum of variables connected in a chain", in Appendix B of R. Howard "Dynamic Probabilistic Systems, volume 1: Markov Chains" (John Wiley and Sons, 1971).
22. McKane, A. J. and Newman, T. J. (2004) Stochastic models in population biology and their deterministic analogs. Phys. Rev. E 70: 041902.
23. McKane, A. J. and Newman, T. J. (2005) Predator-prey cycles from resonant amplification of demographic stochasticity. Phys. Rev. Lett. 94: 218102.
24. Moran, P. A. P. (1958) Random processes in genetics. Proc. Cambridge Philos. Soc. 54: 60–72.
25. Moyal, J. E. (1949) Stochastic processes and statistical physics. J. Roy. Stat. Soc. (London) B 11: 150–210.
26. Nordsieck, A., Lamb, W. E. Jr. and Uhlenbeck, G. E. (1940) On the theory of cosmic-ray showers. I. The Furry model and the fluctuation problem. Physica 7:

344–360.

27. Onsager, L. and Machlup, S. (1953) Fluctuations and irreversible processes. Phys. Rev. 91: 1505–1512.
28. Pauli W. (1928) Probleme der modernen Physik. Festschrift zum 60. Geburtstag A. Sommerfeld, edited by P. Debye (Hirzel, Leipzig). p. 30.
29. Pearson, K. (1905) The problem of the random walk. Nature 72: 294 and 342.
30. Planck, M. (1917) Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie. Sitzungsber. Preuss. Akad. Wiss. pp. 324–341.
31. Rayleigh, Lord (1891) Dynamical problems in illustration of the theory of gases. Phil. Mag. 32: 424–445.
32. Siegert, A. J. F. (1949) On the approach to statistical equilibrium. Phys. Rev. 76: 1708–1714.
33. Smoluchowski, M. v. (1906) Zur kinetishen Theorie der Brownschen Molekularbewegung und der Suspensionen. Ann. Physik 21: 756–780.
34. Smoluchowski, M. v. (1916) Drei Vortage über Diffusion, Brownsche Bewegung, und Koagulation von Kolloidteilchen. Physik Zeits. 17: 557, 585.
35. van Kampen, N. G. (1961) A power series expansion of the master equation. Can. J. Phys. 39: 551–567.
36. van Kampen, N. G. (1981) Itô versus Stratonovich. J. Stat. Phys. 24: 175–187.
37. Wiener, N. (1921a) The average of an analytic functional. Proc. Natl. Acad. Sci. USA 7: 253–260.
38. Wiener, N. (1921b) The average of an analytic functional and the Brownian movement. Proc. Natl. Acad. Sci. USA 7: 294–298.
39. Wright, S. (1931) Evolution in Mendelian populations, Genetics 16: 97–159.
40. Zwanzig, R. (1973) Nonlinear generalized Langevin equations. J. Stat. Phys. 9: 215–220.