In these notes we offer an informal introduction to the Itô integral, an important tool for modelling uncertainty in continuous time. When we use this tool we will concentrate on computing the expectation and the variance of random outcomes.

The informal account is of particular, practical significance. As with the ordinary calculus of integration, so too here it is possible to perform calculations without invoking any definitions of integration. Consequently our initial interest will be in establishing manipulative skills which are all based on a facility with the Itô Formula of Section 6. To illustrate the power of these ideas, we look to the Black-Scholes model in finance. As we show in Section ??, the model allows, among other applications, a mechanism for the pricing of options and a mechanism for determining an optimal growth portfolio. (The latter turns out to be equivalent to the pricing mechanism). Along the way we do our best to gain familiarity with how uncertainty influences expected outcomes, and, since the Itô Formula refers to the second derivative, this unavoidably implies the introduction of convexity or concavity. So it is advisable to review Section ??.

Whilst the mathematics of this application is maintained at an informal level throughout this chapter, our approach points the way to an important special tool which requires some measure theory. This is the notion of a martingale, which is not described until we reach Chapter ??. That tool not only opens the door to richer models, but also gives a solid mathematical foundation for the informal arguments of this chapter. Chapter ?? therefore translates some of the arguments from section ?? into a more formal language in order to set an example of how to achieve mathematical precision in this kind of work.

The definition of the Itô integral builds on many of the measure theory ideas associated with Lebesgue integration. Nevertheless it is in some ways more closely reminiscent of the Riemann-Stieltjes integral because of the form that the partial sums take. On the other hand, it does admit a much more general integrator $dz(t)$. We will think of the argument t as representing time. Thus instead of being a function of time with real number values, the integrator $z(t)$ is a function of time taking as values random variables; likewise the integrand is also allowed to be a 'random variable'-valued function of time. The possible time paths that the random variable integrator $z(t)$ may take are restricted by a probability law so that, for instance, decent-looking differentiable paths are altogether ruled out, as they form a set of zero probability. Equally too, many very indecent paths are ruled out, which thus leaves only an 'in-between' class of continuous paths for consideration.

Although the partial sums are similar in form to the Riemann-Stieltjes sums on account of a finite partition of the time interval, the partial sums here are random variables. Random variables are treated like vectors and so discrepancies between them are computed using norms which whilst they are similar to the usual Pythagorean norm are nevertheless related to variance. These norms ignore differences which might occur on paths ruled out as unlikely by the probability law.

This kind of approach requires convergence notions at a higher level of sophistication because use is made of norms in the vector space of continuous functions . So we begin, by way of motivation, with an informal account of the main issues, leaving a rigorous definition and further developments to the next chapter.

1. Modelling 'rates of return'

We are concerned with a method for modelling a quantity which, despite varying continuously over time (with values remaining close when times are close), nevertheless retains a degree of uncertainty as to its value in the future. We have in mind share prices (or stock prices, to give them their American name); these are quoted on the stock exchange on a second by second basis throughout a trading day. Under normal circumstances it is clear that the price quoted in the next second will be close to the current price, thus it makes mathematical sense to regard both time and price as varying continuously. Nevertheless, on each day, at the opening of trade, there is plenty of uncertainty as to what the closing price of the day will be. To an economist it is almost second nature to study the behaviour of the average 'rate of return' on the share price, that is the percentage, or proportional, growth $\Delta p/p$ of the price p, and to assume a direct dependence between the share price return (known also as growth, or yield) and the time lapse, say in the general form

$$\frac{\Delta p_t}{p_t} = f(t, p_t)\Delta t,$$

where $f(t, p)$ is a continuous function, or maybe just a constant. This model, however, implies that an instantaneous growth, or limiting average return, exists in the shape of dp/dt, equal to $f(t, p)$. For instance when $f = const = \mu$, this entails, on passing to the limit and integrating, that

$$\int \frac{dp}{p} = \int \mu dt,$$

and so $p = p(0) \exp \mu t$. The idea is a convenient one, in that it does allow at the very least to pretend that a price growth of μ has been observed, but unfortunately only after the event,. On the other hand a glance at the jagged features of the graph of any particular share price over time makes the existence of the limit dp/dt seem implausible. Given knowledge of the average growth in the past, one might at best take the practical view that there is only an *anticipated* growth μ and that one might super-impose on top of it an *unanticipated* contribution say in the form of a high frequency wave; for instance at its simplest, with σ, k constant and k large, one might add in

$$\sigma \cos(2\pi kt)\Delta t$$

to obtain

$$\frac{\Delta p_t}{p_t} = \mu\Delta t + \sigma \cos(2\pi kt)\Delta t.$$

The function $\cos(2\pi kt)$ oscillates rapidly for large k (it has periodicity $1/k$, since for $t = 1/k$ we have $2\pi kt = 2\pi$). Passing to the limit and integrating now gives

$$\log p - \log p(0) = \mu t + \frac{\sigma}{2\pi k} z(t), \tag{1.1}$$

where $z(t) = \sin(2\pi kt)$. This strategy enables us to distort the underlying exponential trend with an additional source of noise on top of the basic graph in this case in the form of a scalar multiple of a standard fluctuation as represented by $z(t) = \sin(2\pi kt)$; but, the approach suffers from being deterministic, i.e. future values are known ahead of time. A new idea is needed.

2. The Discrete Random walk

To describe the simplest possible idea, begin by fixing an arbitrarily fine partition $\{t_0, t_1, , ..., t_n\}$ of the interval $[0, T]$, say into n subintervals all of *equal* length Δt so that $t_i = t_0 + i \cdot \Delta t$. We shall call Δt the *mesh* of the subdivision. For each $t = t_i$ select a number x_t which is one of $\pm h$, where the sign of h is obtained by the toss of a fair coin (i.e. each face has equal probability showing up). We will later find tossing a biased coin also helpful. The value of h is left unspecified just now. We will see in the next section a natural standardizing choice.

Thus

$$\frac{\Delta p_t}{p_t} = \mu \Delta t + \sigma x_t, \tag{2.1}$$

where $\Delta p_t = p_{t+\Delta t} - p_t$. Hence summing we obtain

$$\begin{aligned} \sum_{i=0}^{n-1} \frac{\Delta p_{t_i}}{p_{ti}} &= \sum_{i=0}^{n-1} \mu \Delta t_i + \sigma \sum_{i=0}^{n-1} x_{t_i}, \\ &= \mu T + \sigma z_T, \end{aligned}$$

where the standard fluctuation is now $z_{t_j} = x_{t_1} + ... + x_{t_j}$ for $j \geq 1$. With a fine enough partition we are tempted to write when $s = t_m$ that

$$\log p(s) - \log p(0) = \int_{p(0)}^{p(s)} \frac{dp}{p} \simeq \sum_{i=0}^{m-1} \frac{\Delta p_{t_i}}{p_{ti}} = \mu s + \sigma z_s. \tag{2.2}$$

Equality is *not* being asserted here (see (6.3) in Section 5 for comparison), but broadly this is a reasonably similar conclusion to the earlier result (1.1). One may of course select the x_t using other randomizing devices and the reader is invited to follow this through on a computer, eg in Mathematica (see the **exercises**).

The main thing to realize at this stage is that the imprecise relation (2.2) (which we shall improve upon) attempts to draw out a logical connection between two quantities $\log p(s)$ and z_s with (2.1) as starting point. Though we do not know at time 0 what the price $p(s)$ at time s will be, we are nevertheless able to say something about its dependence on the sequence of events x_{t_i}. Whilst we do not know what the actual sequence of x_{t_i} will be, we may form a 'calculus' of expectations about $\log p(s)$ based on averaging over all possible outcomes for the sequence of events x_{t_i}. This implies treating the x_t in the equation (2.1) like a random variable: it is a symbol leading to the construction of any one of the 2^n possible sequences $(x_{t_1}, ..., x_{t_n})$. In other words we are dealing with an n-term sequence of random variables indexed by $t_1, ..., t_n$.

The next step is to ask for a limiting version as $\Delta t \to 0$. Clearly we will need to use continuous functions z_t i.e. functions defined for all times t in $[0, T]$, rather than just at a sequence of times $t_1, ..., t_n$. Again we will want to take averages over all various possible functions z_t that may arise over time. This obviously necessitates a further departure from our original notion of a random entity (described so far by a single vector of the possible numerical outcomes) leading us to embrace a new one, that of a function whose values at any instant are vectors of possible outcomes.

It now appears that we need to generalize from the notion of a 'random variable' X to that of a 'random function' Z. The variable X varied over numbers x, each of which could be a 'realization' of X. When the state is ω the number x determined by the state ω is $X(\omega)$. The symbol Z ranges over a collection of functions z, each one being a possible 'realization' of Z. When the state is ω the function z determined by hye state ω is $Z(\omega)$.

As we are accustomed to thinking about functions in terms of the numerical value $z(t)$ at time t we need to ask for an interpretation of $Z(t)$. Clearly, if the realization is $z = Z(\omega)$, then we want $z(t) = Z(\omega)(t)$. Thus Z is really a function with two arguments: state and time, in some order. Our convention will be to take the opposite order: time and state, and then if z is the realization of Z determined by the state of nature ω we can write $z(t) = Z(t, \omega)$. So for each time

t the symbol $Z(t)$ must clearly denote a rule (function) allocating to each possible state of nature a number $Z(t, \omega)$. This allows us to interpret Z as a collection of random variables Z_t indexed by time t where $Z_t(\omega) = Z(t, \omega)$. We call a Z like this a 'stochastic process'.

Notice that our convention makes it easy to let t descend to a subscript and also entitles us to talk of $Z(t)$ by dropping the reference to the state ω.

What should the states of nature be for our particular modelling needs? We consider an example.

Example. Since the states of nature select continuous functions, let the set of states, which are after all our basic events, be functions. We take Ω to be the continuous functions on $[0, T]$. Our simplest example of a 'stochastic process' Z would then be illustrated by

$$Z(t, \omega) = \omega(t).$$

This is very much an extension of the way in which we model the throw of a die - apart from the presence of time. In the case of a die, the states of nature are the number of dots on a face: $\Omega = \{1, 2, 3, 4, 5, 6\}$. And then the simplest example of random variable X on this population is given by $X(\omega) = \omega$. Only now instead of numbers ω we select functions ω. We will return to this question in a moment.

Comment. The reader may at first be disappointed by this simple example and wonder about its usefulness arguing that any particular state of nature ω will not be manifest until after the passage of time from $t = 0$ to $t = T$. But this is our basic building material. The general answer is that we develop a calculus which enables us to use the partial information about ω revealed as time unfolds to calculate expectations as to the future.

Coming back to our discrete random walk, let us observe that if we are to pass to any limit we must now make some kind of connection between values returned the way we have done initially, that is for discrete time arguments t_i, to values returned for all times t in $[0, T]$. And then we need to investigate the effects of increasing the number of discrete times. In order to compare like with like,when considering the graphs generated by different random walks with different Δt, we can make use of linear interpolation to define z_t for all values of t in between the partitioning points. We do this by defining z_t for $t_k < t < t_{k+1}$ as follows:

$$z_t = z_{t_k} \left(\frac{t - t_k}{\Delta t} \right) + z_{t_{k+1}} \left(\frac{t_{k+1} - t}{\Delta t} \right),$$

with z_0 arbitrary. Thus the graph is a broken line. We say that the graph is **piecewise-linear**. Figure xx shows a typical graph, which may be interpreted as the progress of a random (or drunken) walk of a person who, time after time, takes a step of length h forward or back. Notice that the interpolation makes z_t into a continuous function, one selected at random from among 2^n different versions (constructed according to the n binary choices of sign in $\pm h$).

Figure. Random walk (to come)

We are using as sample space Ω the collection of all continuous paths or functions $\omega(t)$ (also written ω_t) with domain $[0, T]$. However, at this stage we have ignored all but the functions which are piecewise linearly constructed by reference to a sequence of n sign choices. Let R_n denote the set of all 2^n possible piecewise linearly random walks. These are treated as equally likely basic events. In other words we place probabilties of 2^{-n} on each one of the 2^n random walks in R^n

and this creates a probability P_n on Ω. We define the measure P_n on all subsets A of Ω by setting

$$P_n(A) = \frac{\sharp(A \cap R_n)}{2^n},$$

where for subsets B of R_n the symbol $\sharp(B)$ denotes the number of elements in B. A limiting measure P is then sought; without wishing to dwell on this point until a later chapter, passage to the limit requires us, properly speaking, to be much more circumspect about the sets A which we allow into consideration; this is explained in Chapter ?? .

Given the measures P_n we can proceed to compute expected outcomes of any path-related numerical property (and of course property= function f) along the paths ω selected by nature as

$$E[f] = \sum_{\omega \in \Omega} P_n(\omega) f(\omega).$$

Despite its appearance, this is really a finite sum (extending only over R_n). Of utmost importance is the fact that the summation ignores paths outside of R_n.

In seeking a limiting probability P on Ω our aim is to simplify the calculation. Indeed, note that the sum could already be presented equivalently as an integral

$$\int_\Omega f(\omega) dP_n(\omega)$$

and we hope that in the limit this is well approximated by the integral

$$\int_\Omega f(\omega) dP(\omega)$$

which in turn will have simpler properties. This is the prime purpose of this chapter and we seek to achieve it by referring only to a few nice properties of the stochastic process z_t which are derivable from P. We should therefore properly speak of z_t as a process described by the probability P.

The precise notion of limit that needs to be applied to the measures P_n is examined later in Section ?? . From now on and until that section we pass over such niceties and proceed intuitively.

3. Brownian motion - a Limiting Random Walk

We return to our lead example and re-write equation (2.1) for the growth rate as a sum of two components: an *anticipated* term and an *unanticipated* term (known also as the v**olatility term**, or noise)

$$\frac{\Delta p_t}{p_t} = \mu \Delta t + \sigma \Delta z_t, \tag{3.1}$$

where $\Delta z_t = z_{t+\Delta t} - z_t$, and t may be any one of the t_i in the partition. We point out that here z_t denotes not a deterministic function but a random variable whose value depends on the time parameter t. More precisely it is a 'stochastic process' as explained in the last section, but the notation pretends it is just a function. In the background is a probability law describing the properties of the process z_t. We will refer to it as P. What are the basic features of this probability law? Our earlier random selection of x_t leads us to identify first the following property.

Assumption A: *The increments Δz_t are independent and identically distributed under P.*

We have referred to this property as an assumption - because we don't have at our disposal any way of rigorously proving it from our scant knowledge of P.It is however plausible, given our construction in the previous section.

The assumption has several interesting consequences for the variance of the increment Δz_t, as we are about to see. Let us write v for the common value of the variance of the increment (see section ???). Thus $v = var[\Delta z_t] = E[(\Delta z_t)^2] - (E[\Delta z_t])^2$ and hence, since $E[\Delta z_t] = 0$, we have for all t

$$v = E[(\Delta z_t)^2].$$

Secondly, recall the formula for the variance of the sum of two random variables X, Y when they are independent, namely

$$\begin{aligned} var(X+Y) &= var(X) + var(Y) + 2cov(X,Y) \\ &= var(X) + var(Y), \end{aligned}$$

i.e. variance in this case is additive. We thus have by the additivity of variance over the independent increments, that

$$\begin{aligned} var[\sigma(z_T - z_0)] &= var[\sigma\{z_T - z_{T-\Delta t}\} + \sigma\{z_{T-\Delta t} - z_{T-2\Delta t}\} + ...] \\ &= \sum E[(\sigma \Delta z_{t_i})^2] \\ &= \sigma^2 \sum var[\Delta z_t] \\ &= \sigma^2 nv \\ &= \sigma^2 vT/\Delta t. \end{aligned}$$

If, in the limit as $\Delta t \to 0$, the limiting random variable z_t is to have finite variance at time T, the limit of $vT/\Delta t$ must also be finite. This conclusion from Assumption A and the requirement of finite variance at each time leads to the following natural standardization of the limiting process, which as we see in a moment informs a choice for the step-size h mentioned in the previous section.

Assumption B *The process z_t has finite variance under P and*

$$\lim_{\Delta t \to 0} \frac{var(\Delta z_t)}{\Delta t} = 1.$$

We may thus like to think that

$$\Delta z_t = \varepsilon_t \sqrt{\Delta t},$$

where the random variable ε_t has some standardized distribution, yet to be discovered.

Returning to our initial random device in the previous section for selecting Δz_t, we note that it has variance h^2 (equal to $\frac{1}{2}(+h)^2 + \frac{1}{2}(-h)^2$). It now follows that if Assumption B is to hold, we must have $h^2/\Delta t = 1$, i.e. $h = \sqrt{\Delta t}$. Since $\Delta t = T/n$, this means that $h = \sqrt{T/n}$. In this case of course

$$\Delta z_t = \varepsilon_t \sqrt{\Delta t} \text{ and } \varepsilon_t = \pm 1 \text{ with equal probability.}$$

We can at last make a deduction from this observation about the limiting distribution of ε_t as $\Delta t \to 0$. Consider a fixed interval (a,b) with $0 < a < b < T$. We concern ourselves only with refinements of a partition of $[0,T]$ which includes a, b. Suppose the refinement subdivides (a,b) into m sub-intervals of length $\Delta t = T/n$. The binary sequence of choices ± 1 for the forward or backward movements at each time t_i in the sub-division of (a,b) creates a random walk and there are 2^m of these sequences; if each is to be equally likely, the probability of each has to be 2^{-m}. Suppose a sequence gives rise to i choices of $+1$ and $m-i$ choices of -1. To satisfy Assumption B the unit of movement h, introduced in section 2, needs to be $h = \sqrt{\Delta t}$. The total forward movement is $x = \{i(+1) + (m-i)(-1)\}\sqrt{\Delta t} = (2i-m)\sqrt{\Delta t}$. For fixed m any given movement x uniquely determines i. But the number of ways of distributing i choices of $+1$ among the m binary choices is

$$\begin{pmatrix} m \\ i \end{pmatrix}$$

and so the probability of a total movement of x when $x = (2i-m)\sqrt{\Delta t}$ is

$$2^{-m}\begin{pmatrix} m \\ i \end{pmatrix}.$$

This is of course the binomial distribution (see section ??). According to the **de Moivre-Laplace Theorem** (and also the more general central limit theorem), this distribution tends in the limit to the standard normal, i.e. to one with a probability distribution in which the value x occurs with density

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Thus for very fine partitions of mesh-length Δt the movement between time $t = a$ and time $t = b$ may be described by saying that at each partitioning point the next step is a standard normal variable times $\sqrt{\Delta t}$. It is not therefore surprising that in our informal account we settle for a standard normal distribution to generate random movements.

Summary. *We assume there exists a probability law P on the sample space of continuous functions and that there is a stochastic process z_t which under this law satisfies Assumption A and Assumption B. Such a process will be called a Brownian motion under P.*

Remark. Assumption A implies that past changes have no influence over future changes and so one says that the market price displays no memory effects (is Markovian). This assumption has of course been tested by comparing empirical evidence against predictions from the model; whilst there is good agreement with reality, it does seem that there might be some small element of memory in the markets. To model these memory effects other assumptions are required.

4. A Stochastic Differential Equation

Our next step is to ask how to interpret the equation (3.1) in the limit as $\Delta t \to 0$. As was said earlier we may not divide each side of the equation by Δt and pass to the limit. If differentiation is out of the question, then how about integration? After all we would have to solve the differential equation representing the model and that would require integration. So why not go straight for integration? The two sums on the right-hand side of the following equation converge as $\Delta t \to 0$:

$$\sum_{i=0}^{n-1} \frac{\Delta p_{t_i}}{p_{ti}} = \sum_{i=0}^{n-1} \mu \Delta t_i + \sigma \sum_{i=0}^{n-1} \Delta z_{t_i}. \tag{4.1}$$

So we are tempted to write

$$\int \frac{dp_t}{p_t} = \mu T + \sigma(z_T - z_0),$$

or even

$$\int_0^T \frac{dp_t}{p_t} = \int_0^T \mu dt + \int_0^T \sigma dz_t, \tag{4.2}$$

despite not having a definition for the integrations process on the left-hand side (fortunately the right-most symbol has an easy interpretation). The equation is fortunately correct, as we shall see later. This justifies writing the limiting version of the model (4.1) in differential notation as the equation

$$\frac{dp_t}{p_t} = \mu dt + \sigma dz_t, \tag{4.3}$$

so long as we agree that (4.3) is interpreted as (4.2).The equation (4.3) is known as a *stochastic differential equation*. Hovering in the background is the probability P governing the properties of he process z_t. Whenever we wish to draw attention to this probability we will speak of z_t as being a P-stochastic process.

Caution 1. We caution the reader that (4.2) needs careful consideration in view of the following **non-identity**:

$$\int_0^T \frac{dp_t}{p_t} \neq \log p_T - \log p_0 \qquad !!!$$

In fact, given (4.3), it will turn out that

$$\int_0^T \frac{dp_t}{p_t} = \log p_T - \log p_0 + \frac{1}{2}\sigma^2 T. \tag{4.4}$$

Indeed the first symbol is *not* covered by the definitions of integration introduced so far. The trouble is that p_t as a function of t is *not of bounded variation.* We can in fact estimate the variation of p and of z when for all i we have $\Delta z_{t_i} = \pm\sqrt{\Delta t}$. Clearly

$$\sum_{i=0}^{n-1} |\Delta z_{t_i}| = \sum_{i=0}^{n-1} \sqrt{\Delta t} = n\sqrt{\Delta t} = n\sqrt{T/n} = \sqrt{Tn} \to \infty$$

as $n \to \infty$, since $T = n\Delta t$. Similarly, as $\sigma > 0$ and provided Δt is small enough

$$\frac{|\Delta p_{t_i}|}{p_{t_i}} = |\mu\Delta t_i + \sigma\Delta z_{t_i}| = |\sigma \pm \mu\sqrt{\Delta t}|\sqrt{\Delta t} = \sqrt{\Delta t}(\sigma \pm \mu\sqrt{\Delta t}).$$

Observe that if $p_s = 0$ for some time s, then by (3.1) we have $p_t = 0$ for all $t \geq s$. Suppose, however, that we restrict attention to some closed interval in which p_t is non-zero. By continuity the price remains bounded away from zero. Suppose for some $M > 0$ we have $p_t \geq M$; then from (3.1)

$$|\Delta p_{t_i}| \geq |p_{t_i}| \cdot |\mu\Delta t_i + \sigma\Delta z_{t_i}| \geq M \cdot \sqrt{\Delta t}(\sigma \pm \mu\sqrt{\Delta t}).$$

So

$$\sum_{i=0}^{n-1} |\Delta p_{t_i}| \geq M\left(\sigma \sum_{i=0}^{n-1} \sqrt{\Delta t} + \mu \sum_{i=0}^{n-1} \pm\Delta t\right) \geq M\left(\sigma\sqrt{nT} - \mu T\right).$$

Thus on any closed interval in which the price is non-zero both z_t and p_t have unbounded variation.

Caution 2. (The Substitution Rule) Observe that a realization of the random variable z_t when constrained by the condition $\Delta z_t = \pm\sqrt{\Delta t}$ does have bounded 'quadratic' variation, since $\sum_{i=0}^{n-1}(\Delta z_{t_i})^2 = \sum_{i=0}^{n-1}\Delta t = T$ (likewise, so does the corresponding realization of p_t). Thanks to this circumstance a theory of 'stochastic integration' with respect to continuous functions z_t can still be developed (for appropriate integrands of the form $f(t,\, z_t)$). Once again we point out that this integration with respect to t is governed by a probability measure P understood from the context as defined on appropriate subsets of the sample space of possible realizations of the continuous path z_t so that in effect the only functions z_t which are relevant are those which have some limiting resemblance to the random discrete walks governed by the constraint $\Delta z_t = \pm\sqrt{\Delta t}$. This was first realized and considered mathematically by Itô in 1944. (It was also used in an intuitive form by Feynman in 1949 in his work on possible paths taken by elementary particles). The rules of such a 'stochastic calculus' are however much different. Principally the Substitution Rule: $\frac{d}{dt}f(z(t)) = f'(z(t))z'(t)$ is false, and this can be traced to the fact that dz_t/dt does not exist along the limiting random walks. As a result for continuous z_t

$$\int_0^T f'(z_t)dz_t \textbf{ does not equal } f(z_T) - f(z_0).$$

(To have a valid equality the functions z_t need to be of bounded variation.) By contrast the following is valid

$$\int_0^T df(z_t) = f(z_T) - f(z_0),$$

so emphatically

$$df(z_t) \neq f'(z_t)dz_t. \qquad !!!$$

The fundamental theorem of Newtonian calculus (see end of section 17.5 of AO's book) asserts that for $f(t)$ continuous $\int_a^b f'(t)dt = f(b) - f(a)$, but this is of no help, when we cannot invoke the substitution rule.

5. Itô's Formula - intuitive approach

Let us examine the issue of substitution in the context of a random walk with $\Delta z_t = \pm\sqrt{\Delta t}$. Let $f(x)$ be a 'smooth' function, i.e. one with a continuous second derivative. To second order accuracy in Δz we have by Taylor's Theorem that

$$\begin{aligned}
\Delta f &= f(z_{t+\Delta t}) - f(z_t) \\
&= f(z_t + \Delta z_t) - f(z_t) \\
&= f'(z_t)\Delta z_t + \frac{1}{2}f''(z_t)(\Delta z_t)^2 \\
&= f'(z_t)\Delta z_t + \frac{1}{2}f''(z_t)\Delta t.
\end{aligned}$$

So summing we obtain

$$\begin{aligned}
f(z_T) - f(z_0) &= \{f(z_T) - f(z_{T-\Delta t})\} + \{f(z_{T-\Delta t}) - f(z_{T-2\Delta t})\} + \\
&\quad ... + \{f(z_{\Delta t}) - f(z_0)\} \\
&= \sum \Delta f(z_t) \\
&= \sum f'(z_t)\Delta z_t + \frac{1}{2}\sum f''(z_t)\Delta t.
\end{aligned}$$

Assuming one may define the relevant limits, we may hope to obtain the formula.

$$f(z_T) - f(z_0) = \int_0^T f'(z_t)dz_t + \frac{1}{2}\int_0^T f''(z_t)dt. \tag{5.1}$$

As an aside we should acknowledge that actually the 'Taylor expansion' just used ignores an error e_t in each summand, which, thanks to the smoothness assumption, is small relative to Δt i.e.

$$\lim_{\Delta t \to 0} \frac{|e_t|}{\Delta t} = 0. \tag{5.2}$$

(Compare section 17.5.) Fortunately the limiting cumulative error tends to zero, since

$$\sum |e_t| \leq N \max |e_t| \leq T \max \frac{|e_t|}{\Delta t} \to 0$$

as $|e_t|/\Delta t \to 0$ as $\Delta t \to 0$ by (5.2).

Remark. The argument is more complicated when $\Delta z_t = \varepsilon_t \sqrt{\Delta t}$ with ε_t standard normal. It is true that $E[\varepsilon_t^2] = 1$, so $E[(\Delta z_t)^2] = \Delta t$, but intuition balks at the prospect of comparing the two quantities $(\Delta z_t)^2$ and Δt, the first being random and the second, being the expectation of the first, merely a constant. Yet the routine and customary inspection of the variance of $(\Delta z_t)^2$ shows that variance to be of order $(\Delta t)^2$ and in fact:

$$var[(\Delta z_t)^2] = E\left[\left((\Delta z_t)^2 - \Delta t\right)^2\right] = 2(\Delta t)^2\ . \tag{5.3}$$

Thus Δz_t is in the limit close to being non-random. For a proof note that, since $E[\varepsilon_t^4] = 3$, we have $E[(\Delta z_t)^4] = 3(\Delta t)^2$ and so

$$\begin{aligned} E[(\Delta z_t)^4 - 2(\Delta z_t)^2\Delta t + (\Delta t)^2] &= E[(\Delta z_t)^4] - (\Delta t)^2 \\ &= 2(\Delta t)^2. \end{aligned} \tag{5.4}$$

Thus, summing and observing that $N \cdot (\Delta t)^2 = T \cdot \Delta t$, we conclude that the variance of $\sum (\Delta z_t)^2$ is extremely small and tends to zero as $\Delta t \to 0$. This gives more credence to the argument leading to (5.1). It implies that we may need to treat two random variables as identical (more properly: 'equivalent' as discussed in Chapter 2) if they have the same mean and variance. We will see later a generalization of (5.4) lifting the result to a sum of differentials.

We can repeat the argument leading to formula (5.1) by considering Δp instead of Δz when

$$\Delta p_t = a(p_t)\Delta t + b(p_t)\Delta z_t.$$

Here $a(p)$ and $b(p)$ are continuous functions of p. Assuming first that $\Delta z_t = \pm\sqrt{\Delta t}$ we have that

$$(\Delta p_t)^2 = a^2(p_t)(\Delta t)^2 \pm 2a(p_t)b(p_t)(\Delta t)^{3/2} + b^2(p_t)\Delta t, \tag{5.5}$$

or to first-order in Δt

$$(\Delta p_t)^2 = b^2(p_t)\Delta t.$$

Next assuming $\Delta z_t = \varepsilon_t\sqrt{\Delta t}$ with ε_t normally distributed, and modifying (5.5) appropriately, we see that, since $E[\varepsilon_t^2] = 1$, the expected value of $(\Delta p_t)^2$ is

$$E[(\Delta p_t)^2] = a^2(p_t)(\Delta t)^2 + b^2(p_t)\Delta t.$$

To first order this is $b^2(p_t)\Delta t$. Turning to variance analysis we see that $var[(\Delta p_t)^2]$ is evidently of order $(\Delta t)^2$. In fact

$$E[\big((\Delta p_t)^2 - b^2(p_t)\Delta t\big)^2] = 2b^4(p_t)(\Delta t)^2 + 4b^2(p_t)a^2(p_t)(\Delta t)^3 + +a(p_t)^4\Delta t^4.$$

Now we compute using a Taylor expansion of order 2 in Δp. Dropping all but the terms first-order in Δt, we obtain that

$$\begin{aligned} f(p_T) - f(p_0) &= \sum \Delta f \\ &= \sum f'(p_t)\Delta p_t + \frac{1}{2}\sum f''(p_t)(\Delta p_t)^2 \\ &= \sum f'(p_t)\left(a(p_t)\Delta t + b(p_t)\Delta z_t\right) + \frac{1}{2}\sum f''(p_t)b^2(p_t)\Delta t \\ &= \sum \left(f'(p_t)a(p_t) + \frac{1}{2}f''(p_t)b^2(p_t)\right)\Delta t + \sum f'(p_t)b(p_t)\Delta z_t. \end{aligned}$$

Note that the final sum, consists of contributions all of which are of zero expectation so the sum also has zero expectation. Assuming a limit exists we obtain

$$f(p_T) - f(p_0) = \int_0^T f'(p_t)b(p_t)dz_t + \int_0^T \left(f'(p_t)a(p_t) + \frac{1}{2}f''(p_t)b^2(p_t)\right)dt. \tag{5.6}$$

This is known as **Itô's Formula**. It is often quoted in differential form

$$df = bf'dz_t + (af' + \frac{1}{2}f''b^2)dt,$$

but the intended meaning of this notation is as an abbreviation of the related formula with integral signs.

It is worth reflecting on this formula. As in (2.2), only this time exactly, the formula traces the connection between the quantity $f(p_T)$ which is unknown at time 0 and the path z_t which will be selected by nature (of course, granted the assumptions made by the model (2.1) of price incrementation). In particular it opens the way to the calculation of all the expectations one might possibly want by giving them in the form of ordinary integrals; as long as the expectations are of the form $E[f(p_T)]$ for any smooth but otherwise arbitrary function $f()$. All we need to know is that the first integral on the right of (5.6) has zero expectation.

The precise statement of circumstances when the formula (5.6) is valid, known as Itô's Lemma, in particular what wider class of functions $a(), b()$ may be used, requires proper definitions for the 'standard fluctuation' z_t; the argument just given needs to be only slightly adapted to yield the Lemma. See a later Section (below) for an indication.

It is important to realize that the random variables on the two sides of the equation are only 'treated as equal' in a defined sense and that as usual there is lurking in the background a probability measure describing the properties of z_t. (The two sides are of course only 'equivalent' since it is only their expected values and variances that agree under the background measure.)

6. Examples on Itô's Formula

Throughout this section z_t denotes the stochastic process which we have suggested exists as a limit of the random walks and we refer to its as Brownian motion without any indication of the background measure P which describes its behaviour.

We begin by clarifying (4.4). We know that

$$\int_0^T d\log(p_t) = \log p_T - \log p_0,$$

so we apply Itô's Lemma to $f(x) = \log x$.

Example 1. Take $f(x) = \log x$ and recall from (3.1) that

$$dp_t = p_t\mu dt + p_t\sigma dz_t, \tag{6.1}$$

with constant μ, σ. We take $a = \mu p_t, b = \sigma p_t$. So we obtain

$$\begin{aligned}
\log p_T - \log p_0 &= \int_0^T \frac{\sigma p_t dz_t}{p_t} + \int_0^T \left(\frac{\mu p_t}{p_t} - \frac{1}{2}\frac{(\sigma p_t)^2}{{p_t}^2}\right) dt \\
&= \sigma \int_0^T dz_t + \int_0^T \left(\mu - \frac{1}{2}\sigma^2\right) dt \\
&= \sigma(z_T - z_0) + \left(\mu - \frac{1}{2}\sigma^2\right) T. \qquad (6.2)
\end{aligned}$$

Notice that integrating directly from (6.1) gives

$$\int_0^T \frac{dp_t}{p_t} = \mu T + \sigma(z_T - z_0)$$

and so eliminating z_T by way of (6.2) we have, in contrast to Newtonian calculus, that

$$\int_0^T \frac{dp_t}{p_t} = (\log p_T - \log p_0) + \frac{1}{2}\sigma^2 T. \tag{6.3}$$

Example 2. *Find the distribution of the price at time $t = 1$, i.e. of the price p_1.*

Solution. According to (6.2) in the last example we find that

$$\log p_1 = \log p_0 + \left(\mu - \frac{1}{2}\sigma^2\right) + \sigma(z_1 - z_0).$$

The final term has mean zero and variance σ^2. We thus find that $\ln p_1$ has mean m where

$$m = \log p_0 + \left(\mu - \frac{1}{2}\sigma^2\right) \tag{6.4}$$

so that $(\ln p_1 - m)/\sigma = (z_1 - z_0)$ is a standard normal variable.

Example 3. *Find the expected value of the log-normal distribution distribution of the price p_1.*

Solution. Denoting the density symbollically by $g(p_1)$, the expected value is

$$\int_0^\infty p_1 g(p_1) dp_1.$$

Now the transform w of p_1 given by

$$w = \frac{\log p_1 - m}{\sigma}$$

is standard normal and $p_1 = e^{\sigma w + m}$ so we see that

$$\begin{aligned} E[e^{\sigma w+m}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\sigma w+m} e^{-w^2/2} dw \\ &= e^{m+\sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(w^2-2\sigma w+\sigma^2)/2} dw \\ &= \exp(\log p_0 + \mu) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= p_0 e^{\mu}. \end{aligned}$$

In the last but one line we have used a second substitution, the shift $y = w - \sigma$.

Alternatively, we may invoke the moment generating function $M(\sigma) = E[e^{\sigma w}]$ to write, more neatly,

$$E[e^{\sigma w+m}] = e^m E[e^{\sigma w}] = e^m e^{\frac{1}{2}\sigma^2} = e^{\mu + \log p_0} = p_0 e^{\mu}, \tag{6.5}$$

using (6.4).

Notice that the expected price grows at a rate μ, as one might hope. Thus the modelling intentions are satisfied: μ represents the average trend in the behaviour of price.

In the next two examples we assume that

$$dS_t = \alpha S_t dt + \beta S_t dz_t. \tag{6.6}$$

Example 4. *Find df for $f(t, x) = e^{-rt} x$.*
Solution. We have $f_t = -rf, f_x = f/x, f_{xx} = 0$, so

$$\begin{aligned} df &= (-rf + \frac{f}{S_t}\alpha S_t + 0)dt + \frac{f}{S_t}\beta S_t dz_t \\ &= (\alpha - r) f dt + \beta f dz_t. \end{aligned}$$

Example 5. *Find df when $f(S_t, t) = \log(S_t e^{-rt})$.*
Solution. Here $f(x, t) = \ln x - rt$. So we have

$$f_t = -r, f_x = \frac{1}{x}, f_{xx} = -\frac{1}{x^2}.$$

Thus

$$\begin{aligned} df &= (-r + \mu S \frac{1}{S} + \frac{1}{2}\sigma^2 S^2 \frac{-1}{S^2})dt + \sigma S \frac{1}{S} dz \\ &= (\mu - r - \frac{1}{2}\sigma^2)dt + \sigma dz. \end{aligned}$$

Example 6. *Find dV when*

$$V = e^{-rt} \int_0^t \sigma e^{rs} dz_s, \tag{6.7}$$

where r is a constant.

Solution. Here we let $S_t = \int_0^t \sigma e^{rs} dz_s$ so that $dS_t = \sigma e^{rt} dz_t$. i.e. $\alpha = 0$ and $\beta = \sigma e^{rt}$. Now $V_t = f(S_t, t)$ where $f(x, t) = e^{-rt} x$ as in Example 4 above. So

$$dV_t = -rV_t dt + e^{-rt}\sigma e^{rt} dz_t = -rV_t dt + \sigma dz_t.$$

The equation

$$dV_t = -rV_t dt + \sigma dz_t$$

describes what is known as the **Ornstein-Uhlenbeck** process. As may be seen from (6.7)

$$E[V_t] = 0,$$

i.e. the process has mean zero. Note also $E[dV_t] = 0$ and that the process has the interesting property of '*reversion to the mean*', meaning that, if V_t is positive the process value tends to decrease (i.e. moves down towards zero), whereas if $V_t < 0$ it tends to move up towards zero. This is of interest in modelling interest rates.