Chapter 4

Stochastic Processes

4.1 Definition

In the previous chapter we studied random variables as functions on a sample space $X(\omega), \omega \in \Omega$, without regard to how these might depend on parameters. We now want to study more complicated situations in which probability can evolve with time. This defines a *stochastic process* or *random process*.

Recall that a random variable is a function defined on the sample space, it assigns a number to an event $X(\omega) \in R$. A stochastic process is a family of random variables depending on a real parameter, i.e. a stochastic process is a function of two varaiables, one which is a point in the sample space, the other which is a real variable usually the time.

There are three equivalent ways to look on a stochastic process.

i) as a function of two variables $X(\omega, t)$ where $\omega \in \Omega$ and t denotes time.

ii) for a fixed value of t, X is a random variable. For each t it is a different random variable; we can regard X as a family of random variables indexed by the variable t.

iii) we may also consider X as a family of functions of t one for each fixed ω. $X(t)$ for a fixed ω is called a *realization* or *sample function* of the process.

The parameter t above can belong to the real line or belong to some countable subset such as the integers, i.e. be a discrete parameter. In most applications t denotes time.

Example 4.1 *Let X be a random variable and $f(t)$ a given function of time. Then*

$$Y(t) = f(t)X$$

is a stochastic process.

Example 4.2 Let X, ω and δ be random variables then

$$Y(t) = X \sin(\omega t + \delta)$$

is a stochastic process. It corresponds to an oscillation with random amplitude, frequancy and phase. Of course, it will still be a stochastic process if only one or two of these three quantirtties is random.

Example 4.3 *A coin tossed n times. The number of heads is a random variable which depends on the real parameter n. It is therefore a stochastic process in discrete time.*

In general once a random variable X is defined we can define a stochastic process by some mapping f

$$Y_X(t) = f(X(\omega), t) \tag{4.1}$$

A sample function or realization of Y is obtained with $X(\omega) = x$

$$Y_x(t) = f(x, t) \tag{4.2}$$

With $Y_X(t)$ defines as in (4.1) we can form averages knowing the probability density $\rho_X(x)$ of X. For instance

$$\langle Y_X(t) \rangle = \int Y_X(t) \rho_X(x) \mathrm{d}x \tag{4.3}$$

More generally, take n-values $t_1, t_2, \ldots, t_n$ for the time variable and form the n-th moment

$$\langle Y(t_1) Y(t_2) \cdots Y(t_n) \rangle = \int Y_X(t_1) Y_X(t_2) \cdots Y_X(t_n) \rho_X(x) \mathrm{d}x \tag{4.4}$$

Of particular interest is the auto-correlation function of a stochastic process $X(t)$

$$\begin{aligned} C(t_1, t_2) &= \langle [X(t_1) - \langle X(t_1) \rangle] \rangle \, [X(t_2) - \langle X(t_2) \rangle] \rangle \\ &= \langle X(t_1) X(t_2) \rangle - \langle X(t_1) \rangle \langle X(t_2) \rangle \end{aligned} \tag{4.5}$$

For $t_1 = t_2 = t$ it reduces to the time-dependent variance

$$\sigma^2(t) = \langle [X(t) - \langle X(t) \rangle]^2 \rangle \tag{4.6}$$

4.2 Distribution Functions

Usually we cannot write explicit formulas for random variables. Since stochastic processes are sets of random variables, it follows that we shall usually not be able to write explicit expressions for them either as in (4.1). We can characterize them in terms of their probability distribution functions.

Choose some finite set of times $t_1, t_2, \ldots, t_n$. Then $X(t_1), X(t_2), \ldots, X(t_n)$ are a set of random variables, and we can specify their joint probability distribution function. So we have the density functions

$$\begin{aligned} &\rho_1(x_1, t_1) \\ &\rho_2(x_1, t_1; x_2, t_2) \\ &\rho_3(x_1, t_1; x_2, t_2; x_3, t_3) \\ &\vdots \\ &\rho_n(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) \end{aligned} \tag{4.7}$$

Here $\rho_1(x_1, t_1)\mathrm{d}x_1$ is the probability that the random variable $X(t_1)$ has the value x_1 in the range $\mathrm{d}x_1$. $\rho_2(x_1, t_1; x_2, t_2)\mathrm{d}x_1\mathrm{d}x_2$ is the joint probability that the two random variables

$X(t_1)$ and $X(t_2)$ have the values x_1 and x_2 in the ranges $\mathrm{d}x_1$ and $\mathrm{d}x_2$ respectively. The ρ_n with $n > 2$ have similar meanings.

For a process defined as in (4.1) we find

$$\rho_1(y,t) = \int \delta(y - Y_X(t))\rho_X(x)\mathrm{d}x \tag{4.8}$$

and in general

$$\begin{aligned} &\rho_n(y_1,t_1;y_2,t_2;\ldots;y_n,t_n;) \\ = &\int \delta(y_1 - Y_X(t_1))\delta(y_2 - Y_X(t_2))\cdots\delta(y_n - Y_X(t_n))\rho_X(x)\mathrm{d}x \end{aligned} \tag{4.9}$$

From the distribution functions we can calculate averages as

$$\langle X(t_1)X(t_2)\cdots X(t_n)\rangle = \int x_1x_2\cdots x_n\rho_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \tag{4.10}$$

The hierarchy of functions ρ_n obeys the following consistency conditions:

i) $\rho_n \geq 0$,

ii) $\rho_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n) = \rho_n(x_i,t_i;\ldots;x_j,t_j)$, i.e. they are symmetric upon permutation of the variables,

iii)

$$\rho_n(x_1,t_1;\ldots;x_n,t_n) = \int \rho_{n+1}(x_1,t_1;\ldots;x_n,t_n;x_{n+1},t_{n+1})\mathrm{d}x_{n+1}$$

iv) $\int \rho_1(x_1,t_1)\mathrm{d}x_1 = 1$.

It has been proved by Kolmogorov that any set of functions obeying these consistency conditions determines a stochastic process $X(\omega,t)$.

A process is called stationary if

$$\rho_n(x_1,t_1+\tau;x_2,t_2+\tau;\ldots;x_n,t_n+\tau) = \rho_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n) \tag{4.11}$$

for alla n and τ. For a *stationary process*

$$\rho_1(x_1,t_1) = \rho_1(x_1) \tag{4.12}$$

and

$$\langle X(t_1)X(t_2)\rangle = \langle X(0)X(t_2-t_1)\rangle \tag{4.13}$$

In particular, the mean value of a stationary stochastic process is independent of time $\langle X(t)\rangle = \text{const.}$.

If a process has a constant first moment or expectation, and an autocorrelation function $C(t_2,t_1)$ that depends only on $t_2 - t_1$, it is called *wide sense stationary*. This is a weaker condition than stationarity, since it imposes no restrictions on the distribution functions of order greater than two.

We shall also introduce a *conditional probability* density

$\rho_{1,1}(x_2, t_2|x_1, t_1)$ =conditional probability density for $X(t)$ to have the value x_2 at t_2 given that it had x_1 at t_1

which is defined as

$$\rho_{1,1}(x_2t_2|x_1t_1) = \frac{\rho_2(x_1t_1; x_2t_2)}{\rho_1(x_1t_1)} \tag{4.14}$$

Clearly $\rho_{1,1}$ is nonnegative and normalized

$$\int \rho_{1,1}(x_2t_2|x_1t_1)\mathrm{d}x_2 = 1$$

Integrating (4.14) over x_1 also gives

$$\rho_1(x_2t_2) = \int \rho_{1,1}(x_2t_2|x_1t_1)\rho_1(x_1t_1)\mathrm{d}x_1 \tag{4.15}$$

More generally one may fix the values of X at k different times $t_1, \ldots, t_k$ and ask for the joint probability at l other times $t_{k+1}, \ldots, t_{k+l}$. This leads to the general definition of the conditional probability $\rho_{l,k}$

$$\begin{aligned} & \rho_{l,k}(x_{k+1}t_{k+1}; \ldots; x_{k+l}t_{k+l}|x_1t_1; \ldots; x_kt_k) \\ = \quad & \frac{\rho_{k+l}(x_1t_1; \ldots; x_kt_k; x_{k+1}t_{k+1}; \ldots; x_{k+l}t_{k+l})}{\rho_k(x_1t_1; \ldots; x_kt_k)} \end{aligned} \tag{4.16}$$

By definition $\rho_{k,l}$ is symmetric in the set of k pair of variables, and also in the set of l pairs of variables.

4.3 Sample path properties

A stochastic process is not a single function, but a family of functions, its sample functions. In order to describe limiting operations on members of this family, we need to introduce the concept of convergence of sequences of random variables.

There are several types of convergence that arise for random variables. Let $\{X_n\}$ be a sequence of random variables and X som fixed random variable.

1. The sequence $\{X_n\}$ is said to converge to X *almost certainly* if $|X_n - X| \to 0$ for all sufficiently large n, except on some set of events of probability zero. This ode of convergence is sometimes called strong convergence.

2. The sequence $\{X_n\}$ converges to X in probability if

   $$\Pr\{|X_n - X|\} \to 0 \quad \text{as} \quad n \to \infty$$

   This mode of convergence is sometimes called weak convergence.

   The difference between almost certain convergence and convergence in probability is that in almost certain convergence, the set on which the sequence does not approach X settles down to some fixed set at zero probability. For convergence in probability the set where $|X_n - X|$ is not zero also becomes of zero probability, but it need not be fixed. It may move around in the event space Ω as n increases, never settling down.

3. The sequence $\{X_n\} \to X$ in mean square if

$$\langle |X_n - X|^2 \rangle \to 0, \text{as} \quad n \to \infty$$

This mode of convergence is analogous to mean square convergence in vector spaces. It is the kind of convergence for which most of the results of stochastic calclus have been derived.

4. $\{X_n\}$ is said to converge to X in *distribution* if the cumulative distribution function of the X_n approach the cumulative distribution function of X at all continuity points of the latter function, i.e.

$$F_{X_n}(x) \to F_X(x), \quad n \to \infty$$

Since the cumulative distribution functions are ordinary functions their convergence follows from usual analysis. The central limit theorem is an example of convergence in distribution.

We may now ask what does it mean to say that a stochastic process is continous? We shall say that a stochastic process is continous if, with probability one, all of its sample functions are continous functions of t.

Sufficient conditions to determine whether a random process is or is not continous are known, but are not easy to apply; it is often quite difficult to verify the hypotheses of the continuity criteria. However for Markov processes (section 2.6) there is a relatively simple criterion. If

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x-y|>\epsilon} \varrho 1, 1(x, t + \Delta t | yt) \mathrm{d}x = 0 \tag{4.17}$$

for all $\epsilon > 0$, then the process $X(t)$ is continous. This condition means that finite jumps of arbitrarily small size becom very improbable for sufficiently short time intervals.

Example 4.4 *The transition probability*

$$\varrho_{1,1}(xt|y0) = \left(\frac{1}{4\pi D t}\right)^{1/2} \mathrm{e}^{-(x-y)^2/2Dt}$$

satisfies (4.17) and yields sample functions which are very jagged but continous. On the other hand

$$\varrho_{1,1}(x, t + \Delta t | yt) = \frac{\Delta t}{\pi} \frac{1}{(x-y)^2 + (\Delta t)^2}$$

the Cauchy distribution, does not satisfy (4.17) and yields sample functions that have many discontinous jumps of varying amplitudes.

Differentiability of sample functions can be studied by the ordinary methods of analysis since sample functions are ordinary functions.

Integration is another matter. Integrals of random processes are very important, especially in the theory of Brownian motion.

Let $X(t)$ be a random process, $x(t)$, one of its sample functions, and $f(t)$ some fixed function. Then we define

$$\int_a^b \mathrm{d}t f(t) X(t) = \lim_{\delta \to 0} \sum_k f(t_\xi) x(t_\xi)(t_k - t_{k-1}) \quad \text{for all sample functions } x(t) \tag{4.18}$$

where $t_{k-1} < t_\xi < t_k$ and $\delta = |t_k - t_{k-1}|$. Here $X(t)$ is taken over all sample functions, so the integral is a new stochastic process. The quantity on the rhs is the familiar Riemann sum defining the integral of the sample function $x(t)$. If the sample functions are integrable for example if X is continous, then the integral of X defined in (4.18) will exist.

Similarly if the sample functions of $X(t)$ are of bounded variation, i.e.

$$V_X(t) = \lim_{\delta \to 0} \sum_k |x(t_k) - x(t_{k-1})| < \infty$$

one can define the Stieltjes integral

$$\int f(t) \mathrm{d}X(t) = \lim_{\delta \to 0} \sum_k f(t_k) \left[x(t_k) - x(t_{k-1}) \right]$$

However, it is a common occurence, that the process we may want to integrate are not of bounded variation. We usually need to integrate very wildly varying functions. We would like to define integrals of the structure

$$\int \phi(X(t)) \mathrm{d}X(t)$$

where $X(t)$ is not of bounded variation. We might try to define them as limits of Rieman-Stieltjes sums

$$\lim_{\delta \to 0} \sum_k \phi(t_\xi) \left(x(t_k) - x(t_{k-1}) \right)$$

If the limit exists, is independent of the choice of $t_{k-1} < t_\xi < t_k$, then we say that the integral exists and is equal to the limit. Unfortunately for the wildly varying functions arising in Brownian motion theory for instance, while the limit often exists, it is *not* independent of the choice of the intermediate points t_ξ. If $x(t)$ varies very rapidly in the interval $t_k - t_{k-1}$, no matter how small the interval, then the limit of the Rieman-Stieltjes sum will clearly depend on how those intermediate points are chosen. Notice that this problem does not arise if the function ϕ does not depend on X, but is merely a function of t.

This difficulty can be circumvented by specifying how the t_ξ points are to be chosen in constructing the sums. There are several ways of doing this, and each of them will give rise to a different definition of the integral.

In Ito's definition t_ξ is always chosen to be t_k the at the end of the k'th interval. In Stratonovich definition $t_\xi = (t_k + t_{k-1})/2$ i.e. the midpoint of the jk'th interval.

4.4 Fourier analysis

In the analysis of ordinary functions, it is often useful to decompose the function into frequency components, that is to construct either a Fourier series or a Fourier integral representation of the function. The same thing can be done with random processes. Since random processes are families of functions, not single functions, one will only be able to determine statistical properties of the Fourier coefficients or transforms.

The sample functions of a random process are generally not periodic functions. Therefore they cannot dbe developed in Fourier series in the conventional way. Nor do these

functions vanish as $t \to \pm\infty$, and they do not have Fourier integrals in the conventional sense.

Let $x(t)$ be a real sample function of the stochastic process $X(t)$, with $\langle X(t)\rangle = 0$. Let T be some time and define a new function

$$x_T(t) = \begin{cases} x(t), & -\frac{T}{2} \le t \le \frac{T}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.19}$$

In other words x_T is a clipped version of x. The Fourier transform is

$$\begin{aligned} A_T(\omega) &= \int_{-T/2}^{T/2} \mathrm{d}t \mathrm{e}^{-\mathrm{i}\omega t} x_T(t) \\ x_T(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega \mathrm{e}^{\mathrm{i}\omega t} A_T(\omega) \end{aligned} \tag{4.20}$$

Since $x_T(t)$ is real we have $A_T(\omega) = A_T^*(\omega)$. Now

$$x_T(t+s)x_T(t) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}\omega \mathrm{d}\omega' \mathrm{e}^{\mathrm{i}\omega(t+s)} \mathrm{e}^{\mathrm{i}\omega' t)} A_T(\omega) A_T(\omega') \tag{4.21}$$

Suppose that $X(t)$ is a stationary process, then

$$\frac{1}{T}\int_{-T/2}^{T/2} \mathrm{d}t \langle x_T(t+s)x_T(t)\rangle = \frac{1}{T}\int_{-T/2}^{T/2} \mathrm{d}t \langle x_T(s)x_T(0)\rangle = \langle x_T(s)x_T(0)\rangle$$

and so

$$\langle x_T(s)x_T(0)\rangle = \frac{1}{T}\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega s}\langle A_T(\omega)A_T^*(\omega')\rangle \left\{\int_{-T/2}^{T/2} \mathrm{d}t \mathrm{e}^{\mathrm{i}(\omega-\omega')t)}\mathrm{d}t\right\} \mathrm{d}\omega \mathrm{d}\omega' \tag{4.22}$$

As T approaches infinity

$$\int_{-T/2}^{T/2} \mathrm{d}t \mathrm{e}^{\mathrm{i}(\omega-\omega')t)}\mathrm{d}t \to 2\pi\delta(\omega-\omega')$$

We also suppose that

$$\lim_{T\to\infty}\frac{1}{T}\langle |A_T(\omega)|^2\rangle = S(\omega)$$

exists. Then (4.22) becomes

$$\langle x(s)x(0)\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega \mathrm{e}^{\mathrm{i}\omega s} S(\omega) = \frac{1}{\pi}\int_0^{\infty} \mathrm{d}\omega \cos(\omega s) S(\omega) \tag{4.23}$$

The second equality in (4.23) follows since

$$\begin{aligned} S(-\omega) &= \lim_{T\to\infty}\frac{1}{T}\langle |A_T(-\omega)|^2\rangle = \lim_{T\to\infty}\frac{1}{T}\langle A_T(-\omega)A_T^*(-\omega)\rangle \\ &= \lim_{T\to\infty}\frac{1}{T}\langle A_T^*(\omega)A_T(\omega)\rangle = S(\omega) \end{aligned}$$

The quantity $S(\omega)$ is called the *power spectrum* or just the *spectrum* of the process $X(t)$.

Equation (4.23) is the Wiener-Khinchin theorem.

4.5 White noise

A random process whose spectrum $S(\omega)$ is independent of frequency is called *white noise*. The name comes from an analogy with white light, whose spectrum is independent of frequency.

If $X(t)$ is a white noice process, then

$$\langle x(s)x(0)\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega \mathrm{e}^{\mathrm{i}\omega s} S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega \mathrm{e}^{\mathrm{i}\omega s} S = S\delta(t) \tag{4.24}$$

i.e. $\langle x(t)x(0)\rangle$ is a delta function. Then $X(t+s)$ and $X(t)$ are uncorrelated no matter how small s is.

White noise can be considered to be a model for processes with short correlation times.

4.6 Classification of Stochastic Processes

There are many ways in which stochastic processes can be classified. For example they could be classified in terms of continuity, boundedness etc., i.e. properties of their sample functions. They could also be classified with respect to the properties of their distribution functions. However the classification which has been found to be most useful for our purposes is a classification according to their memory.

The simplest kund of process is one which has no memory of the past, or

$$\rho_n(x_1t_1; x_2t_2; \ldots; x_nt_n) = \prod_{i=1}^{n} \rho_i(x_it_i) \tag{4.25}$$

Such a process is called a completely random process. In terms of conditional probabilities this gives

$$\rho_{1,n-1}(x_1t_1|x_2t_2; \ldots; x_nt_n) = \rho_1(x_1t_1) \tag{4.26}$$

The random variables $X(t_1)$ and $X(t_2)$ ar independent when $t_1 \neq t_2$. More complicated processes has correlations at different times. Processes with short memory are called *Markov processes* where the process has memory only at its immediate past.

Let $t_1 < t_2 < \cdots < t_n$ then for a Markov process we have

$$\rho_{1,n-1}(x_nt_n|x_2t_2; \ldots; x_{n-1}t_{n-1}) = \rho_{1,1}(x_nt_n|x_{n-1}t_{n-1}) \tag{4.27}$$

That is the conditional probability density for x_n at t_n is fully determined by the value of x_{n-1} at t_{n-1} and is not affected by any knowledge of the stochastic variable $X(t)$ at earlier times. The conditional probability density $\rho_{1,1}$ is called the *transition probability*

A Markov process is fully determined by the two functions $\rho_1(x_1t_1)$ and $\rho_{1,1}(x_2t_2|x_1t_1)$. For example

$$\begin{aligned} \rho_3(x_3t_3; x_2t_2; x_1t_1) &= \rho_{1,2}(x_3t_3|x_2t_2; x_1t_1)\rho_2(x_2t_2; x_1t_1) \\ &= \rho_{1,1}(x_3t_3|x_2t_2)\rho_{1,1}(x_2t_2|x_1t_1)\rho_1(x_1t_1) \end{aligned} \tag{4.28}$$

In general

$$\begin{aligned} &\rho_n(x_nt_n; \ldots; x_1t_1) = \rho_{1,1}(x_nt_n|x_{n-1}t_{n-1})\rho_{1,1}(x_{n-1}t_{n-1}|x_{n-2}t_{n-2})\cdots \\ &\rho_{1,1}(x_2t_2|x_1t_1)\rho_1(x_1t_1) \end{aligned} \tag{4.29}$$

If we integrate (4.28) over x_2 assuming $t_1 < t_2 < t_3$ we obtain

$$\rho_2(x_3t_3|x_1t_1) = \int \rho_{1,1}(x_3t_3|x_2t_2)\rho_{1,1}(x_2t_2|x_1t_1)\mathrm{d}x_2\rho_1(x_1t_1) \tag{4.30}$$

and then

$$\rho_{1,1}(x_3t_3|x_1t_1) = \int \rho_{1,1}(x_3t_3|x_2t_2)\rho_{1,1}(x_2t_2|x_1t_1)\mathrm{d}x_2 \tag{4.31}$$

This is the *Chapman-Kolmogorov-Smoluchovski equation*. The probability of transition from x_1, t_1 to x_3, t_3 have been broken into two successive steps, first from x_1, t_1 to x_2, t_2 and then from x_2, t_2 to x_3, t_3. The successive steps are statistically independent. From (4.30) we also have

$$\rho_1(x_3t_3) = \int \rho_{1,1}(x_3t_3|x_2t_2)\rho_1(x_2t_2)\mathrm{d}x_2 \tag{4.32}$$

which determines ρ_1 in terms of the transition probability $\rho_{1,1}$.

4.7 The Fokker-Planck Equaton

From the Chapman-Kolmogorov equation (4.31) we have with $t_1 = 0, t_2 = t$ and $t_3 = t+s$,

$$\rho_{1,1}(y, t+s|x0) = \int \mathrm{d}z\rho_{1,1}(y, t+s|zt)\rho_{1,1}(zt|x0) \tag{4.33}$$

We assume now that $\rho_{1,1}(yt+s|zt)$ is a very sharply peaked function of $y-z$ when s is small. That is the system cannot change its state very much in a short time, there are no jumps. Then only values of z near y will contribute to the integral. Take some arbitary smooth function $\phi(y)$ wich vanish as $y \to \pm\infty$. Multiply by this function and integrate over y. We also expand $\phi(y)$ as

$$\phi(y) = \phi(z) + (y-z)\phi'(z) + \frac{1}{2}(y-z)^2\phi''(z) + \cdots$$

Then

$$\int \mathrm{d}y\phi(y)\rho_{1,1}(y, t+s|x0) = \int \mathrm{d}y \int \mathrm{d}z\phi(y)\rho_{1,1}(y, t+s|zt)\rho_{1,1}(zt|x0) \tag{4.34}$$

But if we intrchange the order of integration on the rhs we have

$$\begin{aligned}
& \int \mathrm{d}y\phi(y)\rho_{1,1}(y, t+s|zt) \\
= & \int \mathrm{d}y \left[\phi(z) + (y-z)\phi'(z) + \frac{1}{2}(y-z)^2\phi''(z) + \cdots\right] \rho_{1,1}(y, t+s|zt) \\
= & \phi(z) + \phi'(z) \int \mathrm{d}y(y-z)\rho_{1,1}(y, t+s|zt) \\
+ & \frac{1}{2}\phi''(z) \int \mathrm{d}y(y-z)^2\rho_{1,1}(y, t+s|zt) + \cdots
\end{aligned} \tag{4.35}$$

Let us now assume that the following limits exist

$$\begin{aligned} A(z) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int \mathrm{d}y(y-z)\rho_{1,1}(yt+s|zt) \\ B(z) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int \mathrm{d}y(y-z)^2\rho_{1,1}(yt+s|zt) \\ 0 &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int \mathrm{d}y(y-z)^n\rho_{1,1}(yt+s|zt); \quad n>2 \end{aligned} \tag{4.36}$$

Then for $s=\Delta t\to 0$

$$\begin{aligned} &\int \mathrm{d}y\phi(y)\rho_{1,1}(y,t+\Delta t|x0) \\ =\ &\int \mathrm{d}z\left[\phi(z)+\phi'(z)A(z)\Delta t+\frac{1}{2}\phi''(z)B(z)\Delta t\right]\rho_{1,1}(zt|x0) \end{aligned} \tag{4.37}$$

A rearrangement and integration by parts gives

$$\begin{aligned} &\int \mathrm{d}y\phi(y)\frac{\rho_{1,1}(yt+\Delta t|x0)-\rho_{1,1}(yt|x0)}{\Delta t} \\ =\ &\int \mathrm{d}y\phi(y)\left\{-\frac{\partial}{\partial y}\left[A(y)\rho_{1,1}(yt|x0)\right]+\frac{1}{2}\frac{\partial^2}{\partial y^2}\left[B(y)\rho_{1,1}(yt|x0)\right]\right\} \end{aligned} \tag{4.38}$$

Take $\Delta t\to 0$, and since $\phi(y)$ is an arbitrary smooth function we conclude that

$$\frac{\partial}{\partial t}\rho_{1,1}(yt|x0)=-\frac{\partial}{\partial y}\left[A(y)\rho_{1,1}(yt|x0)\right]+\frac{1}{2}\frac{\partial^2}{\partial y^2}\left[B(y)\rho_{1,1}(yt|x0)\right] \tag{4.39}$$

This is the *Fokker-Planck equation* or the *forward* equation.

We can also derive the *adjoint* or *backward* equation from the Chapman-Kolmogorov equation. From (4.31) we have with $0<t<t_1$

$$\rho_{1,1}(x_1t_1|x0)=\int \mathrm{d}y\rho_{1,1}(x_1t_1|yt)\rho_{1,1}(yt|x0) \tag{4.40}$$

Taking the derivative with respect to t gives

$$0=\int \mathrm{d}y\frac{\partial}{\partial t}\rho_{1,1}(x_1t_1|yt)\rho_{1,1}(yt|x0)+\int \mathrm{d}y\rho_{1,1}(x_1t_1|yt)\frac{\partial}{\partial t}\rho_{1,1}(yt|x0) \tag{4.41}$$

and from the forward equation together with partial integrations

$$\begin{aligned} &\int \mathrm{d}y\left[\frac{\partial}{\partial t}\rho_{1,1}(x_1t_1|yt)+A(y)\frac{\partial}{\partial y}\rho_{1,1}(x_1t_1|yt)\right. \\ +\ &\left.\frac{1}{2}B(y)\frac{\partial^2}{\partial y^2}\rho_{1,1}(x_1t_1|yt)\right]\rho_{1,1}(yt|x0)=0 \end{aligned}$$

which leads to

$$\frac{\partial}{\partial t}\rho_{1,1}(x_1t_1|yt)=-A(y)\frac{\partial}{\partial y}\rho_{1,1}(x_1t_1|yt)-\frac{1}{2}B(y)\frac{\partial^2}{\partial y^2}\rho_{1,1}(x_1t_1|yt) \tag{4.42}$$

4.8 The Wiener Process

A stochastic process $X(t)$, is said to be a process with independent increments if given $n+1$ times $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$ the random variables $X(0), X(t_1)-X(0), \ldots, X(t_n)-X(t_{n-1})$ are pairwise independent. A stochastic process is said to be a Gaussian process if all of its distribution functions are multivariate Gaussian distributions

$$\rho_n(x_1,t_1;x_2,t_2;\ldots,x_n,t_n) = \left[\frac{\det g}{(2\pi)^n}\right]^{1/2} \mathrm{e}^{-\frac{1}{2}x^T g x} \tag{4.43}$$

where x is a column vector with elements $(X(t_1),\ldots,X(t_n))$ and x^T the corresponding row vector. The matrix g is the inverse of the covariance matrix C where

$$C(t_i,t_j) = \langle X(t_i)X(t_j)\rangle$$

Here we have assumed that the stochastic process is centred so that $\langle X(t)\rangle = 0$.

A particular important Gaussian Markov process is the *Wiener process*, which is a set of Gaussian random variables $W(t)$ defined for $t \geq 0$ with the following properties:

i) $\langle W(t)\rangle = 0$,

ii) $W(0) = 0$,

iii) the increments $W(t_i) - W(t_j), t_i, t_j > 0$ are stationary and independent i.e.,

$$P\left[W(t_2+\tau) - W(t_1+\tau)\right] = P\left[W(t_2) - W(t_1)\right]$$

and $W(t_i) - W(t_j), W(t_k) - W(t_l)$ are independent for $t_i > t_j \geq t_k > t_l \geq 0$.

The Wiener process itself is not stationary since $W(0) = 0$. Since $W(t)$ is Gaussian and centered, so also are the increments.

Let us examine the variance of $W(t)$. Since its mean value is zero its variance is equal to its second moment. For any $t_1, t_2 > 0$

$$\begin{aligned}\mathrm{Var}\left[W(t_1+t_2\right] &= \langle[W(t_1+t_2)]^2\rangle = \langle[W(t_1+t_2) - W(t_1) + W(t_1 - W(0)]^2\rangle \\ &= \langle[W(t_1+t_2) - W(t_1)]^2\rangle + \langle[W(t_1) - W(0)]^2\rangle\end{aligned}$$

The last step follows since $W(t_1+t_2) - W(t_1)$ and $W(t_1) - W(0)$ are independent. From the stationary property of the increments the last equation gives

$$\mathrm{Var}\left[W(t_1+t_2\right] = \mathrm{Var}\left[W(t_2\right] + \mathrm{Var}\left[W(t_1)\right] \tag{4.44}$$

If we write

$$\mathrm{Var}\left[W(t)\right] = c(t) \tag{4.45}$$

equation (4.44) yields

$$c(t_1+t_2) = c(t_1) + c(t_2) \tag{4.46}$$

Hence for r a positive integer $c(r) = rc(1)$, and it immediately follows that this is true for r the quotient of two positive integers. It is also true for a positive irrational number, if we take this as the intersection of two limiting sequaences of rational numbers, $c(t)$ being

continous because $W(t)$ is. The function $c(t)$ is therefore proportional to t and we may express (4.45) as

$$\mathrm{Var}\left[W(t)\right] = \sigma^2 t \tag{4.47}$$

where σ^2 is positive and is connected to a diffusion constant. More generally

$$\mathrm{Var}\left[W(t) - W(s)\right] = \mathrm{Var}\left[W(t-s) - W(0)\right] = \mathrm{Var}\left[W(t-s)\right]$$

or

$$\mathrm{Var}\left[W(t-s)\right] = \sigma^2 |t-s| \tag{4.48}$$

since the variance is nonnegative. Since the increments are Gaussian we have the characteristic function

$$\phi(k; W(t) - W(s)) = \mathrm{e}^{-\frac{1}{2}\sigma^2 |t-s| k^2} \tag{4.49}$$

Consider now $W(s)$ and $W(t)$ with $0 \le s \le t$, then

$$\begin{aligned} \langle W(s)W(t)\rangle &= \langle W(s)[W(s) + W(t) - W(s)]\rangle \\ &= \langle W(s)W(s)\rangle + \langle [W(s) - W(0)][W(t) - W(s)]\rangle \\ &= \mathrm{Var}[W(s)] = \sigma^2 s \end{aligned}$$

Similarly when $0 \le t \le s$

$$\langle W(s)W(t)\rangle = \mathrm{Var}[W(s)] = \sigma^2 t$$

Both cases therefore give

$$\langle W(s)W(t)\rangle = \sigma^2 \min(s,t)$$

Let us now take $t_1 < t_2$ and $t_1' < t_2'$ and consider the case $t_1 < t_1' < t_2 < t_2'$. Then

$$\begin{aligned} & \langle [W(t_2) - W(t_1)]\left[W(t_2') - W(t_1')\right]\rangle = \langle W(t_2)W(t_2')\rangle + \langle W(t_1)W(t_1')\rangle \\ - \;& \langle W(t_2)W(t_1')\rangle - \langle W(t_1)W(t_2')\rangle = \sigma^2\left(t_2 + t_1 - t_1' - t_1\right) = \sigma^2(t_2 - t_1') \\ = \;& \sigma^2\left[(t_2 - t_1) \cap (t_2' - t_1')\right] \end{aligned} \tag{4.50}$$

This relation is generally true for other combinations of the intervals.

If we now write

$$\begin{aligned} & t_1 = t, t_2 = t + \mathrm{d}t; t_1' = t', t_2' = t' + \mathrm{d}t' \\ & W(t + \mathrm{d}t) - W(t) = \mathrm{d}W(t) \\ & W(t' + \mathrm{d}t') - W(t') = \mathrm{d}W(t') \end{aligned}$$

we deduce from (4.50)

$$\langle \mathrm{d}W(t)\mathrm{d}W(t')\rangle = \sigma^2\left(\mathrm{d}t \cap \mathrm{d}t'\right) \tag{4.51}$$

and in particular for $t = t'$

$$\langle \mathrm{d}W(t)^2\rangle = \sigma^2 \mathrm{d}t \tag{4.52}$$

This relation holds for the actual paths i.e. $[\mathrm{d}W(t)]^2 = \sigma^2 \mathrm{d}t$ which shows that $W(t)$ is nowhere differentiable.

Let us now choose an arbitrary non-stochastic function $f(t,t')$ continouos in t and t' and consider the integral

$$\begin{aligned} & \int\int f(t,t')\langle \mathrm{d}W(t)\mathrm{d}W(t')\rangle = \sigma^2 \int\int f(t,t') \left(\mathrm{d}t \cap \mathrm{d}t'\right) \\ = \quad & \sigma^2 \int f(t,t)\mathrm{d}t = \sigma^2 \int\int f(t,t')\delta(t-t')\mathrm{d}t\mathrm{d}t' \end{aligned}$$

Therefore we may write this as

$$\langle \frac{\mathrm{d}W(t)}{\mathrm{d}t} \frac{\mathrm{d}W(t')}{\mathrm{d}t'} \rangle = \sigma^2 \delta(t-t') \tag{4.53}$$

and even if $\mathrm{d}W(t)/\mathrm{d}t$ does not exist its variance is a delta-function.