m3f33chVI

Chapter V. STOCHASTIC PROCESSES IN CONTINUOUS TIME

§1. Brownian Motion.

The Scottish botanist Robert Brown observed pollen particles in suspension under a microscope in 1828 and 1829 (though this had been observed before),[1] and observed that they were in constant irregular motion.

In 1900 L. Bachelier considered Brownian motion a possible model for stock-market prices:
BACHELIER, L. (1900): Théorie de la spéculation. *Ann. Sci. Ecole Normale Supérieure* **17**, 21-86
– the first time Brownian motion had been used to model financial or economic phenomena, and before a mathematical theory had been developed.

In 1905 Albert Einstein considered Brownian motion as a model of particles in suspension, and used it to estimate *Avogadro's number* ($N \sim 6 \times 10^{23}$), based on the diffusion coefficient D in the *Einstein relation*

$$var X_t = Dt \qquad (t > 0).$$

In 1923 Norbert Wiener defined and constructed Brownian motion rigorously for the first time. The resulting stochastic process is often called the *Wiener process* in his honour, and its probability measure (on path-space) is called *Wiener measure*.

We define *standard Brownian motion* on $\mathbb{R}$, BM or $BM(\mathbb{R})$, to be a stochastic process $X = (X_t)_{t \geq 0}$ such that
1. $X_0 = 0$,
2. X has *independent increments*: $X_{t+u} - X_t$ is independent of $\sigma(X_s : s \leq t)$ for $u \geq 0$,
3. X has *stationary increments*: the law of $X_{t+u} - X_t$ depends only on u,
4. X has *Gaussian increments*: $X_{t+u} - X_t$ is normally distributed with mean 0 and variance u,

$$X_{t+u} - X_t \sim N(0, u),$$

5. X has *continuous paths*: X_t is a continuous function of t, i.e. $t \mapsto X_t$ is continuous in t.

For time t in a finite interval – $[0, 1]$, say – we can use the following filtered

[1]The Roman author Lucretius observed this phenomenon in the gaseous phase – dust particles dancing in sunbeams – in antiquity: *De rerum natura*, c. 50 BC.

space: (i) $\Omega = C[0,1]$, the space of all continuous functions on $[0,1]$; (ii) the points $\omega \in \Omega$ are thus random functions, and we use the coordinate mappings: X_t, or $X_t(\omega)$, $= \omega_t$; (iii) the filtration is given by $\mathcal{F}_t := \sigma(X_s : 0 \leq s \leq t)$, $\mathcal{F} := \mathcal{F}_1$; (iv) P is the measure on $(\Omega, \mathcal{F})$ with finite-dimensional distributions specified by the restriction that the increments $X_{t+u} - X_t$ are stationary independent Gaussian $N(0,u)$.

Theorem (WIENER, 1923). Brownian motion exists.

The best way to prove this is by construction, and one that reveals some properties. The result below is originally due to Paley, Wiener and Zygmund (1933) and Lévy (1948), but is re-written in the modern language of *wavelet* expansions. We omit the proof; for this, see e.g. [BK] 5.3.1, or SP L20-22. The *Haar system* $(H_n) = (H_n(.))$ is a complete orthonormal system (cons) of functions in $L^2[0,1]$. The *Schauder system* Δ_n is obtained by integrating the Haar system. Consider the triangular function (or 'tent function')

$$\Delta(t) := 2t \quad \text{on} \quad [0, \frac{1}{2}), \qquad 2(1-t) \quad \text{on } [\frac{1}{2}, 1], \qquad 0 \quad \text{else}.$$

With $\Delta_0(t) := t$, $\Delta_1(t) := \Delta(t)$, define the nth *Schauder function* Δ_n by

$$\Delta_n(t) := \Delta(2^j t - k) \qquad (n = 2^j + k \geq 1).$$

Note that Δ_n has

$$\int_0^t H(u)du = \frac{1}{2}\Delta(t),$$

and similarly

$$\int_0^t H_n(u)du = \lambda_n \Delta_n(t),$$

where $\lambda_0 = 1$ and for $n \geq 1$,

$$\lambda_n = \frac{1}{2} \times 2^{-j/2} \qquad (n = 2^j + k \geq 1).$$

The Schauder system (Δ_n) is again a complete orthogonal system on $L^2[0,1]$. We can now formulate the next result; for proof, see the references above.

Theorem (PWZ theorem: Paley-Wiener-Zygmund, 1933). For $(Z_n)_0^\infty$ independent $N(0,1)$ random variables, λ_n, Δ_n as above,

$$W_t := \sum_{n=0}^{\infty} \lambda_n Z_n \Delta_n(t)$$

converges uniformly on $[0,1]$, a.s. The process $W = (W_t : t \in [0,1])$ is Brownian motion.

Thus the above description does indeed define a stochastic process $X = (X_t)_{t\in[0,1]}$ on $(C[0,1], \mathcal{F}, (\mathcal{F}_t), P)$ (the uniform limit of continuous functions is continuous). The construction gives X on $C[0,n]$ for each $n = 1, 2, \cdots$, and combining these: X exists on $C[0,\infty)$. It is also unique (a stochastic process is uniquely determined by its finite-dimensional distributions and the restriction to path-continuity).

No construction of Brownian motion is easy: one needs both some work and some knowledge of measure theory. But *existence* is really all we need, and we assume this. For background, see any measure-theoretic text on stochastic processes. The classic is Doob's book, quoted above (see VIII.2 there). Excellent modern texts include Karatzas & Shreve [KS] (see particularly §2.2-4 for construction and §5.8 for applications to economics), Revuz & Yor [RY], Rogers & Williams [RW1] (Ch. 1), [RW2] Itô calculus – below).

We denote standard Brownian motion $BM(\mathbb{R})$ – or just BM for short – by $B = (B_t)$ (B for Brown), though $W = (W_t)$ (W for Wiener) is also common. Standard Brownian motion $BM(\mathbb{R}^d)$ in d dimensions is defined by $B(t) := (B_1(t), \cdots, B_d(t))$, where $B_1, \cdots, B_d$ are *independent* standard Brownian motions in one dimension (*independent copies* of $BM(\mathbb{R})$).

Zeros.

It can be shown that Brownian motion *oscillates*:

$$\limsup_{t\to\infty} X_t = +\infty, \qquad \liminf_{t\to\infty} X_t = -\infty \qquad a.s.$$

Hence, for every n there are zeros (times t with $X_t = 0$) of X with $t \geq n$ (indeed, infinitely many such zeros). So if

$$Z := \{t \geq 0 : X_t = 0\}$$

denotes the zero-set of $BM(\mathbb{R})$:

1. Z is an *infinite* set.

Next, if t_n are zeros and $t_n \to t$, then by path-continuity $B(t_n) \to B(t)$; but $B(t_n) = 0$, so $B(t) = 0$:
2. Z is a *closed* set (Z contains its limit points).
Less obvious are the next two properties:
3. Z is a *perfect* set: every point $t \in Z$ is a limit point of points in Z. So there are *infinitely many* zeros in *every* neighbourhood of *every* zero (so the paths must oscillate amazingly fast!).
4. Z is a (Lebesgue) *null* set: Z has Lebesgue measure zero.

In particular, the diagram above (or any other diagram!) grossly distorts Z: *it is impossible to draw a realistic picture of a Brownian path.*

Brownian Scaling.

For each $c \in (0, \infty)$, $X(c^2 t)$ is $N(0, c^2 t)$, so $X_c(t) := c^{-1} X(c^2 t)$ is $N(0, t)$. Thus X_c has all the defining properties of a Brownian motion (check). So, X_c **IS** a Brownian motion:

Theorem. If X is BM and $c > 0$, $X_c(t) := c^{-1} X(c^2 t)$, then X_c is again a BM.

Corollary. X is *self-similar* (reproduces itself under scaling), so a Brownian path $X(.)$ is a *fractal*. So too is the zero-set Z.

Brownian motion owes part of its importance to belonging to *all* the important classes of stochastic processes: it is (strong) Markov, a (continuous) martingale, Gaussian, a diffusion, a Lévy process (process with stationary independent increments), etc.

§2. Filtrations; Finite-Dimensional Distributions

The underlying set-up is as before, but now *time is continuous rather than discrete*; thus the time-variable will be $t \geq 0$ in place of $n = 0, 1, 2, \ldots$. The information available at time t is the σ-field $\mathcal{F}_t$; the collection of these as $t \geq 0$ varies is the filtration, modelling the information flow. The underlying probability space, endowed with this filtration, gives us the stochastic basis (filtered probability space) on which we work,

We assume that the filtration is *complete* (contains all subsets of null-sets as null-sets), and *right-continuous*: $\mathcal{F}_t = \mathcal{F}_{t+}$, i.e.

$$\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$$

(the 'usual conditions' – right-continuity and completeness – in Meyer's terminology).

A stochastic process $X = (X_t)_{t \geq 0}$ is a family of random variables defined on a filtered probability space with X_t $\mathcal{F}_t$-measurable for each t: thus X_t is known when $\mathcal{F}_t$ is known, at time t.

If $\{t_1, \cdots, t_n\}$ is a finite set of time-points in $[0, \infty)$, $(X_{t_1}, \cdots, X_{t_n})$, or $(X(t_1), \cdots, X(t_n))$ (for typographical convenience, we use both notations interchangeably, with or without ω: $X_t(\omega)$, or $X(t, \omega)$) is a random n-vector, with a distribution, $\mu(t_1, \cdots, t_n)$ say. The class of all such distributions as $\{t_1, \cdots, t_n\}$ ranges over all finite subsets of $[0, \infty)$ is called the class of all *finite-dimensional distributions* of X. These satisfy certain obvious consistency conditions:
(i) deletion of one point t_i can be obtained by 'integrating out the unwanted variable', as usual when passing from joint to marginal distributions,
(ii) permutation of the t_i permutes the arguments of the measure $\mu(t_1, \cdots, t_n)$ on $\mathbb{R}^n$.
Conversely, a collection of finite-dimensional distributions satisfying these two consistency conditions arises from a stochastic process in this way (this is the content of the DANIELL-KOLMOGOROV Theorem: P. J. Daniell in 1918, A. N. Kolmogorov in 1933).

Important though it is as a general existence result, however, the Daniell-Kolmogorov theorem does not take us very far. It gives a stochastic process X as a random function on $[0, \infty)$, i.e. a random variable on $\mathbb{R}^{[0,\infty)}$. This is a vast and unwieldy space; we shall usually be able to confine attention to much smaller and more manageable spaces, of functions satisfying regularity conditions. The most important of these is *continuity*: we want to be able to realise $X = (X_t(\omega))_{t \geq 0}$ as a random *continuous* function, i.e. a member of $C[0, \infty)$; such a process X is called *path-continuous* (since the map $t \to X_t(\omega)$ is called the sample path, or simply path, given by ω) – or more briefly, *continuous*. This is possible for the extremely important case of *Brownian motion*, for example, and its relatives. Sometimes we need to allow our random function $X_t(\omega)$ to have jumps. It is then customary, and convenient, to require X_t to be *right-continuous with left limits* (rcll), or càdlàg (continu à droite, limite à gauche) – i.e. to have X in the space $D[0, \infty)$ of all such functions (the *Skorohod space*). This is the case, for instance, for the *Poisson process* and its relatives.

General results on *realisability* – whether or not it is possible to *realise*, or obtain, a process so as to have its paths in a particular function space – are known, but it is usually better to *construct* the processes we need directly on the function space on which they naturally live.

Given a stochastic process X, it is sometimes possible to improve the regularity of its paths without changing its distribution (that is, without changing its finite-dimensional distributions). For background on results of this type (separability, measurability, versions, regularization, ...) see e.g. Doob's classic book [D].

The continuous-time theory is technically much harder than the discrete-time theory, for two reasons:
(i) questions of path-regularity arise in continuous time but not in discrete time,
(ii) *uncountable* operations (like taking sup over an interval) arise in continuous time. But measure theory is constructed using *countable* operations: uncountable operations risk losing measurability.

Filtrations and Insider Trading

Recall that a filtration models an information flow. In our context, this is the information flow on the basis of which market participants – traders, investors etc. – make their decisions, and commit their funds and effort. All this is information in the *public* domain – necessarily, as stock exchange prices are publicly quoted. Again necessarily, many people are involved in major business projects and decisions (an important example: mergers and acquisitions, or M&A) involving publicly quoted companies. Frequently, this involves price-sensitive information. People in this position are – rightly – prohibited by law from profiting by it directly, by trading on their own account, in publicly quoted stocks but using private information. This is rightly regarded as theft at the expense of the investing public.[2] Instead, those involved in M&A etc. should seek to benefit legitimately (and indirectly) – enhanced career prospects, commission or fees, bonuses etc.

The regulatory authorities (Securities and Exchange Commission – SEC – in US; Financial Conduct Authority (FCA) and Prudential Regulation Authority (PRA, part of the Bank of England (BoE) in UK) monitor all trading electronically. Their software alerts them to patterns of suspicious trades. The software design (necessarily secret, in view of its value to criminals) involves all the necessary elements of Mathematical Finance in exaggerated

[2] The plot of the film *Wall Street* revolves round such a case, and is based on real life – recommended!

form: economic and financial insight, plus: mathematics, probability and stochastic processes; statistics (especially pattern recognition, data mining and machine learning); numerics and computation.

§3. Classes of Processes.

1. *Martingales.*

The martingale property in continuous time is as in discrete time:

$$E[X_t|\mathcal{F}_s] = X_s \qquad (s < t),$$

and similarly for submartingales and supermartingales. There are regularization results, under which one can take X_t right-continuous in t. The convergence results, and UI mgs – important as they occur in RNVF – are similar. Among the contrasts: the Doob-Meyer decomposition, easy in discrete time (IV.8), is deep in continuous time. For background, see e.g.
MEYER, P.-A. (1966): *Probabilities and potentials*. Blaisdell
– and subsequent work by Meyer and the French school (Dellacherie & Meyer, *Probabilités et potentiel*, I-V, etc.)

2. Gaussian Processes.

Recall the multivariate normal distribution $N(\mu, \mathbf{\Sigma})$ in n dimensions. If $\mu \in \mathbb{R}^n$, $\mathbf{\Sigma}$ is a non-negative definite $n \times n$ matrix, $\mathbf{X}$ has distribution $N(\mu, \mathbf{\Sigma})$ if it has characteristic function

$$\phi_{\mathbf{X}}(\mathbf{t}) := E \exp\{i\mathbf{t}^T.\mathbf{X}\} = \exp\{i\mathbf{t}^T.\mu - \frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}\} \qquad (\mathbf{t} \in \mathbb{R}^n).$$

If further $\mathbf{\Sigma}$ is positive definite (so non-singular), $\mathbf{X}$ has density (*Edgeworth's Theorem* of 1893: F. Y. Edgeworth (1845-1926), English statistician)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{1}{2}n}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\}.$$

A process $X = (X_t)_{t \geq 0}$ is *Gaussian* if all its finite-dimensional distributions are Gaussian. Such a process can be specified by:
(i) a measurable function $\mu = \mu(t)$ with $EX_t = \mu(t)$,
(ii) a non-negative definite function $\sigma(s,t)$ with

$$\sigma(s,t) = cov(X_s, X_t).$$

Gaussian processes have many interesting properties. Among these, we quote *Belayev's dichotomy*: with probability one, the paths of a Gaussian process are either continuous, or extremely pathological: for example, unbounded above and below on any time-interval, however short. Naturally, we shall confine attention in this course to continuous Gaussian processes.

3. Markov Processes.

X is *Markov* if for each t, each $A \in \sigma(X_s : s > t)$ (the 'future') and $B \in \sigma(X_s : s < t)$ (the 'past'),

$$P(A|X_t, B) = P(A|X_t).$$

That is, if you know where you are (at time t), how you got there doesn't matter so far as predicting the future is concerned. Equivalently, past and future are conditionally independent given the present.

The same definition applied to Markov processes in discrete time.

X is said to be *strong Markov* if the above holds with the *fixed* time t replaced by a *stopping time* T (a random variable). This is a real restriction of the Markov property in continuous time (though not in discrete time) – another instance of the difference between the two.

4. Diffusions.

A *diffusion* is a path-continuous strong-Markov process such that for each time t and state x the following limits exist:

$$\mu(t,x) := \lim_{h\downarrow 0} \frac{1}{h} E[(X_{t+h} - X_t)|X_t = x],$$

$$\sigma^2(t,x) := \lim_{h\downarrow 0} \frac{1}{h} E[(X_{t+h} - X_t)^2|X_t = x].$$

Then $\mu(t,x)$ is called the *drift*, $\sigma^2(t,x)$ the *diffusion coefficient*. Then $p(t,x,y)$, the density of transitions from x to y in time t, satisfies the parabolic PDE

$$Lp = \partial p/\partial t, \qquad L := \frac{1}{2}\sigma^2 D^2 + \mu(x)D, \qquad D := \partial/\partial x.$$

The (2nd-order, linear) differential operator L is called the *generator*. *Brownian motion* is the case $\sigma = 1$, $\mu = 0$, and gives the *heat equation* ($L = \frac{1}{2}D^2$ in one dimension, half the *Laplacian* Δ in higher dimensions).

It is not at all obvious, but it is true, that this definition does indeed capture the nature of physical diffusion. Examples: heat diffusing through a metal; smoke diffusing through air; dye diffusing through liquid; pollutants diffusing through air or liquid.

§4. Quadratic Variation (QV) of Brownian Motion; Itô's Lemma

Recall that for $\xi\ N(\mu, \sigma^2)$, ξ has moment-generating function (MGF)

$$M(t) := E\exp\{t\xi\} = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}.$$

Take $\mu = 0$ below; for $\xi\ N(0, \sigma^2)$,

$$\begin{aligned} M(t) := E\exp\{t\xi\} &= \exp\{\frac{1}{2}\sigma^2 t^2\} \\ &= 1 + \frac{1}{2}\sigma^2 t^2 + \frac{1}{2!}(\frac{1}{2}\sigma^2 t^2)^2 + O(t^6) \\ &= 1 + \frac{1}{2!}\sigma^2 t^2 + \frac{3}{4!}\sigma^4 t^4 + O(t^6). \end{aligned}$$

So as the Taylor coefficients of the MGF are the moments (hence the name!),

$$E(\xi^2) = var\xi = \sigma^2, \qquad E(\xi^4) = 3\sigma^4, \qquad \text{so} \qquad var(\xi^2) = E(\xi^4) - [E(\xi^2)]^2 = 2\sigma^4.$$

For $B\ BM$, this gives in particular

$$EB_t = 0, \qquad varB_t = t, \qquad E[(B_t)^2] = t, \qquad var[(B_t)^2] = 2t^2.$$

In particular, for $t > 0$ *small*, this shows that the variance of B_t^2 is negligible compared with its expected value. Thus, the *randomness* in B_t^2 is negligible compared to its mean for t small.

This suggests that if we take a fine enough partition $\mathcal{P}$ of $[0, T]$ – a finite set of points

$$0 = t_0 < t_1 < \cdots < t_k = T$$

with $|\mathcal{P}| := \max|t_i - t_{i-1}|$ small enough – then writing

$$\Delta B(t_i) := B(t_i) - B(t_{i-1}), \qquad \Delta t_i := t_i - t_{i-1},$$

$\Sigma(\Delta B(t_i))^2$ will closely resemble $\Sigma E[(\Delta B(t_i)^2]$, which is $\Sigma\Delta t_i = \Sigma(t_i - t_{i-1}) = T$. This is in fact true a.s.:

$$\Sigma(\Delta B(t_i))^2 \to \Sigma\Delta t_i = T \qquad \text{as} \qquad \max|t_i - t_{i-1}| \to 0.$$

This limit is called the *quadratic variation* V_T^2 of B over $[0, T]$:

Theorem. The quadratic variation of a Brownian path over $[0, T]$ exists and equals T, a.s.

For details of the proof, see e.g. [BK], §5.3.2, SP L22, SA L7,8.

If we increase t by a small amount to $t + dt$, the increase in the QV can be written symbolically as $(dB_t)^2$, and the increase in t is dt. So, formally we may summarise the theorem as

$$(dB_t)^2 = dt.$$

Suppose now we look at the *ordinary* variation $\Sigma|\Delta B_t|$, rather than the *quadratic* variation $\Sigma(\Delta B_t)^2$. Then instead of $\Sigma(\Delta B_t)^2 \sim \Sigma\Delta t \sim t$, we get $\Sigma|\Delta B_t| \sim \Sigma\sqrt{\Delta t}$. Now for Δt small, $\sqrt{\Delta t}$ is of a larger order of magnitude that Δt. So if $\Sigma\Delta t = t$ converges, $\Sigma\sqrt{\Delta t}$ diverges to $+\infty$. This suggests – what is in fact true – the

Corollary. The paths of Brownian motion are of infinite variation - their variation is $+\infty$ on every interval, a.s.

The QV result above leads to Lévy's 1948 result, the Martingale Characterization of BM. Recall that B_t is a continuous martingale with respect to its natural filtration $(\mathcal{F}_t)$ and with QV t. There is a remarkable converse; we give two forms.

Theorem (Lévy; Martingale Characterization of Brownian Motion). If M is any continuous local $(\mathcal{F}_t)$-martingale with $M_0 = 0$ and quadratic variation t, then M is an $(\mathcal{F}_t)$-Brownian motion.

Theorem (Lévy). If M is any continuous $(\mathcal{F}_t)$-martingale with $M_0 = 0$ and $M_t^2 - t$ a martingale, then M is an $(\mathcal{F}_t)$-Brownian motion.

For proof, see e.g. [RW1], I.2. Observe that for $s < t$,

$$B_t^2 = [B_s + (B_t - B_s)]^2 = B_s^2 + 2B_s(B_t - B_s) + (B_t - B_s)^2,$$

$$E[B_t^2|\mathcal{F}_s] = B_s^2 + 2B_s E[(B_t - B_s)|\mathcal{F}_s] + E[(B_t - B_s)^2|\mathcal{F}_s] = B_s^2 + 0 + (t - s) :$$

$$E[B_t^2 - t|\mathcal{F}_s] = B_s^2 - s :$$

$B_t^2 - t$ is a martingale.

Quadratic Variation (QV).

The theory above extends to *continuous* martingales (bounded continuous martingales in general, but we work on a finite time-interval $[0, T]$, so continuity implies boundedness). We quote (for proof, see e.g. [RY], IV.1):

Theorem. A continuous martingale M is of finite quadratic variation $\langle M \rangle$, and $\langle M \rangle$ is the unique continuous increasing adapted process vanishing at zero with $M^2 - \langle M \rangle$ a martingale.

Corollary. A continuous martingale M has infinite variation.

Quadratic Covariation. We write $\langle M, M \rangle$ for $\langle M \rangle$, and extend $\langle \ \rangle$ to a bilinear form $\langle ., . \rangle$ with two different arguments by the *polarization identity*:

$$\langle M, N \rangle := \frac{1}{4}(\langle M + N, M + N \rangle - \langle M - N, M - N \rangle).$$

If N is of *finite* variation, $M \pm N$ has the same QV as M, so $\langle M, N \rangle = 0$.

Itô's Lemma.

We discuss Itô's Lemma in more detail in §6 below; we pause here to give the link with quadratic variation and covariation. We quote: if $f(t, x_1, \cdots, x_d)$ is C^1 in its zeroth (time) argument t and C^2 in its remaining d space arguments x_i, and $M = (M^1, \cdots, M^d)$ is a continuous vector martingale, then (writing f_i, f_{ij} for the first partial derivatives of f with respect to its ith argument and the second partial derivatives with respect to the ith and jth arguments) $f(M_t)$ has stochastic differential

$$df(M_t) = f_0(M)dt + \Sigma_{i=1}^d f_i(M_t) dM_t^i + \frac{1}{2} \Sigma_{i,j=1}^d f_{ij}(M_t) d\langle M^i, M^j \rangle_t.$$

Integration by Parts. If $f(t, x_1, x_2) = x_1 x_2$, we obtain

$$d(MN)_t = N dM_t + M dN_t + \frac{1}{2}\langle M, N \rangle_t.$$

Similarly for stochastic integrals (defined below): if $Z_i := \int H_i dM_i$ $(i = 1, 2)$, $d\langle Z_1, Z_2 \rangle = H_1 H_2 d\langle M_1, M_2 \rangle$.

Note. The integration-by-parts formula – a special case of Itô's Lemma, as above – is in fact *equivalent* to Itô's Lemma: either can be used to derive the other. Rogers & Williams [RW1, IV.32.4] describe the integration-by-parts formula/Itô's Lemma as 'the cornerstone of stochastic calculus'.

Fractals Everywhere.

As we saw, a Brownian path is a *fractal* – a *self-similar* object. So too is its zero-set Z. Fractals were studied, named and popularised by the French mathematician Benôit B. Mandelbrot (1924-2010). See his books, and Michael F. Barnsley: *Fractals everywhere*. Academic Press, 1988.

Fractals *look the same at all scales* – diametrically opposite to the familiar functions of Calculus. In Differential Calculus, a differentiable function has a tangent; this means that locally, its graph *looks straight*; similarly in Integral Calculus. While most continuous functions we encounter are differentiable, at least piecewise (i.e., except for 'kinks'), there is a sense in which the typical, or generic, continuous function is *nowhere differentiable*. Thus Brownian paths may look pathological at first sight – but in fact they are typical!

Hedging in continuous time.

Imagine hedging an option in continuous time. In discrete time, this involves repeatedly rebalancing our portfolio between cash and stock; in continuous time, this has to be done continuously. The relevant stochastic processes (Ch. VII) are *geometric Brownian motion (GBM)*, relatives of BM, which, like BM, have *infinite variation* (finite QV). This makes the rebalancing problematic – indeed, impossible in these terms. Analogy: a cyclist has to rebalance continuously, but does so smoothly, not with infinite variation! Or, think of continuous-time control of a manned space-craft (Kalman filter). In practice, hedging has to be done discretely (as in Ch. V). Or, we can use price processes with *jumps* – finite variation, but now the markets are incomplete, so prices are no longer unique.

§5. Stochastic Integrals (Itô Calculus)

Stochastic integration was introduced by K. ITÔ in 1944, hence its name Itô calculus. It gives a meaning to $\int_0^t X dY = \int_0^t X_s(\omega) dY_s(\omega)$, for suitable stochastic processes X and Y, the *integrand* and the *integrator*. We shall confine our attention here to the basic case with integrator Brownian motion: $Y = B$. Much greater generality is possible: for Y a continuous martingale, see [KS] or [RY]; for a systematic general treatment, see
MEYER, P.-A. (1976): Un cours sur les intégrales stochastiques. *Séminaire*

de Probabilités X: Lecture Notes on Math. **511**, 245-400, Springer.

The first thing to note is that stochastic integrals with respect to Brownian motion, *if* they exist, must be *quite different* from the measure-theoretic integral of III.2. For, the Lebesgue-Stieltjes integrals described there have as integrators the difference of two monotone (increasing) functions (by Jordan's theorem), which are locally of *finite (bounded) variation, FV*. But we know from §4 that Brownian motion is of *infinite (unbounded)* variation on every interval. So Lebesgue-Stieltjes and Itô integrals must be fundamentally different.

In view of the above, it is quite surprising that Itô integrals can be defined at all. But if we take for granted Itô's fundamental insight that they *can* be, it is obvious how to begin and clear enough how to proceed. We begin with the simplest possible integrands X, and extend successively much as we extended the measure-theoretic integral of Ch. III.

1. *Indicators.*

If $X_t(\omega) = I_{[a,b]}(t)$, there is exactly one plausible way to define $\int XdB$:

$$\int_0^t XdB, \quad \text{or} \quad \int_0^t X_s(\omega)dB_s(\omega), := \begin{cases} 0 & \text{if } t \leq a, \\ B_t - B_a & \text{if } a \leq t \leq b, \\ B_b - B_a & \text{if } t \geq b. \end{cases}$$

2. *Simple functions.* Extend by linearity: if X is a linear combination of indicators, $X = \Sigma c_i I_{[a_i,b_i]}$, we should define

$$\int_0^t XdB := \Sigma c_i \int_0^t I_{[a_i,b_i]}dB.$$

Already one wonders how to extend this from constants c_i to suitable random variables, and one seeks to simplify the obvious but clumsy three-line expressions above. It turns out that finite sums are not essential: one can have infinite sums, but now we take the c_i uniformly bounded.

We begin again, calling X *simple* if there is an infinite sequence

$$0 = t_0 < t_1 < \cdots < t_n < \cdots \to \infty$$

and uniformly bounded $\mathcal{F}_{t_n}$-measurable random variables ξ_n ($|\xi_n| \leq C$ for all n and ω, for some C) if $X_t(\omega)$ can be written in the form

$$X_t(\omega) = \xi_0(\omega)I_{\{0\}}(t) + \Sigma_{i=0}^{\infty}\xi_i(\omega)I_{(t_i,t_{i+1}]}(t) \qquad (0 \leq t < \infty, \omega \in \Omega).$$

The only definition of $\int_0^t XdB$ that agrees with the above for finite sums is, if n is the unique integer with $t_n \le t < t_{n+1}$,

$$\begin{aligned} I_t(X) := \int_0^t XdB &= \Sigma_0^{n-1} \xi_i (B(t_{i+1}) - B(t_i)) + \xi_n (B(t) - B(t_n)) \\ &= \Sigma_0^{\infty} \xi_i (B(t \wedge t_{i+1}) - B(t \wedge t_i)) \qquad (0 \le t < \infty). \end{aligned}$$

We note here some properties of the stochastic integral defined so far:
A. $I_0(X) = 0 \qquad P - a.s.$
B. *Linearity.* $I_t(aX + bY) = aI_t(X) + bI_t(Y)$.
Proof. Linear combinations of simple functions are simple.
C. $E[I_t(X)|\mathcal{F}_s] = I_s(X) \qquad P - a.s. \qquad (0 \le s < t < \infty)$:
$I_t(X) = \int_0^t XdB$ is a *continuous martingale.*
Proof. There are two cases to consider.
(i) Both s and t belong to the same interval $[t_n, t_{n+1})$. Then

$$I_t(X) = I_s(X) + \xi_n (B(t) - B(s)).$$

But ξ_n is $\mathcal{F}_{t_n}$-measurable, so $\mathcal{F}_s$-measurable ($t_n \le s$), so independent of $B(t) - B(s)$ (independent increments property of B). So

$$E[I_t(X)|\mathcal{F}_s] = I_s(X) + \xi_n E[B(t) - B(s)|\mathcal{F}_s] = I_s(X).$$

(ii) $s < t$ and t belong to different intervals: $s \in [t_m, t_{m+1})$ for $m < n$. Then

$$\begin{aligned} E[I_t(x)|\mathcal{F}_s] &= E(E[I_t(X)|\mathcal{F}_{t_n}]|\mathcal{F}_s) \qquad \text{(iterated conditional expectations)} \\ &= E(I_{t_n}(X)|\mathcal{F}_s), \end{aligned}$$

since ξ_n $\mathcal{F}_{t_n}$-measurable and independent increments of B give

$$E[\xi_n (B(t) - B(t_n))|\mathcal{F}_{t_n}] = \xi_n E[B(t) - B(t_n)|\mathcal{F}_{t_n}] = \xi_n .0 = 0.$$

Continuing in this way, we can reduce successively to t_{m+1}:

$$E[I_t(X)|\mathcal{F}_s] = E[I_{t_m}(X)|\mathcal{F}_s].$$

But $I_{t_m}(X) = I_s(X) + \xi_m (B(s) - B(t_m))$; taking $E[.|\mathcal{F}_s]$ the second term gives zero as above, giving the result. //

Note. The stochastic integral for simple integrands is essentially a martingale transform, and the above is essentially the proof of Ch. III that martingale transforms are martingales.

We pause to note a property of martingales which we shall need below. Call $X_t - X_s$ the *increment* of X over $(s,t]$. Then for a *martingale* X, *the product of the increments over disjoint intervals has zero mean.* For, if $s < t \leq u < v$,

$$\begin{aligned} E[(X_v - X_u)(X_t - X_s)] &= E[E[(X_v - X_u)(X_t - X_s)|\mathcal{F}_u]] \\ &= E[(X_t - X_s)E[(X_v - X_u)|\mathcal{F}_u]], \end{aligned}$$

taking out what is known (as $s, t \leq u$). The inner expectation is zero by the martingale property, so the LHS is zero, as required.

D (*Itô isometry*). $E[(I_t(X))^2]$, or $E[(\int_0^t X_s dB_s)^2]$, $= E[\int_0^t X_s^2 ds]$.
Proof. The LHS above is $E[I_t(X).I_t(X)]$, i.e.

$$E[(\Sigma_{i=0}^{n-1}\xi_i(B(t_{i+1}) - B(t_i)) + \xi_n(B(t) - B(t_n)))^2].$$

Expanding the square, the cross-terms have expectation zero by above, so

$$E[\Sigma_{i=0}^{n-1}\xi_i^2(B(t_{i+i} - B(t_i))^2 + \xi_n^2(B(t) - B(t_n))^2].$$

Since ξ_i is $\mathcal{F}_{t_i}$-measurable, each ξ_i^2-term is independent of the squared Brownian increment term following it, which has expectation $var(B(t_{i+1}) - B(t_i)) = t_{i+1} - t_i$. So we obtain

$$\Sigma_{i=0}^{n-1}E[\xi_i^2](t_{i+1} - t_i) + E[\xi_n^2](t - t_n).$$

This is $\int_0^t E[X_u^2]du = E[\int_0^t X_u^2 du]$, as required.

E. *Itô isometry (continued).* $I_t(X) - I_s(X) = \int_s^t X_u dB_u$ satisfies

$$E[(\int_s^t X_u dB_u)^2] = E[\int_s^t X_u^2 du] \qquad P - a.s.$$

Proof: as above.

F. *Quadratic variation.* The QV of $I_t(X) = \int_0^t X_u dB_u$ is $\int_0^t X_u^2 du$.

This is proved in the same way as the case $X \equiv 1$, that B has quadratic variation process t.

Integrands.

The properties above suggest that $\int_0^t XdB$ should be defined only for processes with

$$\int_0^t E[X_u^2]du < \infty \qquad \text{for all} \quad t.$$

We shall restrict attention to such X in what follows. This gives us an L_2-theory of stochastic integration (compare the L_2-spaces introduced in Ch. II), for which Hilbert-space methods are available.

3. *Approximation.*

Recall steps 1 (indicators) and 2 (simple integrands). By analogy with the integral of Ch. III, we seek a suitable class of integrands suitably approximable by simple integrands. It turns out that:
(i) The suitable class of integrands is the class of left-continuous adapted processes X with $\int_0^t E[X_u^2]du < \infty$ for all $t > 0$ (or all $t \in [0, T]$ with finite time-horizon T, as here),
(ii) Each such X may be approximated by a sequence of simple integrands X_n so that the stochastic integral $I_t(X) = \int_0^t XdB$ may be defined as the limit of $I_t(X_n) = \int_0^t X_n dB$,
(iii) The stochastic integral $\int_0^t XdB$ so defined still has properties A-F above.

It is not possible to include detailed proofs of these assertions in a course of this type [recall that we did not construct the measure-theoretic integral of Ch. III in detail either – and this is harder!]. The key technical ingredient needed is the *Kunita-Watanabe inequalities*. See e.g. [KS], §§3.1-2.

One can define stochastic integration in much greater generality.
1. *Integrands.* The natural class of integrands X to use here is the class of *predictable* processes. These include the left-continuous processes to which we confine ourselves above.
2. *Integrators.* One can construct a closely analogous theory for stochastic integrals with the Brownian integrator B above replaced by a *continuous local martingale* integrator M (or more generally by a *local martingale*: see below). The properties above hold, with D replaced by

$$E[(\int_0^t X_u dM_u)^2] = E[\int_0^t X_u^2 d\langle M\rangle_u].$$

See e.g. [KS], [RY] for details.

One can generalise further to *semimartingale* integrators: these are processes expressible as the sum of a local martingale and a process of (locally) finite variation. Now C is replaced by: stochastic integrals of local martingales are local martingales. See e.g. [RW1] or Meyer (1976) for details.

§6. Stochastic Differential Equations (SDEs) and Itô's Lemma

Suppose that U, V are adapted processes,with U locally integrable (so $\int_0^t U_s ds$ is defined as an ordinary integral, as in Ch. III), and V is left-continuous with $\int_0^t E[V_u^2]du < \infty$ for all t (so $\int_0^t V_s dB_s$ is defined as a stochastic integral, as in §5). Then

$$X_t := x_0 + \int_0^t U_s ds + \int_0^t V_s dB_s$$

defines a stochastic process X with $X_0 = x_0$. It is customary, and convenient, to express such an equation symbolically in differential form, in terms of the *stochastic differential equation*

$$dX_t = U_t dt + V_t dB_t, \qquad X_0 = x_0. \qquad (SDE)$$

Now suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is a function, continuously differentiable once in its first argument (which will denote time), and twice in its second argument (space): $f \in C^{1,2}$. The question arises of giving a meaning to the stochastic differential $df(t, X_t)$ of the process $f(t, X_t)$, and finding it.

Recall the Taylor expansion of a smooth function of several variables, $f(x_0, x_1, \cdots, x_d)$ say. We use suffices to denote partial derivatives: $f_i := \partial f/\partial x_i$, $\quad f_{i,j} := \partial^2 f/\partial x_i \partial x_j$ (recall that if partials not only exist but are continuous, then the order of partial differentiation can be changed: $f_{i,j} = f_{j,i}$, etc.). Then for $x = (x_0, x_1, \cdots, x_d)$ near u,

$$f(x) = f(u) + \Sigma_{i=0}^d (x_i - u_i) f_i(u) + \frac{1}{2} \Sigma_{i,j=0}^d (x_i - u_i)(x_j - u_j) f_{i,j}(u) + \cdots$$

In our case (writing t_0 in place of 0 for the starting time):

$$f(t, X_t) = f(t_0, X(t_0)) + (t - t_0) f_1(t_0, X(t_0)) + (X(t) - X(t_0)) f_2 + \frac{1}{2}(t - t_0)^2 f_{11} +$$

$$(t - t_0)(X(t) - X(t_0)) f_{12} + \frac{1}{2}(X(t) - X(t_0))^2 f_{22} + \cdots,$$

which may be written symbolically as

$$df(t, X(t)) = f_1 dt + f_2 dX + \frac{1}{2} f_{11} (dt)^2 + f_{12} dt dX + \frac{1}{2} f_{22} (dX)^2 + \cdots .$$

In this, we
(i) substitute $dX_t = U_t dt + V_t dB_t$ from above,
(ii) substitute $(dB_t)^2 = dt$, i.e. $|dB_t| = \sqrt{dt}$, from §4:

$$df = f_1 dt + f_2 (U dt + V dB) + \frac{1}{2} f_{11} (dt)^2 + f_{12} dt (U dt + V dB) + \frac{1}{2} f_{22} (U dt + V dB)^2 + \cdots$$

Now using $(dB)^2 = dt$,

$$\begin{aligned} (U dt + V dB)^2 &= V^2 dt + 2UV dt dB + U^2 (dt)^2 \\ &= V^2 dt + \text{higher-order terms} : \end{aligned}$$

$$df = (f_1 + U f_2 + \frac{1}{2} V^2 f_{22}) dt + V f_2 dB + \text{higher-order terms.}$$

Summarising, we obtain *Itô's Lemma*, the analogue for the Itô or stochastic calculus of the chain rule for ordinary (Newton-Leibniz) calculus:

Theorem (Itô's Lemma). If X_t has stochastic differential

$$dX_t = U_t dt + V_t dB_t, \qquad X_0 = x_0,$$

and $f \in C^{1,2}$, then $f = f(t, X_t)$ has stochastic differential

$$df = (f_1 + U f_2 + \frac{1}{2} V^2 f_{22}) dt + V f_2 dB_t.$$

That is, writing f_0 for $f(0, x_0)$, the initial value of f,

$$f(t, X_t)) = f_0 + \int_0^t (f_1 + U f_2 + \frac{1}{2} V^2 f_{22}) dt + \int_0^t V f_2 dB.$$

This important result may be summarised as follows: use Taylor's theorem formally, together with the rule

$$(dt)^2 = 0, \qquad dt dB = 0, \qquad (dB)^2 = dt.$$

Itô's Lemma extends to higher dimensions, as does the rule above:

$$df = (f_0 + \Sigma_{i=1}^d U_i f_i + \frac{1}{2} \Sigma_1^d V_i^2 f_{ii}) dt + \Sigma_1^d V_i f_i dB_i$$

(where U_i, V_i, B_i denote the ith coordinates of vectors U, V, B, f_i, f_{ii} denote partials as above); here the formal rule is

$$(dt)^2 = 0, \qquad dtdB_i = 0, \qquad (dB_i)^2 = dt, \qquad dB_i dB_j = 0 \qquad (i \neq j).$$

Corollary. $E[f(t, X_t)] = f_0 + \int_0^t E[f_1 + Uf_2 + \frac{1}{2}V^2 f_{22}]dt$.

Proof. $\int_0^t V f_2 dB$ is a stochastic integral, so a martingale, so its expectation is constant ($= 0$, as it starts at 0). //

Note. Powerful as it is in the setting above, Itô's Lemma really comes into its own in the more general setting of semimartingales. It says there that if X is a semimartingale and f is a smooth function as above, then $f(t, X(t))$ is also a semimartingale. The ordinary differential dt gives rise to the bounded-variation part, the stochastic differential gives rise to the martingale part. This closure property under very general non-linear operations is very powerful and important.

Example: The Ornstein-Uhlenbeck Process.

The most important example of a SDE for us is that for geometric Brownian motion (VII.1 below). We close here with another example.

Consider now a model of the velocity V_t of a particle at time t ($V_0 = v_0$), moving through a fluid or gas, which exerts
(i) a frictional drag, assumed propertional to the velocity,
(ii) a noise term resulting from the random bombardment of the particle by the molecules of the surrounding fluid or gas. The basic model is the SDE

$$dV = -\beta V dt + c dB, \qquad (OU)$$

whose solution is called the *Ornstein-Uhlenbeck* (velocity) process with *relaxation time* $1/\beta$ and *diffusion coefficient* $D := \frac{1}{2}c^2/\beta^2$. It is a stationary Gaussian Markov process (not stationary-increments Gaussian Markov like Brownian motion), whose limiting (ergodic) distribution is $N(0, \beta D)$ (the *Maxwell-Boltzmann distribution* of Statistical Mechanics) and whose limiting correlation function is $e^{-\beta|\cdot|}$.

If we integrate the OU velocity process to get the OU *displacement process*, we lose the Markov property (though the process is still Gaussian). Being non-Markov, the resulting process is much more difficult to analyse.

The OU process is the prototype of processes exhibiting *mean reversion*,

or a *central push*: frictional drag acts as a restoring force tending to push the process back towards its mean. It is important in many areas, including
(i) statistical mechanics, where it originated,
(ii) mathematical finance, where it appears in the *Vasicek model* for the term-structure of interest-rates (the mean represents the 'natural' interest rate),
(iii) *stochastic volatility* models, where the volatility σ itself is now a stochastic process σ_t, subject to an SDE of OU type.

Theory of interest rates.

This subject dominates the mathematics of *money markets*, or *bond markets*. These are more important in today's world than stock markets, but are more complicated, so we must be brief here. The area is crucially important in *macro-economic policy*, and in political decision-making, particularly after the financial crisis ("credit crunch"). Government policy is driven by fear of speculators in the bond markets (rather than aimed at inter-governmental cooperation against them). The mathematics is infinite-dimensional (at each time-point t we have a whole *yield curve* over future times), but reduces to finite-dimensionality: bonds are only offered at discrete times, with a *tenor* structure (a finite set of maturity times).

Mean reversion is used in models, to reflect the underlying 'natural interest rate', from which deviations may occur due to short-term pressures.

Note.

The 'short-term pressures' arising from the Crash or Credit Crunch of 2007-8 and on have now lasted a decade! Interest rates have been historically low (to the benefit of borrowers such as mortgage-holders, and the detriment of savers, for example). In the last days of September 2017, the Governor of the Bank of England, Mark Carney, said that bank rate may well rise (we shall see – the decision is taken by the Monetary Policy Committee, on which the Governor has one vote out of nine). You may be interested to compare this with the actions of the Fed in recent years. etc.