Chapter 2

A Brief Review of Mathematical Background

2.1 Introduction

In the previous chapter, we saw that the deterministic formulation of the hydrodynamic dispersion requires us to make linearizing assumptions, which give rise to the coefficients that are dependent on the scale of the problem. This also makes us think of ways by which we could model the randomness introduced by the system, geometry of porous media, in the case of the hydrodynamic dispersion, without having to resort to any linearizing assumptions. Recent advances in stochastic calculus provide us with useful concepts and tools to model systems in much more natural way. We will attempt to review the pertinent concepts and methods in stochastic calculus prior to discussing stochastic differential equations later in Chapter 4 in some detail. However, to motivate the discussion on stochastic calculus we will first present the behavior of a simple model based on stochastic calculus. These type of examples can be found in Kloden and Platen (1992) and Øksendal (1998). We will outline the differences between the solutions of stochastic differential equations and ordinary differential equations using an example from population dynamics. In the next chapter we will discuss the models by using computer simulations to show the complexity of the solutions even though the solutions themselves appear to be straightforward mathematical expressions.

Let us consider a population growth model where the assumption is made that the rate of change of a population ($X(t)$) is linearly related to the population itself. We can write the equation as,

$$\frac{dX(t)}{dt} = \alpha(t) X(t), \tag{2.1}$$

where $\alpha(t)$ is a time dependent coefficient.

If $\alpha(t)$ is a constant ($\alpha(t) = \alpha_0$) then equation (2.1) has the solution,

$$X(t) = X_0 e^{\alpha_0 t}, \tag{2.2}$$

where X_0 is the initial population (i.e. the population at $t = 0$).

If $\alpha(t)$ is affected by unknown processes of the system to make it "noisy", then it can be expressed as the sum of an average rate constant $r(t)$ and a noise term with an amplitude $\sigma(t)$:

$$\alpha(t) = r(t) + \sigma(t)\, W_t \,. \tag{2.3}$$

The "noise" term W_t will be different at different times, but is not uniquely determined by the t value. That is why, for the time being, we use a subscript with W rather than a function notation, and will expand on that below. We generalize the discussion by considering the case where the rate at which the population changes, depends on the population in a more complicated way than a mere proportionality; arriving at the equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) W_t \,. \tag{2.4}$$

Equation (2.1) together with equation (2.3) represents the special case of equation (2.4) when $b(t,X_t) = r(t)\, X_t$ and $\sigma(t,X_t) = \sigma(t)\, X_t$. In equation (2.4) we have changed from $X(t)$ to X_t ; this is to signify that once $\alpha(t)$ is "noisy", it influences X to be "noisy" as well. In other words, X will no longer have a unique value at a given time and is not a simple function any more.

It is appropriate at this time to introduce the concepts of stochastic variables and processes. An ordinary differential equation, such as equation (2.1), defines a functional relationship between X and t that allows us to calculate a deterministic variable X uniquely for a given time, by a formula such as equation (2.2).

A stochastic variable Y, on the other hand, is one that does not have a unique value; it can have any one out of a collection of values. We assign a unique label ω to each possible value of the stochastic variable, and set Ω to denote the collection of all such values. In some cases, such as when Y represents the outcome of throwing dice, Ω may be a finite set of discrete numbers. In others, such as when Y is the instantaneous position of a particle undergoing

Brownian motion, it may be a continuous range of real numbers. If a particular value y_ω is observed for Y, this is called an event F. In fact, this is only the simplest prototype of an event; other possibilities might be that the value of Y is observed not to be y_ω (the complementary event), or that a value within a certain range of ω values is observed. The set of all possible events is denoted by $\mathcal{F}$. The final ingredient needed for Y to qualify as a stochastic variable, is that even though the outcome of a particular observation of Y is unpredictable, the probability of observing y_ω must be determined by a probability function $P(\omega)$. By using the standard methods of probability calculus, this implies that a probability $P(F)$ can also be assigned to compound events F e.g. by appropriate summation or integration over ω values. For this to work, $\mathcal{F}$ must satisfy the criteria that for any event F in $\mathcal{F}$, its complement F^c must also belong to $\mathcal{F}$, and that for any subset of F's the union of these must also belong to $\mathcal{F}$. The explanation above of what it means to call Y a stochastic variable, is encapsulated in formal mathematical language by saying "Y is defined on a probability space $(\omega, \mathcal{F}, P)$" .

In describing physical systems, deterministic variables usually depend on additional parameters such as time. Similarly, a stochastic variable may depend on an additional parameter t (for example, the probability may change with time, i.e. $P(y_\omega,t)$). The collection of stochastic variables, Y_t , is termed a stochastic process. The word 'process' suggests temporal development and is particularly appropriate when the parameter t has the meaning of time, but mathematically it is equally well used for any other parameter, usually assumed to be a real number in the interval $[0,\infty]$. The label ω is often explicitly included in writing the notation $Y_t(\omega)$, for an individual value obtained from the set of Y-values at a fixed t. Conversely, we might keep ω fixed, and let t vary; a natural notation would be to write $Y_\omega(t)$. In physical terms, one may think of this as the set of values obtained from a single experiment to observe the time development of the stochastic variable Y. When the experiment is repeated, a different set of observations is obtained; those may be labeled by a different value of ω. Each such sequence of observed Y-values is called a realization (or sometimes a path) of the stochastic process, and from this perspective ω may be considered as labeling the realizations of the process. It is seen that it is somewhat arbitrary which of ω and t is considered to be a label, and which an independent variable; this is sometimes expressed by writing the stochastic process as $Y(t,\omega)$. In the light of these concepts and definitions we return to a consideration of the noise term W_t in equation (2.4). One can make a number of plausible assumptions

about the properties ascribed to W_t, which provide us with a framework to model the noise term:

- Considering that the noise represents deviations away from the deterministic rate, the expected value of W_t over all realizations, is zero $(E(W_t) = 0)$, i.e. the mean of the noise term at a given instant is negligible.

- W_t at a particular time does not depend on the noise at any other time, i.e., W_t is an independent stochastic process.

- W_t is stationary, i.e. the joint probability distribution of $\{ W_{t1} , W_{t2} ,\ldots W_{ti} , \ldots\}$ is time invariant. Unless we have a large set of "noise" data we cannot be certain about this assumption. But many applications suggest that this assumption is reasonable. If the system components and/or processes causing noise are invariant, then one could expect this assumption to hold.

- All paths of W_t are continuous. This is needed to ensure that the underlying variable, i.e. $\alpha(t)$ in equation (2.1), remains continuous within any single realization. Although the sources of noise are assumed unknown and have different values in each realization, it is reasonable to assume that they are finite and thus can not cause discontinuous jumps in any system variable.

These assumptions, even though appearing reasonable, cannot be met by any known stochastic process. A stochastic process meeting the first three requirements, cannot have continuous paths (Øksendal, 1998). Intuitively, the problem is that if at every moment the stochastic variable changes independently, any given experiment is bound to show discontinuous fluctuations of the measured value, no matter how small the time steps are made.

The resolution of this dilemma is to assume that not the variable itself, but rather its increments, are independent at different times. This state of affairs is reminiscent of the physical phenomenon of Brownian motion. The botanist Robert Brown first observed that pollen grains suspended in liquid, undergo irregular motion. Centuries later, it was realized that the physical explanation of this is that the pollen grain is continually bombarded by molecules of the liquid traveling with different speeds in different directions. Over a time scale that is large compared with the intervals between molecular impacts, these will average out and no net force is exerted on the grain. However, this will not happen over a small time interval; and if the mass of the grain is small

enough to undergo appreciable displacement in the small time interval as the result of molecular impacts, an observable erratic motion results. The crucial point to notice in the present context is that while the impacts and therefore the individual displacements suffered by the grain can be considered independent at different times, the actual position of the grain can only change continuously. Clearly this phenomenon can serve as a template on which to model the description of the stochastic noise term and its mathematical description should be understood in more detail.

Let us examine the solution to the stochastic differential equation we have discussed before. We take equation (2.4) in the simple case where the population growth rate and noise amplitude are constant, i.e. $r(t) = r$ and $\sigma(t) = \sigma$. The solution to the stochastic differential equation is given by (Øksendal, 1998):

$$X_t = X_0 \exp\left(\sigma W_t\right) \exp\left(\left(r - \frac{\sigma^2}{2}\right)t\right). \tag{2.5}$$

Equation (2.5) gives a realization of the solution and each "experiment" will produce a different path, since W_t will assume a different set of values.

Comparing the stochastic solution with the deterministic solution of equation (2.2), we notice two important differences. The last exponential factor in equation (2.5) shows that the presence of noise in the rate coefficient α in the population growth equation (2.1) modifies the effective growth rate. But moreover, the first exponential factor that will develop differently for each realization, could mask completely the underlying deterministic growth if $\sigma >> r$. This also suggests that apparently random series of data may come from a simple deterministic relationship with the presence of noise in its coefficients. These remarks will be further explored by computer simulation later.

To summarize, the effects of including random effects in differential equations can manifest in two different ways which require two fundamentally different methods of analysis. The first type of equation has random coefficients, a random initial value or is forced by a fairly regular stochastic process, or when some combination of these holds. These equations are called random differential equations and are solved sample path by sample path (Kloeden and Platen, 1992). On the other hand, if the underlying noise is modeled by an irregular stochastic process such as Gaussian white noise, the equations are described as stochastic differential

equations (SDE) which should be solved by employing Ito or Stratonovich stochastic integrals. We will discuss Ito integral later in this chapter and show how it differs from the standard integral, Riemann integral. The sample paths of random differential equations are differentiable but those of SDEs are not differentiable. The choice of an appropriate model to represent noise is a crucial factor in deciding the solution methods. However, most variability in natural systems appears to be irregular whether the sources of noise be the random inputs, irregular boundary conditions or unpredictable distribution of material properties. Therefore flows and variables in most natural systems can be represented as stochastic variables, and deterministic descriptions can be considered as a subset of the pertinent stochastic models.

Before we develop models using stochastic approaches, it is useful to summarily present useful concepts and methods of stochastic calculus. This discussion is brief and intended only to make reader familiar with the ideas and apply them in appropriate situations.

2.2 Elementary Stochastic Calculus

The review presented in the sections that follows this section should not be considered as an exhaustive treatment of the subject; we carefully selected concepts and definitions that are fundamental to understand stochastic calculus and processes without being overly concerned about covering the entire domain of knowledge that could be useful in developing models. Instead we present a starting point to study this area, often perceived as a difficult one in applied mathematics by the scientists and engineers who are oriented towards solving practical problems. There are many excellent books available to reader to study into the subject deeper: Klebaner (1998) gives an excellent introduction to stochastic calculus; Øksendal (1998), Kloeden and Platen (1992), Steele (2001) and Durrett (1996) provide rigorous treatments of the subject for the more mathematically oriented readers. We intend to discuss sufficient number of topics to facilitate the developments and discussions we are going to have later in the book.

2.3 What is Stochastic Calculus?

In standard calculus we deal with differentiable functions which are continuous except perhaps in certain locations of the domain under consideration. To understand the continuity of the functions better we make use of the definitions of the limits. We call a function f, a continuous function at the point $t = t_0$ if

$$\lim_{t \to t_0} f(t) = f(t_0)$$

regardless of the direction t approaches t_0 . A right-continuous function at t_0 has a limiting value only when t approaches t_0 from the right direction, i.e. t is larger than t_0 in the vicinity of t_0. We will denote this as

$$f(t+) = \lim_{t \downarrow t_0} f(t) = f(t_0) .$$

Similarly a left-continuous function at t_0 can be represented as

$$f(t-) = \lim_{t \uparrow t_0} f(t) = f(t_0) .$$

These statements imply that a continuous function in both right-continuous and left-continuous at a given point of t. Often we encounter functions having discontinuities; hence the need for the above definitions. To measure the size of a discontinuity, we define the term "jump" at any point t to be a discontinuity where the both $f(t+)$ and $f(t-)$ exist and the size of the jump be $\Delta f(t) = f(t+) - f(t-)$. The jumps are the discontinuities of the first kind and any other discontinuity is called a discontinuity of the second kind. Obviously a function can only have countable number of jumps in a given range. From the mean value theorem in calculus it can be shown that we can differentiate a function in a given interval only if the function is either continuous or has a discontinuity of the second kind during the interval. Stochastic calculus is the calculus dealing with often non-differentiable functions having jumps without discontinuities of the second kind. One such example of a function is the Wiener process (Brownian motion). One realization of the standard Wiener process is given in Figure 2.1.

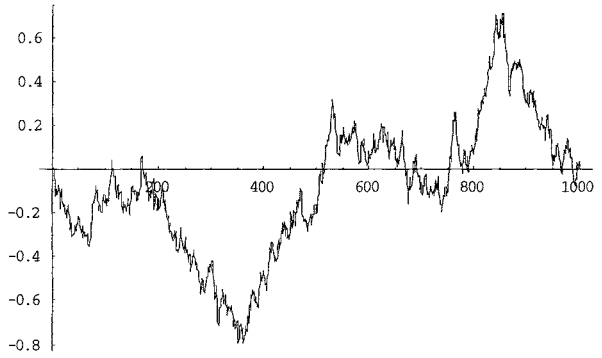

Figure 2.1 An example of a function dealt in stochastic calculus. This function is continuous but not differentiable at any point.

Without going into details of how we computed this function- we will do that in Chapter 3 – we can see that the increments are irregular and we can not define a derivative according to the mean value theorem. This is because of the fact that the function changes erratically within small intervals, however small that interval may be, and we can not define a derivative at a given point in the conventional sense. Therefore we have to devise new mathematical tools that would be useful in dealing with these irregular non-differentiable functions.

2.4 Variation of a Function

Variation of a function f on $[a,b]$ is defined as

$$V_f([a,b]) = \lim_{\delta_n \to 0} \sum_{i=1}^{n} \left| f(t_i^n) - f(t_{i-1}^n) \right|, \tag{2.6}$$

where $\delta_n = \max_{1 \le i \le n} (t_i - t_{i-1})$.

If $V_f([a,b])$ is finite such as in continuous differentiable functions then f is called a function of finite variation on $[a,b]$. Variation of a function is a measure of the total change in the function value within the interval considered. An important result (Theorem 1.7, Klebaner (1998)) is that a

function of finite variation can only have a countable number of jumps. Furthermore, if f is a continuous function, f' exists and $\int |f'(t)|dt < \infty$ then f is a function of finite variation. This implies that a function of finite variation on $[a,b]$ is differentiable on $[a,b]$.

Another quantity that plays a major role in stochastic calculus is the quadratic variation. In stochastic calculus the quadratic variation of a function f over the interval $[0,t]$ is given by

$$[f](t) = \lim \sum_{i=1}^{n} (g(t_i^n) - g(t_{i-1}^n))^2 , \tag{2.7}$$

where the limit is taken over the partitions:

$$0 = t_0^n < t_1^n < \ldots < t_n^n = t ,$$

with $\delta_n = \max_{1 \le i \le n} (t_i^n - t_{i-1}^n) \to 0.$

It can be proved that quadratic variation of a continuous function with finite variation is zero. However, the functions having zero quadratic variation may have infinite variation such as zero energy processes (Klebaner, 1998). If a function or process has a finite positive quadratic variation within an interval, then its variation is infinite. It also means that such functions are continuous but not differentiable.

We do not use quadratic variation in standard calculus which deals with continuous and smooth functions. We also define quadratic covariation of functions f and g on $[0, t]$ by extending equation (2.7):

$$[f, g](t) = \lim \sum_{i=0}^{n-1} (f(t_{i+1}^n) - f(t_i^n))^2 (g(t_{i+1}^n) - g(t_i^n))^2 . \tag{2.8}$$

when the limit is taken over partitions $\{t_i^n\}$ of $[0,t]$ with $\delta_n = \max_{1 \le i \le n} (t_{i+1}^n - t_i^n) \to 0.$ It can be shown that if both the functions are continuous and one is of finite variation, the quadratic covariation is zero.

Covariation of two functions, f and g, has the following properties:

Polarization Identity

Polarization identity expresses the quadratic covariation, $[f, g](t)$, in terms of quadratic variation of individual functions.

$$[f,g](t)=\frac{1}{2}([f+g,f+g](t)-[f,f](t)-[g,g](t)). \tag{2.9}$$

Symmetry

$$[f,g](t) = [g,f](t). \tag{2.10}$$

Linearity

Using polarization identity and symmetry one can show that covariation is linear for any constants a and b,

$$[af+bg,h](t)=a[f,h](t)+b[g,h](t). \tag{2.11}$$

Quadratic variation of a function $[f](t)$ and covariation $[f, g](t)$ are measures of change in the functional values over a given range $[0, t]$.

2.5 Convergence of Stochastic Processes

In many situations where stochastic processes are involved, we would like to know the limiting values of useful random variables, i.e. whether they approach some sort of a "steady state" or asymptotic behavior. However the steady state of random variables have to be defined within a probabilistic context. Therefore, in stochastic processes we discuss the convergence of random variables using four different criteria.

Almost Sure Convergence

Random variables $\{X_n\}$ converges to $\{X\}$ with probability one:

$$P(\{\omega \in \Omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\}) = 1. \tag{2.12}$$

Mean-square Convergence

$\{X_n\}$ converges to $\{X\}$ such a way that $E(X_n^2) < \infty$ for $n = 1,2,...,n$, $E(X) < \infty$ and

$$\lim_{n\to\infty} E(|X_n - X|^2) = 0. \tag{2.13}$$

Convergence in Probability

$\{X_n\}$ converges to $\{X\}$ with zero probability of having a difference between the two processes:

$$\lim_{n\to\infty} P(\{\omega \in \Omega; |X_n(\omega) - X(\omega)| \geq \varepsilon) = 0, \tag{2.14}$$

for all $\varepsilon > 0$.

Convergence in probability is called stochastic convergence as well.

Convergence in Distribution

Distribution function of $\{X_n\}$ converges to that of $\{X\}$ at any point of continuity of the limiting distribution (i.e. the distribution function of $\{X\}$).

These four criteria add another dimension to our discussion of the asymptotic behavior of a process. These arguments can be extended to the comparison of stochastic processes with each other. For example, in many instances one stochastic process is said to be equal to another in convergence in probability. That means for the two stochastic processes $\{X_1\}$ and $\{X_2\}$,

$$\lim_{n\to\infty} P(\{\omega\in\Omega;\left|X_1(\omega)-X_2(\omega)\right|\geq\varepsilon)=0 \tag{2.15}$$

for all $\varepsilon>0$.

Unlike in deterministic variables where any asymptotic behavior can clearly be identified either graphically or numerically, stochastic variables do require adherence to one of the convergence criteria mentioned above which are called the "criteria for strong convergence". There are weakly converging stochastic processes and we do not discuss the weak convergence criteria as they are not relevant to the development of the material in this book.

2.6 Riemann and Stieltjes Integrals

In standard calculus we have continuous functions with discontinuities at finitely many points and we integrate them using the definition of Riemann integral of a function $f(t)$ over the interval $[a,b]$:

$$\int_a^b f(t)\,dt = \lim_{\delta\to 0}\sum_{i=1}^{n} f(\xi_i^n)\,(t_i^n - t_{i-1}^n), \tag{2.16}$$

where t_i^n's represents partitions of the interval,

$$a=t_0^n<t_1^n<t_2^n....<t_n^n=b,$$

$$\delta=\max_{1\leq i\leq n}(t_i^n-t_{i-1}^n),\ \text{and}\quad t_{i-1}^n\leq\xi_i^n\leq t_i^n.$$

Riemann integral is used extensively in standard calculus where continuous functions are the main concern. The integral converges regardless of the chosen ξ_i^n within $[t_{i-1}^n, t_i^n]$.

A generalization of Riemann integral is Stieltjes integral which is defined as the integral of $f(t)$ with respect to a monotone function $g(t)$ over the interval $[a,b]$:

$$\int_a^b f(t)\,dg(t) = \lim_{\delta \to 0} \sum_{i=1}^{n} f(\xi_i)(g(t_i^n) - g(t_{i-1}^n)), \tag{2.17}$$

with the same definitions as above for δ and t_i^n's. It can be shown that for the integral to exist for any continuous function $f(t)$, $g(t)$ must be a function with finite variation on $[a,b]$. This means that if $g(t)$ has infinite variation on $[a,b]$ then for such a function, integration has to be defined differently. This is the case in stochastic integration based on Brownian motion which has infinite variation in any given interval however small the interval is, and therefore, can not be integrated using Stieltjes integral. Before we discuss alternative forms of integration that can be applied to the functions of positive quadratic variation, i.e. the functions of infinite variation, we will present important properties and some useful results associated with Brownian motion or the Wiener process.

2.7 Brownian Motion and Wiener Processes

In the physical Brownian motion, there are small but nevertheless finite intervals between the impulses of molecules colliding with the pollen grain. Consequently, the path that the grain follows, consists of a sequence of straight segments forming an irregular but continuous line – a so-called random walk. Each straight segment can be considered an increment of the momentary position of the grain.

The mathematical idealization of this, similar to that which describes a circle as the limiting case of a polygon when the number of sides approaches infinity, is to let the interval between increments approach zero. The resulting process – called Brownian motion or a Wiener process – is difficult to conceptualize: for example, consideration shows that the resulting position is everywhere continuous, but nowhere differentiable. This means that while the particle has a position at any moment, and since this position is changing - it

is moving - yet no velocity can be defined. Nevertheless as discussed by Stroock and Varadhan (1979) a consistent mathematical description is obtained by defining the position as a stochastic process $B(t,\omega)$ with the following properties that are suggested as a model for a pure noise process:

P1: $B(0,\omega) = 0$, i.e. choose the position of the particle at the arbitrarily chosen initial time $t = 0$ as the coordinate origin;

P2: $B(t,\omega)$ has independent increments, i.e. $B(t_1,\omega)$, $\{B(t_2,\omega)- B(t_1,\omega)\},\ldots,$ $\{B(t_k,\omega)- B(t_{k-1},\omega)\}$ are independent for all $0 \leq t_1 < t_2 \ldots < t_k$;

P3: $\{B(t_{i+1},\omega)-B(t_i,\omega)\}$ is normally distributed with mean 0 and variance $(t_{i+1} - t_i)$;

P4: The stochastic variation of $B(t,\omega)$ at fixed time t is determined by a Gaussian probability;

P5: The Gaussian has a zero mean, $E\,[B(\,t,\omega)\,] = 0$ for all values of t;

P6: $B\,(t,\omega)$ are continuous functions of t for $t \geq 0$;

P7: The covariance of Brownian motion is determined by a correlation between the values of $B(t,\omega)$ at times t_i and t_j (for fixed ω), given by

$$E[\,B(t_i,\omega)\,B(t_j,\omega)\,] = \min(t_i,t_j)\,. \tag{2.18}$$

When applied to $t_i = t_j = t$, P7 reduces to the statement that

$$Var[B\,(t,\omega)] = t, \tag{2.19}$$

where '*Var*' means statistical variance. For Brownian motion this can be interpreted as the statement that the radius within which the particle can be found increases proportional to time. This is a plausible behavior for a random walk phenomenon, and is of such fundamental importance in what follows it is explored in more detail.

Consider a particle restricted to one-dimensional motion along the x-axis, starting from an initial position $x=x_0$. It is acted upon by independent impacts (e.g. from gaseous molecules impinging on it) at an average rate of f impacts per unit time. Its displacement $b(\tau)$ after a time τ, is given by

$$b(\tau) = x_0 + \sum_{i=1}^{N} x_i \,, \tag{2.20}$$

where $x_i = v_i \, \Delta t_i$ is the distance traveled in interval i as a result of the velocity v_i it acquires in the i-th impact occurring at a discrete time t_i. In terms of the previous terminology, the x_i s are the increments of the position. The total number of impacts N is obviously given by $N = f \, \tau$. The quantities v_i and Δt_i have probability distributions which will be determined by the physics of the situation, but are not further specified except for the assumption that the average value of v_i and consequently also of x_i, are zero. Considering each x_i to be an independent stochastic variable, the probability distribution of $b(\tau)$ is determined by the so-called Central Limit Theorem (CLT) of elementary statistics (Kenney 1966, or any standard statistics textbook). According to the CLT, the distribution of a sum of stochastic variables approaches a normal (i.e. Gaussian) distribution, with its mean and variance equal to the sum of means and variances of the individual variables, as the number of terms approaches infinity. This applies for any non-pathological distribution of the individual variables. Hence $b(\tau)$ has a Gaussian probability distribution with zero mean, and its variance is N times that of an individual position increment x_i.

For a fixed average impact frequency, this means that $Var(b) \propto \tau$ as long as $\tau >> 1/f$ so that $N >> 1$. On the other hand, suppose we keep τ fixed and let f increase without changing the distribution of the impact velocities. For example, in the actual experiment the density of the gas may be increased without changing the temperature. Then, although N increases proportional to f, the value and therefore also the variance of each x_i decreases in the same ratio because the Δt_i decreases proportional to $1/f$. Therefore it is reasonable to assume $Var(b)$ is independent of f and we can take it as proportional to τ even in the limit as $f \to \infty$, in which case the discrete step Brownian motion becomes a Wiener process.

In this way the set of Wiener process properties stipulated above are seen to arise naturally from consideration of a random walk. In particular, the assumption of a Gaussian distribution for B is relatively independent of the detailed statistical properties of the increments.

Note that in the Brownian motion example, τ is multiplied by a proportionality constant containing the average impact frequency and the variance of individual increments, but in the Wiener process the time

constant is one. To achieve that in the Brownian motion, either the position variable or the time needs to be appropriately rescaled. In adopting the standard Wiener process definition, this scaling has been hidden from view. As often done in mathematical discussions, all variables are essentially assumed to be dimensionless. This convention needs to be remembered when applying the theory to a physical situation.

A consistent way to do this is to transform all physical variables occurring in the applicable deterministic differential equations to dimensionless ratios, by dividing them by appropriate scale constants, before introducing the stochastic terms to the equation. In choosing scales one should recognize that the Wiener definition itself introduces the new scale constant explained above. In our Brownian motion example, the rate at which the particle wanders away from its starting position will clearly depend on the magnitude of the velocity imparted to it in individual impacts, i.e. on the mass of the particle and the temperature of the gas in which it is immersed. This demonstrates that physical stochastic processes can take place on different time scales, and an appropriate one should be used to reduce a particular problem to the universal time scale assumed for a Wiener process.

In the previous discussion, for the sake of clarity a distinction was made between Brownian motion where there are random increments at discrete time steps, and the Wiener process which is the limit in which the intervals between increments approach zero. Many authors do not make this distinction and use the terms Brownian motion and Wiener process interchangeably for the mathematical idealization. We will also use the terms Brownian motion and Wiener process interchangeably and by doing so we refer to the same stochastic process.

Because the Wiener process is defined by the independence of its increments, it is for some purposes convenient to reformulate the variance stipulation of a Wiener process in terms of the variance of the increments:

From P3, for $t_i < t_j$:

$$\mathrm{var}[B(t_j,\omega) - B(t_i,\omega)] = t_j - t_i \ . \tag{2.21}$$

Bearing in mind that the statistical definition of the variance of a quantity X reduces to the expectation value expression $Var(X) = E(X^2)\text{-}E^2(X)$ and that the expectation value or mean of a Wiener process is zero, we can rewrite this as

$$E[\{B(t_2,\omega)-B(t_1,\omega)\}^2]=\text{var}[B(t_2,\omega)-B(t_1,\omega)]$$

$$\text{i.e.} \quad E[\Delta B\cdot\Delta B]=\Delta t \tag{2.22}$$

where Δ is defined to mean the time increment for a fixed realization ω. The connection between the two formulations is established by similarly rewriting equation (2.21) and then applying equation (2.18):

$$\begin{aligned} Var[B(t_1,\omega)-B(t_2,\omega)] &= E[\{B(t_1,\omega)-B(t_2,\omega)\}^2] \\ &= E[B^2(t_1,\omega)+B^2(t_1,\omega)-2B(t_1,\omega)B(t_2,\omega)] \\ &= t_1+t_2-2\min(t_1,t_2) \\ &= t_1-t_2 \qquad \text{for } t_1>t_2\,. \end{aligned}$$

2.8 Relationship between White Noise and Brownian Motion

Consider a stochastic process $X(t,\omega)$ having a stationary joint probability distribution and $E(X(t,\omega))=0$, i.e. the mean value of the process is zero. The Fourier transform of $Var(X(t,\omega))$ can be written as,

$$S(\lambda,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} Var(X(\tau,\omega)\,e^{-i\lambda\tau}d\tau \tag{2.23}$$

$S(\lambda,\omega)$ is called the spectral density of the process $X(t,\omega)$ and is also a function of angular frequency λ. The inverse of the Fourier transform is given by

$$Var(X(\tau,\omega))=\int_{-\infty}^{\infty} S(\lambda,\omega)\,e^{i\lambda\tau}d\lambda \tag{2.24}$$

and when $\tau = 0$,

$$Var(X(0,\omega))=\int_{-\infty}^{\infty} S(\lambda,\omega)d\lambda\,. \tag{2.25}$$

Therefore, variance of $X(0,\omega)$ is the area under a graph of spectral density $S(\lambda,\omega)$ against λ:

$$Var(X(0,\omega)) = E(X^2(0,\omega)), \tag{2.26}$$

because $E(X(t,\omega))=0$.

Spectral density $S(\lambda,\omega)$ is considered as the "average power" per unit frequency at λ which gives rise to the variance of $X(t,\omega)$ at $\tau = o$. If the average power is a constant which means that the power is distributed uniformly across the frequency spectrum, such as the case for white light, then $X(t,\omega)$ is called white noise. White noise is often used to model independent random disturbances in engineering systems, and the increments of Brownian motion have the same characteristics as white noise. Therefore white noise $(\zeta(t))$ is defined as

$$\zeta(t) = \frac{dB(t)}{dt},$$

$$dB(t) = \zeta(t)dt. \tag{2.27}$$

We will use this relationship to formulate stochastic differential equations.

2.9 Relationships Among Properties of Brownian Motion

As shown before, the relationships among the properties mentioned above can be derived starting from P1 to P7. For example, let us evaluate the covariance of Brownian motions of $B(t_i,\omega)$ and $B(t_j,\omega)$:

$$Cov(B(t_i,\omega)B(t_j,\omega)) = E(B(t_i,\omega)\,B(t_j,\omega)). \tag{2.28}$$

Assuming $t_i < t_j$ we can express

$$B(t_j,\omega) = B(t_i,\omega) + B(t_j,\omega) - B(t_i,\omega). \tag{2.29}$$

Therefore,

$$E(B(t_i,\omega)\,B(t_j,\omega)) = E(B(t_i,\omega)(B(t_i,\omega) + B(t_j,\omega) - B(t_i,\omega)),$$

$$= E(B^2(t_i,\omega) + B(t_i,\omega)B(t_j,\omega) - B^2(t_i,\omega)),$$

$$= E(B^2(t_i,\omega) + B(t_i,\omega)(B(t_j,\omega) - B(t_i,\omega)))\,,$$

$$= E(B^2(t_i,\omega)) + E(B(t_i,\omega)(B(t_j,\omega) - B(t_i,\omega)))\,. \tag{2.30}$$

From P2, $B(t_i,\omega)$ and $(B(t_j,\omega) - B(t_i,\omega))$ are independent processes and therefore we can write

$$E(B(t_i,\omega)(B(t_j,\omega) - B(t_i,\omega))) = E(B(t_i,\omega))E(B(t_j,\omega) - B(t_i,\omega))\,. \tag{2.31}$$

According to P3 and P5,

$E(B(t_i,\omega)) = 0$ and

$E(B(t_j,\omega) - B(t_i,\omega) = 0\,.$

Therefore, from equation (2.31)

$E(B(t_i,\omega)B(t_j,\omega) - B(t_i,\omega))) = 0\,.$

This leads equation (2.30) to

$E(B(t_i,\omega)B(t_j,\omega)) = E(B^2(t_i,\omega))$, and

$$E(B^2(t_i,\omega)) = E((B(t_i,\omega)) - 0)^2)\,. \tag{2.32}$$

From P3, $\{\, B(t_i,\omega) - B(0,\omega)\,\}$ is normally distributed with a variance $(t_i - 0)$, and equation (2.32) becomes,

$$E(B^2(t_i,\omega)) = t_i \tag{2.33}$$

and, therefore,

$$Cov(B(t_i,\omega)B(t_j,\omega)) = t_i\,. \tag{2.34}$$

Using a similar approach it can be shown that if $t_i > t_j$,

$$Cov(B(t_i,\omega)B(t_j,\omega)) = t_j\,. \tag{2.35}$$

This leads to P7:

$$E(B(t_i,\omega)B(t_j,\omega)) = \min(t_i,t_j)\,. \tag{2.36}$$

The above derivations show the relatedness of the variance of an independent increment, $Var\{B(t_1,\omega)-B(t_2,\omega)\}$ to the properties of Brownian motion given by P1 to P7. The fact that $\{B(t_{i+1},\omega)-B(t_i,\omega)\}$ is a Gaussian random variable with zero mean and $\{t_{i+1}-t_i\}$ variance can be used to construct Brownian motion paths on computer. If we decide the time interval $[0,t]$ into n equidistant parts having length Δt, and at the end of each segment we can randomly generate a Brownian increment using the Normal distribution with mean 0 and variance Δt. This increment is simply added to the value of Brownian motion at the point considered and move on to the next point. When we repeat this procedure starting, from $t=\Delta t$ to $t=t$ and taking the fact that $B(0,\omega)=0$ into account, we can generate a realization of Brownian motion. We can expect these Brownian motion realizations to have properties quite distinct from other continuous functions of t. We will briefly discuss some important characteristics of Brownian motion realizations next as these results enable us to utilize this very useful stochastic process effectively.

2.10 Further Characteristics of Brownian Motion Realizations

1. *$B(t,\omega)$ is a continuous, nondifferentiable function of t.*

2. *The quadratic variation of $B(t,\omega)$, $[B(t,\omega),B(t,\omega)](t)$ over $[0,t]$ is t.*

Using the definition of covariation of functions,

$$\begin{aligned}[B(t,\omega),B(t,\omega)](t) &= [B(t,\omega),B(t,\omega)]([0,t]) \\ &= \lim_{\delta_n\to 0}\sum_{i=1}^{n}[B(t_i^n)-B(t_{i-1}^n)]^2\end{aligned} \tag{2.37}$$

where $\delta_n = \max(t_{i+1}^n - t_i^n)$ and $\{t_i^n\}_{i=1}^n$ is a partition of $[0\,,t]$, as $n\to\infty, \delta_n\to 0$.

Taking the expectation of the summation,

$$E(\sum(B(t_i^n)-B(t_{i-1}^n))^2) = \sum(E((B(t_i^n)-B(t_{i-1}^n))^2)) \tag{2.38}$$

$E((B(t_i^n)-B(t_{i-1}^n))^2)$ is the variance of an independent increment $\{B(t_i^n)-B(t_{i-1}^n)\}$.

As seen before,

$$Var[B(t_i^n)-B(t_{i-1}^n)]=(t_i^n-t_{i-1}^n). \tag{2.39}$$

Therefore,

$$\begin{aligned} E\Big(\sum(B(t_i^n)-B(t_{i-1}^n))^2\Big) &= \sum Var[B(t_i^n)-B(t_{i-1}^n)], \\ &= \sum_{i=1}^{n}(t_i^n-t_{i-1}^n)=t-0=t. \end{aligned} \tag{2.40}$$

Let us take the variance of $\sum(B(t_i^n)-B(t_{i-1}^n))^2$:

$$Var\Big(\sum(B(t_i^n)-B(t_{i-1}^n))^2\Big)=\sum 3(t_i^n-t_{i-1}^n)^2 \le 3\ \max\ (t_i^n-t_{i-1}^n)\ t=3t\delta_n. \tag{2.41}$$

As $n\to\infty, \delta_n\to 0, \sum Var(B(t_i^n)-B(t_{i-1}^n))^2\to 0$.

Summarizing the results,

$$E(\sum(B(t_i^n)-B(t_{i-1}^n))^2)=t$$

and

$$Var(\sum(B(t_i^n)-B(t_{i-1}^n))^2)\to 0 \text{ as } n\to\infty.$$

This implies that, according to the monotone convergence theories that $\sum(B(t_i^n)-B(t_{i-1}^n))^2\to t$ almost surely as $n\to\infty$.

Therefore, the quadratic variation of Brownian motion $B(t,\omega)$ is t:

$$[B(t,\omega),B(t,\omega)](t)=t. \tag{2.42}$$

Omitting t and ω, $[B,B](t)=t$.

3. Brownian Motion $(B(t,\omega))$ is a Martingale.

A stochastic process, $\{X(t)\}$ is a martingale, when the future expected value of $\{X(t)\}$ is equal to $\{X(t)\}$. In mathematical the notation, $E(X(t+s)|F_t) = X(t)$ with converging almost surely, and F_t is the information about $\{X(t)\}$ up to time t. We do not give the proof of these martingale characteristics of Brownian motion here but it is easy to show that $E(B(t+s)|F_t) = B(t)$.

It can also be shown that $\{B(t,\omega)^2 - t\}$ and $\{\exp(\alpha B(t,\omega) - \frac{\alpha^2}{2}t)\}$ are also martingales. These martingales can be used to characterize Brownian motion as well and more details can be found in Klebaner (1998).

4. Brownian motion has Markov property.

Markov property simply states that the future of a process depends only on the present state. In other words, a stochastic process having Markov property does not "remember" the past and the present state contains all the information required to drive the process into the future states.

This can be expressed as

$$P(X(t+s) \le y \mid F_t) = P(X(t+s) \le y \mid X(t)), \tag{2.43}$$

almost surely .

From the very definition of increments of the Wiener process (Brownian motion), for the discretized intervals of [0,t], $\{B(t_{i+1}^n) - B(t_i^n)\}$, the Brownian motion increment behaves independently to its immediate processor $\{B(t_i^n) - B(t_{i-1}^n)\}$. In other words $\{B(t_{i+1}^n) - B(t_i^n)\}$ does not remember the behavior of $\{B(t_{i-1}^n) - B(t_{i-1}^n)\}$ and only element common to both increments is $B(t_i^n)$. One can now see intuitively why Brownian motion should behave as a Markov process. This can be expressed as

$$P(B(t_i+s) \le x_i \mid \{B(t_i), B(t_{i-1})...0)\}) = P(B(t_i+s) \le x_i \mid B(t_i)), \tag{2.44}$$

which is another way of expressing the previous equation (2.43).

2.11 Generalized Brownian Motion

The Wiener process as defined above is sometimes called the standard Wiener process, to distinguish it from that obtained by the following generalized equation (2.45):

$$E[\, B(t_i,\omega)\, B(t_j,\omega)\,] = \int_0^{\min(t_i,t_j)} q(\tau)\, d\tau \quad . \tag{2.45}$$

The integral kernel $q(\tau)$ is called the correlation function and determines the correlation between stochastic process values at different times. The standard Wiener process is the simple case that $q(\tau) \equiv 1$, i.e. full correlation over any time interval; the generalized Wiener process includes, for example, the case that q decreases, and there is progressively less correlation between stochastic values in a given realization as the time interval between them increases.

2.12 Ito Integral

At this point of our discussion, we need to define the integration of stochastic process with respect to the Wiener process $(B(t,\omega))$ so that we understand the conditions under which this integral exists and what kind of processes can be integrated using this integral. As we restrict the definition to Ito integration we denote the integral as

$$I[X](\omega) = \int_S^T X(t,\omega)\, dB(t,\omega)\,. \tag{2.46}$$

$I[X](\omega)$ implies that the integration of $X[t,\omega]$ is along a realization ω and with respect to Brownian motion which is a function of t. $I[X](\omega)$ is also a stochastic process in its own right and have properties stemming out of the definition of the integral. It is natural to expect $I[X](\omega)$ to be equal to $c(B(t,\omega) - B(s,\omega))$ when $X(t,\omega)$ is a constant c. If $X(\ t\ )$ is a deterministic process, which can be expressed as a sequence of constants over small intervals, we can define Ito integral as follows:

$$\begin{aligned} I[X] &= \int_S^T X(t)\, dB(t) \\ &= \sum_{i=0}^{n-1} c_i((B(t_{i+1}) - B(t_i))) \end{aligned} \tag{2.47}$$

where $X(t) = \begin{cases} c_0, & t=S \\ c_i, & t_i < t \le t_{i+1} \end{cases}$ $i = 0, \cdots, n-1$.

The time interval $[S,T]$ has been discretized into n intervals: $S = t_0 < t_1 < \cdots < t_n = T$.

Using the property of independent increments of Brownian motion, we can show that the mean of $I[X](\omega)$ is zero and,

$$Variance = Var(I[X]) = \sum_{i=0}^{n-1} c_i^2 (t_{i+1} - t_i) . \tag{2.48}$$

It turns out that if $X(t,\omega)$ is a continuous stochastic process and its future values are solely dependant on the information of this process only up to t, Ito integral $I[X](\omega)$ exists. The future states of a stochastic process, $X(t,\omega)$, is only dependent on F_t then it is called an adapted process. A left-continuous adapted process $X(t,\omega)$ is defined as a predictable process and it satisfies the following condition: $\int_0^T X^2(t,\omega)\, dt < \infty$ almost surely.

As we are only concerned about continuous processes driven by the past events, we limit our discussion of predictable processes to the subclass of left-continuous and adapted processes. Reader may want to refer to Øksendal (1998) and Klebaner (1998) for more rigorous discussion of these concepts.

We can now define Ito integral $I[X](\omega)$ similarly to equation (2.47). If $X(t,\omega)$ is a continuous and adapted process then $I[X](\omega)$ can be defined as

$$\sum_{i=0}^{n-1} X(t_i^n,\omega)(B(t_{i+1}^n,\omega) - B(t_i^n,\omega)), \tag{2.49}$$

and this sum converges in probability.

Dropping ω for convenience and adhering to the same discretization of interval $[S, T]$ as in equation (2.47),

$$I[X] = \int_S^T X(t)dB(t) = \sum_{i=0}^{n-1} X(t_i^n)(B(t_{i+1}^n) - B(t_i^n)) . \tag{2.50}$$

Equation (2.50) expresses an approximation of $\int_S^T X(t)\,dB(t)$ based on probabilistic convergence. We take equation (2.49) as the definition of Ito integral for the purpose of this book. As stated earlier $I[X](\omega)$ is a stochastic process and it has the following properties (see, for example, Øksendal (1998) for more details):

1. Ito integral is linear.

If $X(t)$ and $Y(t)$ are predictable processes and α and β as some constants, then

$$I[\alpha X + \beta Y](\omega) = \alpha\, I[X](\omega) + \beta\, I[Y](\omega). \tag{2.51}$$

2. Ito integral has zero-mean property.

$$E(I[X](\omega)) = 0. \tag{2.52}$$

3. Ito integral is isometric.

$$E[(\int_S^T X(t)\,dB(t))^2] = \int_S^T E(X^2(t))\,dt. \tag{2.53}$$

The isometry property says that the expected value of the square of Ito integral is the integral with respect to t of the expectation of the square of the process $X(t)$.

Since $E[(\int_S^T X(t)\,dB(t))^2] = 0$ from zero mean property, we can express the left hand side of equation (2.53) as

$$\begin{aligned} &E((\int_S^T X(t)dB(t))^2 - E(\int_S^T X(t)dB(t))) \\ &= E[\int_S^T X(t)dB(t) - E(\int_S^T X(t)dB(t))]^2 = Var(\int_S^T X(t)dB(t)). \end{aligned} \tag{2.54}$$

Therefore the variance of Ito integral process is $\int_s^T E(X^2(t))dt$ and this result will be useful to us in understanding the behavior of Ito integral process. We say that Ito integral is square integrable. According to Fubuni's Theorem, which states that, for a stochastic process $X(t)$, with continuous realizations,

$$E(\int_s^T X(t)dt) = \int_s^T E(X(t))dt, \tag{2.55}$$

$$\int_s^T E(X^2(t))dt = E(\int_s^T (X^2(t)dt). \tag{2.56}$$

4. Ito integral is a martingale.

It can be shown that $E(I[X(t)] \mid F_t) = I[X(t)]$. Strictly speaking $X(t)$ should satisfy $\int_S^T X^2(t)dt < \infty$ and $\int_S^T E(X^2(t))dt < \infty$ for martingale property to be true. Therefore Ito integrals are square integrable martingales.

5. Ito integral of a deterministic function X (t) *is a Guassian process with zero mean and covariance function.*

$$Cov(I[X(t)], I[X(t+t_0)]) = \int_S^t X^2(s)ds, \; t_0 \geq 0. \tag{2.57}$$

$I[X(t)]$ is a square integrable martingale.

6. Quadratic variation of Ito integral is given by

$$[I[X], P[X]](t) = \int_S^T X^2(t)dt. \tag{2.58}$$

We see that Ito integral has a positive quadratic variation making it a process with infinite variation i.e. it is a nondifferentiable continuous function of t.

7. Quadratic covariation of Ito integral with respect to processes $X_1(t)$ and $X_2(t)$ is given by

$$[I[X_1], I[X_2]](f) = \int_S^T X_1(t)X_2(t)dt. \tag{2.59}$$

Armed with these properties we can proceed to discuss mechanics of stochastic calculus such as stochastic chain rule, which is also known as Ito formula.

2.13 Stochastic Chain Rule (Ito Formula)

2.13.1 Differential notation

As we have seen previously, quadratic variations of Brownian motion, $[B(t,\omega), B(t,\omega)](t)$, is the limit in probability over the interval $[0, t]$:

$$[B(t,\omega),B(t,\omega)](t) = \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} (B(t_{i+1}^n) - B(t_i^n))^2, \tag{2.60}$$

$\delta_n = \max(t_{i+1}^n - t_i^n) \to 0$. Using the differential notation,

$$\Delta B = B(t_{i+1}^n) - B(t_i^n),$$

and summation as an integral,

$$[B(t,\omega),B(t,\omega)](t) = \int_s^t (dB(s))^2 . \tag{2.61}$$

We have shown that $[B,B](t) = t$, and therefore, $\int_s^t (dB(s))^2 = t$.

For our convenience and also to make the notation similar to the one in standard differential calculus, we denote

$$\int_o^t (dB(s))^2 = t \tag{2.62}$$

as $(dB(t))^2 = dt$. (2.63)

This equation does not have a meaning outside the integral equation (2.62) and should not be interpreted in any other way.

Similarly for any other continuous function $g(t)$,

$$g(t)(dB(t))^2 = g(B(t))dt\,, \tag{2.64}$$

which means,

$$\int_o^t g(t)dB(s))^2 = \int_o^t g(B(s))ds. \tag{2.65}$$

This equation is an expression of the approximation, in probability, of

$$\lim_{\delta_n \to 0} \sum_{i=0}^{n-1} g(t_i^n)(B(t_{i+1}^n) - B(t_i^n))^2 = \int_o^t g(B(s))ds\,. \tag{2.66}$$

As quadratic variation of a continuous and differentiable function is zero,

$$[t,t]\,(t) = 0\,. \tag{2.67}$$

This equation in integral notation,

$$\int_0^t (dt)^2 = 0\ ,$$

and in differential notation,

$$(dt)^2 = 0\ . \tag{2.68}$$

Similarly, quadratic covariation of t (a continuous and differentiable function) and Brownian notion,

$$[t,B](t) = 0\,. \tag{2.69}$$

This relationship can be proved by expressing quadratic covariation as

$$[t,B](t) = \lim_{\delta_n \to o} \sum_{i=0}^{n-1} (t_{i+1}^n - t_i^n)(B(t_{i+1}^n) - B(t_i^n)), \tag{2.70}$$

$$\delta_n = \max(t_{i+1}^n - t_i^n),$$

$$\begin{aligned} [t,B](t) &\le \delta_n \sum_{i=0}^{n-1} (B(t_{i+1}^n) - B(t_i^n)), \\ &\le \delta_n B(t). \end{aligned} \tag{2.71}$$

Therefore as $n \to \infty, \delta_n \to 0$ (because t is a function of finite variation),

$[t, B](t) \to 0 \quad \text{as} \quad n \to \infty.$

Hence, $[t, B](t) = 0$ and in integral notation,

$$\int_0^t dt\, dB = 0 \ . \tag{2.72}$$

This can be written in differential notation,

$$dt.dB = 0. \tag{2.73}$$

Therefore, we can state the following rules in differential notation,

$$dt.dt = 0;\ dt.dB = 0;\ dB.dt = 0, \text{ and } dB.dB = dt. \tag{2.74}$$

2.13.2 Stochastic Chain Rule

In order to come to grips with the interpretation of the differential properties of dB_t, it is useful to consider the chain rule of differentiation. This will also lead us to formulas that are often more useful to apply in calculating Ito integrals than the basic definition as the limit of a sum. Consider first the case in ordinary calculus of a function g (x,t), where x is also a function of t. We can write the change in g as t changes, as follows:

$$\Delta g = \frac{\partial g(t,x)}{\partial t}\Delta t + \frac{\partial g(t,x)}{\partial x}\Delta x + \tfrac{1}{2}\frac{\partial^2 g(t,x)}{\partial x^2}(\Delta x)^2 + \ldots$$

From this, an expression for dg/dt is obtained by taking the limit $\Delta t \to 0$ of the ratio $(\Delta g/\Delta t)$.

Since $\Delta x = (dx/dt)\, \Delta t$, when $\Delta t \to 0$ the 2nd derivative term shown is of order $(\Delta t)^2$ and falls away together with all higher derivatives, and the well-known chain rule formula for the total derivative (dg/dt) is obtained. However, if instead of x we have a Wiener process B_t, we get

$$\Delta g = \frac{\partial g(t,B_t)}{\partial t}\Delta t + \frac{\partial g(t,B_t)}{\partial x}\Delta B_t + \tfrac{1}{2}\frac{\partial^2 g(t,B_t)}{\partial x^2}(\Delta B_t \cdot \Delta B_t) + \ldots$$

If the expectation value of this expression over all realizations is taken, the above shows that the second derivative term is now only of order Δt and cannot be ignored. Since this holds for the expectation value, for consistency we also cannot neglect the term if the limit $\Delta t \rightarrow 0$ is taken without considering the expectation value. Unlike the case of ordinary calculus where all expressions containing products of differentials higher than 1 is neglected, in Ito calculus we therefore have the rules given by equation (2.74).

Recall that in standard calculus chain rule is applied to composite functions. For example, if $Y=f(t)$ then $g(Y)$ is a function of Y.

Then

$$\frac{dg}{dt} = \frac{dg}{dY}.\frac{dY}{dt}.$$

In differential notation,

$$dg = \frac{dg}{dY}.df\ .$$

By integrating

$$g(f(t)) = g(0) + \int_0^t g'(f(t))df\ . \tag{2.75}$$

Suppose say $f(t) = B(t)$ (Brownian motion) and $g(x)$ is twice continuously differentiable continuous function. Then by using stochastic Taylor series expansion,

$$g(B(t)) = g(0) + \int_0^t g'(B(s))dB(s) + \frac{1}{2}\int_0^t g''(B(s))ds\ . \tag{2.76}$$

Comparing equation (2.76) and the corresponding stochastic chain rule, we can see that the second derivative term of the Taylor series plays a significant role in changing the chain rule in the standard calculus to the stochastic one. For example, let

$$g(x) = e^x\ .$$

Therefore,

$$e^{B(t)} = e^{(0)} + \int_0^t e^{B(s)} dB(s) + \frac{1}{2}\int_0^t e^{B(s)} ds \ . \tag{2.77}$$

In differential notation (which is only a convention)

$$d(e^{B(t)}) = e^{B(t)} dB(t) + \frac{1}{2} e^{B(t)} dt \ . \tag{2.78}$$

As another example, let $g(x) = x^2$.

Therefore, from the chain rule

$$(B(t))^2 = (B(0))^2 + 2\int_s^t B(s) dB(s) + \frac{1}{2}\int_0^t 2ds \,, \tag{2.79}$$

$$\int_0^t B(s) dB(s) = \frac{1}{2}(B(t))^2 - \frac{1}{2}t \,. \tag{2.80}$$

This is quite a different result from the standard integration. In differential convention,

$$B(t)\, dB(t) = \frac{1}{2} d((B(t))^2) - \frac{1}{2} dt \,. \tag{2.81}$$

In other words, the stochastic process $\int_0^t B(s) dB(s)$ can be calculated by evaluating $\{\frac{1}{2}(B(t))^2 - \frac{1}{2}t\}$. We will show how this process behaves using computer simulations in Chapter 3.

2.13.4 Ito Processes

We extend Ito integral by defining on stochastic process, which is similar form to stochastic differential equation (2.4) when written in differential from. We write Ito integral as

$$Y(t) = \int_0^t \sigma(s)dB(s). \tag{2.82}$$

The we can add a "drift term" to the "diffusion term" given by equation (2.82):

$$Y(t) = Y(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s). \tag{2.83}$$

We recall that $\sigma(s)$ should be a predictable process and is subjected to the condition $\int_0^T \sigma^2(t)dt < \infty$ almost surely. $\mu(t)$ is, on the other hand, an adapted continuous process of finite variation. In equation (2.83) $\int_0^t \sigma(s)dB(s)$ represents the diffusion part of the process and $\int_0^t \mu(s)ds$ does not contain the noise; therefore it represents the drifting of the process. $Y(t)$ is called an Ito process and in differential notation we can write,

$$dY(t) = \mu(t)dt + \sigma(t)dB(t). \tag{2.84}$$

Equation (2.84) can be quite useful in practical applications where the main driving force is perturbed by an irregular noise. A particle moving through a porous medium is such an example. In this case, advection gives rise to the drift term and hydrodynamic dispersion and micro diffusion give rise to the "diffusion" term. In the population dynamics example we previously discussed, the diffusion term is a direct result of noise in the proportionality constant. Therefore it is important to study Ito process further in order to apply it in modeling situations. $\mu(t)$ is called the drift coefficient and $\sigma(t)$ the diffusion coefficient and they can depend on $Y(t)$ and/or $B(t)$. For example, we can write in pervious result (equation (2.81)),

$$d(B(t)^2) = dt + 2B(t)dB(t). \tag{2.85}$$

This is an Ito process with the drift coefficient of 1 and the diffusion coefficient of $2B(t)$. Quadratic variation of Ito process on $[0,T]$

$$Y(t) = Y(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s) \tag{2.86}$$

is given by

$$[Y,Y](t) = \int_0^t \sigma^2(s)ds\,. \tag{2.87}$$

This can be deduced from the fact that $\int_0^t \mu(s)ds$ is a continuous function with finite variation and using quadratic variation of Ito integral. In differential notation,

$$\begin{aligned}(dY(t))^2 &= dY(t).dY(dt),\\ &= \mu^2(t)(dt)^2 + 2\mu\,\sigma dtdB + \sigma^2(dB)^2,\\ &= \sigma^2(t)dt.\end{aligned} \tag{2.88}$$

The chain rule given in equation (2.76) gives us a way to compute the behavior of a function of Brownian motion. It is also useful to know the chain rule to compute a function of a given Ito process. Suppose Ito process is given by a general form,

$$dX(t) = \mu dt + \sigma dB(t)\,. \tag{2.89}$$

where μ is the drift coefficient and B is the diffusion coefficient and let $g\ (t,\ x)$ is a twice differentiable continuous function. Let $Y\ (t) = g(t,\ X\ (t))$. Here $Y\ (t)$ is a function of t and Ito process $X\ (t)$, and also a stochastic process. $Y\ (t)$ can also be expressed as an Ito process. Then Ito formula states,

$$dY(t) = \frac{dg}{dt}(t, X(t))dt + \frac{\partial g}{\partial x}(t, X(t))dX(t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X(t)).(dX(t))^2 \tag{2.90}$$

where ,

$$(dX(t))^2 = d(X(t)).d(X(t))\,, \tag{2.91}$$

and is evaluated according to the rules given by equation (2.74).

As an example, consider the Ito process

$$dX(t) = dt + 2B(t)dB(t) \tag{2.92}$$

where $\mu = 1$ and $\sigma = 2B(t)$.

Assume $g(t,x) = x^2$, therefore

$$\frac{\partial g}{\partial t} = 0;\ \frac{\partial g}{\partial x} = 2x;\ \frac{\partial^2 g}{\partial^2 x} = 2. \tag{2.93}$$

Substituting to Ito formula above,

$$\begin{aligned} dg &= 2X(t)dX(t) + dX(t).dX(t), \\ &= 2(dt + 2B(t)dB(t))X(t) + 4B^2(t)dt. \end{aligned} \tag{2.94}$$

Therefore,

$$dX^2(t) = (2X(t) + 4B^2(t))dt + 4X(t)B(t)dB(t). \tag{2.95}$$

As seen above $dX^2(t)$ is also an Ito process with $u = 2X(t) + 4B^2(t)$ (drift coefficient), a function of $X(t)$ and $B(t)$, and $v = 4X(t)B(t)$ (diffusion coefficient), also a function of $X(t)$ and $B(t)$.

Substituting $X(t) = B^2(t)$ to equation (2.95),

$$\begin{aligned} d(B^4(t)) &= 2(B^2(t) + 2B^2(t)dt + 4(B^2(t))B(t)dB(t), \\ &= 6B^2(t)dt + 4B^3(t)dB(t). \end{aligned} \tag{2.96}$$

We can derive this from chain rule for a function of $B(t)$ as well.
Let $g(x) = x^4$ and from Ito formula (equation (2.76)):

$$\begin{aligned} dg &= g'(B(t))dB(t) + \frac{1}{2}g''(t)dt, \\ &= 4B^3(t)dB(t) + \frac{1}{2}4.3.B^2(t)dt. \end{aligned} \tag{2.97}$$

$$d(B^4(t)) = 6B^2(t)dt + 4B^3(t)dB(t). \tag{2.98}$$

This is the same Ito process as in equation (2.98). Let us consider another example which will be useful later in this book. Consider the function $g(x) = \ln x$ and the Ito process

$$dX(t) = \frac{1}{2}X(t) + X(t)dB(t). \tag{2.99}$$

For this Ito process $\mu = \frac{1}{2}X(t)$ and $\sigma = X(t)$.

From Ito formula (equation (2.90)),

$$\begin{aligned} d(\ln X(t)) &= \frac{1}{X(t)}dX(t) + \frac{1}{2}\left(-\frac{1}{X^2(t)}\right)(X^2(t)dt), \\ &= \frac{1}{X(t)}\left(\frac{1}{2}X(t)dt + X(t)dB(t)\right) - \frac{1}{2}dt, \\ &= \frac{1}{2}dt + dB(t) - \frac{1}{2}dt, \\ &= dB(t). \end{aligned} \tag{2.100}$$

By convention, the above stochastic differential is given by the following integral equation:

$$\ln X(t) = \ln X(0) + \int_0^t dB(t), \tag{2.101}$$

$$\ln\left[\frac{X(t)}{X(0)}\right] = B(t), \tag{2.102}$$

$$X(t) = X(0)e^{B(t)}.$$

We can show that $X(t) = X(0)e^{B(t)}$ satisfies $dX(t) = \frac{1}{2}X(t)dt + X(t)dB(t)$. In other words $X(t) = X(0)e^{B(t)}$ is a "solution" to the stochastic differential

$$dX(t) = \frac{1}{2}X(t)dt + X(t)dB(t).$$

This idea of having a solution to a stochastic differential is similar to having a solution to differential equations in standard calculus.

2.13.5 Stochastic Product Rule

Suppose $X_1(t)$ and $X_2(t)$ are Ito processes given by the following differentials:

$$dX_1(t) = \mu_1(t)dt + \sigma_1(t)dB(t), \tag{2.103}$$

$$dX_2(t) = \mu_2(t)dt + \sigma_2(t)dB(t). \tag{2.104}$$

Quadratic covariation is given by

$$\begin{aligned} d[X_1, X_2] &= dX_1(t).dX_2(t) \\ &= \mu_1\mu_2(dt)^2 + \mu_1\,\sigma_2 dt.dB(t) + \mu_2\,\sigma_1 dt.dB(t) + \sigma_1\sigma_2(dB(t))^2 \end{aligned}$$

and

$$(dt)^2 = dt.dB(t) = 0.$$

$$\begin{aligned} d[X_1, X_2] &= \sigma_1(t)\sigma_2(t)(dB(t))^2, \\ &= \sigma_1(t)\sigma_2(t)dt. \end{aligned} \tag{2.105}$$

The stochastic product rule is given by,

$$X_1(t)\,X_2(t) - X_1(0)X_2(0) = \int_0^t X_1(s)dX_2(s) + \int_0^t X_2(s)dX_1(s) + [X_1, X_2](t). \tag{2.106}$$

If at least one of X_1 and X_2 is a continuous function with finite variation, then $[X_1, X_2](t) = 0$ and equation (2.106) reduces to the integration by parts formula in the standard calculus.

Stochastic product rule can be expressed in differential form:

$$d(X_1(t)\,X_2(t)) = X_1(t)dX_2(t) + X_2(t)dX_1(t) + \sigma_1(t)\sigma_2(t)dt. \tag{2.107}$$

As an example, consider $Y(t) = t\,B(t)$,

$$Y(t) = X_1(t)X_2(t),$$

where $X_1(t) = t$, a continuous function with finite variation and $\sigma_1 = 0$, and $X_2(t) = B(t)$, Brownian motion with infinite variation and $\sigma_2 = 1$.

From the product rule,

$$d(Y(t)) = tdB(t) + B(t)dt + (0)(1)dt,$$

$$d(tB(t)) = tdB(t) + B(t)dt. \tag{2.108}$$

This is the same result we obtain if we use the standard product rule. The reason for this is that quadratic covariation of a continuous function and a function with infinite variation is zero as we have discussed previously.
Suppose

$$dX_1(t) = tdB(t) + B(t)dt, \text{ and}$$
$$dX_2(t) = \frac{1}{2}X_2(t)dt + X_2(t)dB(t),$$

where

$$\mu_1(t) = B(t); \sigma_1(t) = t; \sigma_2(t) = X_2(t); \text{ and } \mu_2(t) = \frac{1}{2}X_2.$$

From the product rule,

$$\begin{aligned} d(X_1(t)X_2(t)) &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + \sigma_1\sigma_2 dt, \\ &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + tX_2(t)dt. \end{aligned} \tag{2.109}$$

By substitution,

$$\begin{aligned} d(X_1(t)X_2(t)) &= X_1(t)(\frac{1}{2}X_2dt + X_2(t)dB(t)) \\ &\quad + X_2(t)(tdB(t) + B(t)dt) + tX_2(t)dt, \\ &= \left(\frac{1}{2}X_1(t)X_2(t) + X_2(t)B(t) + tX_2(t)\right)dt \\ &\quad + (X_1(t)X_2(t) + tX_2(t))dB(t). \end{aligned} \tag{2.110}$$

This is again an Ito process.

$$\begin{aligned} d(X_1(t)X_2(t)) &= \left(\frac{1}{2}X_1(t)X_2(t) + tX_2(t) + X_2(t)B(t)\right)dt \\ &\quad + \left(X_1(t)+t\right)X_2(t)dB(t), \\ &= X_2(t)\left(\frac{1}{2}X_1(t) + t + B(t)\right)dt \\ &\quad + X_2(t)\left(X_1(t)+t\right)dB(t)). \end{aligned} \tag{2.111}$$

As an integral equation,

$$X_1(t)X_2(t) - X_1(0)X_2(0) = \int_0^t X_2(t)(\frac{1}{2}X_1(t) + t + B(t))dt + \int_0^t X_2(t)(X_1(t)+t)dB(t).$$

2.13.6 Ito Formula for Functions of Two Variables

If $g(x_1, x_2)$ is a continuous and twice differentiable function of x_1 and x_2 and Ito processes are of the forms

$$dX_1(t) = \mu_1 dt + \sigma_1 dB(t) \tag{2.112}$$

and

$$dX_2(t) = \mu_2 dt + \sigma_2 dB(t) \ . \tag{2.113}$$

Then $g(X_1(t), X_2(t))$ is also an Ito process and given by the following differential form:

$$\begin{aligned} dg(X_1(t), X_2(t)) &= \frac{\partial g(X_1)}{\partial x_1}dX_1(t) + \frac{\partial g(X_2)}{\partial x_2}dX_2(t) + \frac{1}{2}\frac{\partial^2 g(X_1)}{\partial^2 x_1}(dX_1(t))^2 \\ &\quad + \frac{1}{2}\frac{\partial^2 g(X_2)}{\partial^2 x_2}(dX_2(t))^2 + \frac{\partial^2 g(X_1 X_2)}{\partial x_1 x_2}dX_1(t).dX_2(t). \end{aligned} \tag{2.114}$$

Using quadratic variation and covariation of Ito processes,

$$(dX_1(t))^2 = dX_1(t).dX_1(t) = \sigma_1^2 dt,$$

$$(dX_2(t))^2 = dX_2(t).dX_2(t) = \sigma_2^2 dt, \text{ and} \tag{2.115}$$

$$dX_1(t).dX_2(t) = \sigma_1\sigma_2 dt.$$

These can be considered as a generalization of the rules on differentials given by equation (2.74). We use this generalized Ito formula for a function of two Ito processes in the following example.

We will express the stochastic process $X(t) = 2 + t + e^{B(t)}$ as an Ito process having the standard form, $dX(t) = \mu dt + \sigma dB(t)$.
We can consider

$$X(t) = g(t, B(t)) = 2 + t + e^{B(t)}. \tag{2.116}$$

Therefore, $g(t, y) = 2 + t + e^{y}$,

where $X_1(t) = t$,
$X_2(t) = y = B(t)$.

These equations give, $dX_1 = dt$ and $dX_2 = dB(t)$,
where $\mu_1 = 1; \sigma_1 = 0; \mu_2 = 0;$ and $\sigma_2 = 1$.

Using equation (2.114),

$$\begin{aligned} dg &= (1)dt + e^{B(t)}dB(t) + \frac{1}{2}(0)(dB(t))^2 + \frac{1}{2}e^{B(t)}(dB(t))^2 \\ &\quad + (0)dt.dB(t), \\ &= dt + e^{B(t)}dB(t) + \frac{1}{2}e^{B(t)}dt. \end{aligned} \tag{2.117}$$

Using $dB(t))^2 = dt$,

$$\begin{aligned} dg &= dX(t), \\ &= \left(1 + \frac{1}{2}e^{B(t)}\right)dt + e^{B(t)}dB(t). \end{aligned} \tag{2.118}$$

From a previous example,

$$d(e^{B(t)}) = e^{B(t)}dB(t) + \frac{1}{2}e^{B(t)}dt\,. \tag{2.119}$$

Therefore

$$\begin{aligned} dX(t) &= dt + (\frac{1}{2}e^{B(t)}dt + e^{B(t)}dB(t)), \\ &= dt + d(e^{B(t)}). \end{aligned} \tag{2.120}$$

From the integral notation,

$$X(t) = X(0) + \int_0^t dt + \int_0^t d(e^{B(t)})\,, \tag{2.121}$$

$$X(t) = (0) + t + e^{B(t)} - 1,$$

$$X(t) = (X(0) - 1) + t + e^{B(t)}\,. \tag{2.122}$$

Comparing with

$$X(t) = 2 + t + e^{B(t)},$$
$$X(0) - 1 = 2,$$
$$X(0) = 3.$$

$X(t)$ = constant + t + $e^{B(t)}$ can be considered as a solution process to the stochastic differential,

$$dX(t) = (1 + \frac{1}{2}e^{B(t)})dt + e^{B(t)}dB(t)\,.$$

As we can see in the above solution, the solution process contains the characteristics of both the drift and diffusion phenomena. In this case, diffusion phenomenon dominates as t increases because of the expected value of the exponential of Brownian motion increases at a faster rate in general. If we examine the drift term of the stochastic differential above, we see that the drift term is also affected by the Brownian motion, so the final solution is always a result of complex interactions between the drift term and the diffusion term.

2.14 Stochastic Population Dynamics

We will now move back to discuss the population dynamics example equipped with the knowledge of Ito process and formula:

$$\frac{dx(t)}{dt} = \alpha(t)x(t). \tag{2.123}$$

If $\alpha(t) = r(t) + \sigma(t)W_t$,

where

W_t = white noise,

then

$dX(t) = (r(t)dt + \sigma(t)dB(t)).X(t)$,

and Brownian motion increments $dB(t) = W_t dt$.

Therefore ,

$$dX(t) = \left(r(t)dt + \sigma(t)W_t dB(t)\right)X(t),$$
$$dX(t) = r(t)X(t)dt + \sigma(t)X(t)d(t). \tag{2.124}$$

As seen from the above equation (2.124), $X(t)$ is an Ito process.

Consider the case with $r(t) = r$, a constant and $\sigma(t) = \sigma$, a constant then the process $X_1(t)$ can be written in the differential form:

$$dX(t) = rX(t)dt + \sigma X(t)dB(t). \tag{2.125}$$

Assume $g(x)$=ln x, and using the Ito formula,

$$dg(x(t)) = \frac{\partial(\ln x)}{\partial x}dX(t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dX(t))^2,$$

$$d(\ln(X(t)) = \frac{dX(t)}{X(t)} + \frac{1}{2}\left(\frac{-1}{X(t)^2}\right)(\sigma^2 X^2(t)(dB(t))^2),$$

$$= \frac{dX(t)}{X(t)} - \frac{\sigma^2}{2}dt = \frac{1}{X(t)}(rX(t)dt + \sigma X(t)dB(t)) - \frac{\sigma^2}{2}dt.$$

$$d(\ln(X(t)) = rdt + \sigma dB(t) - \frac{\sigma^2}{2}dt,$$

$$= \left(r - \frac{\sigma^2}{2}\right)dt + \sigma dB(t).$$

Converting back to the integral form,

$$\ln(X(t)) = \ln(X(0)) + \int_0^t \left(r - \frac{\sigma^2}{2}\right)dt + \int_0^t \sigma dB(t),$$

$$\ln\left(\frac{X(t)}{X(0)}\right) = \left(r - \frac{\sigma^2}{2}\right)t + \sigma B(t),$$

$$X(t) = X(0)\exp(\sigma B(t))\exp\left((r - \frac{\sigma^2}{2})t\right). \tag{2.126}$$

$X(t)$ process, therefore, satisfies the Ito process,

$$dX(t) = rX(t)dt + \sigma X(t)dB(t),$$

and equation (2.126) can be considered as a solution to the stochastic differential equation. As discussed earlier, this solution significantly different from its deterministic counter part.

In this chapter we have revised the essential results from stochastic calculus and presented the results which could be useful in developing models and solving stochastic differential equations. While analytical expressions are quite helpful to understand stochastic processes, computer simulation provides us with an intuitive "feel" for the simulated phenomena. Sometimes it is revealing to simulate a number of realizations of a process and visualize them on computers to understand the behavior of the process. We will devote the next chapter to computer simulation of Brownian motions, Ito integrals, Ito processes and the solutions of a limited number of stochastic differential equations.