sphandout.tex

STOCHASTIC PROCESSES: HANDOUT, December 2010

The following material is not examinable. Think of it as (roughly) what I would have included had I had 11 weeks rather than 10. I include it here as
(a) I have it to hand;
(b) you may find it useful as back-up to the course as taught, and/or background to other courses, or the MSc programme as a whole.

1. Likelihood Estimation for Diffusions [Lecture 20] (reference: Bingham & Kiesel, 5.9).

We consider diffusions

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad (*)$$

with W_t a standard Brownian motion, and $\theta \in \Theta$ with Θ a compact parameter space. We require the diffusions to be time-homogeneous and stationary and work under the following additional assumptions.

Assumption 1.
For all $\theta \in \Theta$ there exists a strong solution of $(*)$. This may be guaranteed by imposing Lipschitz and linear growth bounds:

$$\begin{aligned}|\mu(x;\theta) - \mu(y;\theta)| &\le K|x-y|,\\ |\sigma(x;\theta) - \sigma(y;\theta)| &\le K|x-y|,\\ |\mu(x;\theta)|^2 &\le K^2(1+|x|^2),\\ |\sigma(x;\theta)|^2 &\le K^2(1+|x|^2),\end{aligned}$$

where K is some fixed constant independent of θ.

Assumption 2.
Θ is compact and the true parameter value θ_0 is an element of Θ, i.e. $\theta \in \Theta$.

Given observations $X_0, X_{t_1}, \dots, X_{t_n}$ (which we may assume to be equidistant, i.e. to be observations at $n+1$ equidistant time points $t_0 = 0, t_1, \dots, t_n$, with $\Delta_t^n = t_k - t_{k-1}$) the likelihood function of the discrete-time data is

$$\begin{aligned} l_n(\theta) &= p(X_0, X_{t_1}, \dots, X_{t_n};\theta)\\ &= p(X_0, t_0;\theta)\prod_{k=0}^{n-1} p(X_{t_k}, t_k; X_{t_{k+1}}, t_{k+1};\theta)\\ &= p(X_0, t_0;\theta)\prod_{k=0}^{n-1} p(X_{t_k}, \Delta_t^n; X_{t_{k+1}};\theta),\end{aligned}$$

where $p(.;\theta)$ denotes the joint density; the first equality follows from the Markovian nature, and the second equality from the assumption that we have a time-homogeneous process.

We shall be concerned with the log - likelihood function $\ell_n(\theta)$,

$$\ell_n(\theta) = \log l_n(\theta) = \log p(X_0, t_0; \theta) + \sum_{k=0}^{n-1} \log p(X_{t_k}, \Delta_t^n; X_{t_k+1}; \theta).$$

Assumption 3.
The likelihood function ℓ_n is twice continuously differentiable in θ in the interior of Θ. Furthermore the class of random matrices so obtained

$$\left\{ \left[\frac{\partial^2 \ell_n(\theta)}{\partial \theta^2} \right], \theta \in \Theta^o \right\}$$

is uniformly bounded.
So the observed information $J_n(\theta) = -\left[\frac{\partial^2 \ell_n(\theta)}{\partial \theta^2}\right]$ exists.
Assumption 4.
The *expected information matrix*

$$I_n(\theta) = E_\theta \left(\left[\frac{\partial \ell_n(\theta)}{\partial \theta} \right] \left[\frac{\partial \ell_n(\theta)}{\partial \theta} \right]' \right)$$

has full rank and is uniformly bounded for $\theta \in \Theta$.
Assumption 5.
For every vector $\lambda \in \mathbf{R}^d$ we have that the quadratic form $\lambda' I_n(\theta) \lambda \to \infty$.
These assumption are needed to assure that the maximum likelihood estimator $\hat{\theta}$ can be obtained from the (true) likelihood function and is consistent and asymptotically normal. We quote:

Theorem. Under assumptions 1–5 the maximum likelihood estimator $\hat{\theta}_n$ exists and is consistent and asymptotically normal (CAN), i.e. with θ_0 the true parameter

$$\hat{\theta}_n \to \theta_0 \ (n \to \infty) \ \text{ in probability,}$$

and

$$I_n(\theta_0)^{\frac{1}{2}} (\hat{\theta}_n - \theta_0) \to N \ (n \to \infty) \text{ in distribution, } \ N \sim N(0,1).$$

Indirect Inference. A naive approach to the above estimation problem is to use a discrete Euler approximation:

$$x_t^{(\Delta)} = x_{k\Delta}^{(\Delta)} \qquad \text{for} \qquad k\Delta \le t \le (k+1)\Delta,$$

where

$$x^{(\Delta)}_{(k+1)\Delta} = x^{(\Delta)}_{k\Delta} + \Delta\mu(x^{(\Delta)}_{k\Delta};\theta) + \sigma(x^{(\Delta)}_{k\Delta};\theta)\sqrt{\Delta}\epsilon^{(\Delta)}_k$$

with $\epsilon^{(\Delta)}_k$ Gaussian white noise. We quote that the process $x^{(\Delta)}_t$ converges weakly to $X(t)$ for $\Delta \to 0$ at a sufficient rate. Thus a naive approach to estimate θ would be to use an approximation of Euler type (usually $\Delta = 1$ is chosen) and compute the maximum likelihood estimator. However, it is known that such an estimator is inconsistent.

2. Brownian Motion in Stochastic Modeling [L23] (ref.: B&K, 5.3.4).

To begin at the beginning: Brownian motion is named after Robert Brown (1773–1858), the Scottish botanist who in 1828 observed the irregular and haphazard – apparently random – motion of pollen particles suspended in water. Similar phenomena are observed in gases – witness the familiar sight of dust particles dancing in sunbeams. During the 19th C., it became suspected that the explanation was that the particles were being bombarded by the molecules in the surrounding medium – water or air. Note that this picture requires *three* different scales: microscopic (water or air molecules), mesoscopic (pollen or dust particles) and macroscopic (you, the observer). These ideas entered the kinetic theory of gases, and statistical mechanics, through the pioneering work of Maxwell, Gibbs and Boltzmann. However, some scientists still doubted the existence of atoms and molecules (not then observable directly). Enter the birth of the quantum age in 1900 with the quantum hypothesis of Max Planck (1858–1947).

Louis Bachelier (1870–1946) introduced Brownian motion into the field of economics and finance in his thesis *Théorie de la spéculation* of 1900. His work lay dormant until much later; we will pick up its influence on Itô, Samuelson, Merton and others below.

Albert Einstein (1879–1955), in his work of 1905, attacked the problem of demonstrating the existence of molecules, and for good measure estimating *Avogadro's number* (c. 6.02×10^{23}) experimentally. Einstein realized that what was informative was the *mean square displacement* of the Brownian particle – its *diffusion coefficient*, in our terms. This is proportional to time, and the constant D of proportionality,

$$var\ W_t = Dt,$$

is informative about Avogadro's number (which, roughly, gives the scale-factor in going from the microscopic to the macroscopic scale). This *Ein-*

stein relation is the prototype of a class of results now known in statistical mechanics as *fluctuation-dissipation theorems*.

All this was done without any proper mathematical underpinning. This was provided by Wiener in 1923, as mentioned earlier.

Quantum mechanics emerged in 1925–28 with the work of Heisenberg, Schrodinger and Dirac, and with the 'Copenhagen interpretation' of Bohr, Born and others, it became clear that the quantum picture is both inescapable at the subatomic level and intrinsically probabilistic. The work of Richard P. Feynman (1918–1988) in the late 1940s on quantum electrodynamics (QED), and his approach to quantum mechanics via 'path integrals', introduced Wiener measure squarely into quantum theory. Feynman's work on quantum mechanics was made mathematically rigorous by Mark Kac (1914–1984) (QED is still problematic!); the Feynman-Kac formula (giving a stochastic representation for the solutions of certain PDEs) stems from this.

Subsequent developments involve *Itô calculus*, and we shall consider them in Ch. IV below. Suffice it to say here that Itô's work of 1944 picked up where Bachelier left off, and created the machinery needed to use Brownian motion to model stock prices successfully (note: stock prices are nonnegative – positive, until the firm goes bankrupt – while Brownian motion changes sign, indeed has lots of sign changes, as we saw above when discussing its zero-set Z). The economist Paul SAMUELSON (1915-2009) in 1965 advocated the Itô model – geometric Brownian motion – for financial modelling. Then in 1973 Black and Scholes gave their famous formula, and the same year Merton derived it by Itô calculus. Today Itô calculus is a fundamental tool in stochastic modeling generally, and the modelling of financial markets in particular.

In sum: wherever we look – statistical mechanics, quantum theory, economics, finance – we see a random world, in which much that we observe is driven by random noise, or random fluctuations. Brownian motion gives us an invaluable model for describing these, in a wide variety of settings. This is statistically natural. The ubiquitous nature of Brownian motion is the dynamic counterpart of the ubiquitous nature of the normal distribution. This rests ultimately on the Central Limit Theorem (CLT – II.9, L12) – known to physicists as the Law of Errors – and is, fundamentally, why statistics works.

Economic Interpretation. Suppose X is used as a driving noise process in a financial market model for asset prices (example: $X = BM$ in the Black-Scholes-Merton model). If prices move continuously, the Brownian model is appropriate: among Lévy processes, only Brownian motions have continuous paths ($\mu = 0$, so there are no jumps). If prices move by intermittent jumps, a compound Poisson (FA) model is appropriate – but this is more suitable for modelling economic shocks, or the effects of big transactions. For the more common case of the everyday movement of traded stocks under the competitive effects of supply and demand, numerous small trades predominate, economic agents are price takers and not price makers, and a model with infinite activity (IA) is appropriate.

There is a parallel between the financial situation above – the IA case (lots of small traders) as a limiting case of the FA case (a few large ones) and the applied probability areas of queues and dams. Think of work arriving from the point of view of you, the server. It arrives in large discrete chunks, one with each arriving customer. As long as there is work to be done, you work non-stop to clear it; when no one is there, you are idle. The limiting situation is that of a dam. Raindrops may be discrete, but one can ignore this from the water-engineering viewpoint. When water is present in the dam, it flows out through the outlet pipe at constant rate (unit rate, say); when the dam is empty, nothing is there to flow out.

3. Wavelets [L23] (ref.: B&K, p.166, 5.3.1).

The Haar system (H_n), and the Schauder system (Δ_n) obtained by integration from it, are examples of *wavelet systems*. The original function, H or Δ, is a *mother wavelet*, and the 'daughter wavelets' are obtained from it by dilation and translation. The expansion of the theorem is the *wavelet expansion of BM* with respect to the Schauder system (Δ_n). For any $f \in C[0,1]$, we can form its wavelet expansion

$$f(t) = \sum_{n=0}^{\infty} c_n \Delta_n(t),$$

with *wavelet coefficients* c_n. Here c_n are given by

$$c_n = f\left(\frac{k+\frac{1}{2}}{2^j}\right) - \frac{1}{2}\left[f\left(\frac{k}{2^j}\right) + f\left(\frac{k+1}{2^j}\right)\right].$$

This is the form that gives the $\Delta_n(.)$ term its correct triangular influence, localized on the dyadic interval $[k/2^j, (k+1)/2^j]$. Thus for f BM, $c_n = \lambda_n Z_n$, with λ_n, Z_n as above. The wavelet construction of BM above is, in modern language, the classical 'broken-line' construction of BM due to Lévy in his book of 1948 - the *Lévy representation* of BM using the Schauder system, and extended to general cons by Cieselski in 1961. The earliest expansion of BM – 'Fourier-Wiener expansion' – used the trigonometric cons (Paley and Zygmund 1930–32, Paley, Wiener and Zygmund 1932).

Note. We shall see that Brownian motion is a *fractal*, and wavelets are a useful tool for the analysis of fractals more generally.

From the mathematical point of view, Brownian motion owes much 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 etc. From an applied point of view, as its diverse origins – Brown's work in botany, Bachelier's in economics, Einstein's in statistical mechanics etc. – suggest, Brownian motion has a universal character, and is ubiquitous both in theory and in applied modeling. The universal nature of Brownian motion as a stochastic process is simply the dynamic counterpart – where we work with evolution in time – of the universal nature of its static counterpart, the normal (or Gaussian) distribution – in probability, statistics, science, economics etc. Both arise from the same source, the central limit theorem. This says that when we average large numbers of independent and comparable objects, we obtain the normal distribution in a static context, or Brownian motion in a dynamic context. What the central limit theorem really says is that, when what we observe is the result of a very large number of individually very small influences, the normal distribution or Brownian motion will inevitably and automatically emerge. This explains the central role of the normal distribution in statistics – basically, this is why statistics works. It also explains the central role of Brownian motion as the basic model of random fluctuations, or *random noise* as one often says. As the word noise suggests, this usage comes from electrical engineering and the early days of radio. When we come to studying the dynamics of stochastic processes by means of stochastic differential equations (Ch. IV), we will usually find a 'driving noise' term. The most basic driving noise process is Brownian motion; its role is to represent the 'random buffeting' of the object under study by a myriad of influences which we have no hope of studying in detail – and indeed, no need to. By using the central limit theorem, we make the very complexity of the situation work on our side: Brownian motion is a comparatively simple and tractable process to work with – vastly simpler than the underlying random buffeting whose effect it approximates and represents.

The precise circumstances in which one obtains the normal or Gaussian distribution, or Brownian motion, have been much studied (this was the predominant theme in Lévy's life's work, for instance). One needs means and variances to exist (which is why the mean μ and the variance σ^2 are needed to parametrize the normal or Gaussian family). One also needs either independence, or something not too far removed from it, such as suitable martingale dependence or Markov dependence.

4. Zero set of BM [L23] (ref.: B&K p.172-3, 5.3.3).

We write Z for the zero set of BM: Z is the random time-set where BM vanishes, $Z := \{t : B_t = 0\}$.

Using time-inversion, we see that – as the zero-set of Brownian motion $Z := \{t \geq 0 : W_t = 0\}$ is unbounded (contains infinitely many points increasing to infinity), it must also contain infinitely many points decreasing to zero. That is, any zero of Brownian motion (e.g., time $t = 0$, as we are choosing to start our BM at the origin) produces an 'echo' – an infinite sequence of zeros at *positive* times decreasing to zero. How can we hope to graph such a function? (We can't!) How on earth does it manage to *escape* from zero, when hitting zero at one time, u say, forces zero to be hit infinitely many times in any time-interval $[u, u+\epsilon]$ $(\epsilon > 0)$? The answer to these questions involves *excursion theory*, one of Itô's great contributions to probability theory (1970). When BM is *at* zero, it is as likely to leave to the right as to the left, by symmetry – but it will leave, immediately, with probability one. These 'excursions away from zero' – above and below – happen according to a *Poisson random measure* governing the excursions – the *excursion measure* – on path-space. As there are infinitely many excursions in finite time-intervals, the excursion measure has *infinite mass* – it is σ-finite but not finite. We will not pursue excursion theory here. Note however that, far from being pathological as one might at first imagine, the behaviour described above is what one expects of a normal, well-behaved process: the technical term is '$\{0\}$ is regular for 0', and 'regular' is used to describe good, not bad, behaviour.

Since Brownian motion has continuous paths, its zero-set Z is *closed*. Since each zero is, by above, a limit-point of zeros, Z is a *perfect* set. The zero-set is also *uncountable* ('big', in one sense), but Lebesgue-null – has Lebesgue measure zero ('small', in another sense). The machinery for measuring the size of small sets such as Z is that of *Hausdorff measures*. The

Hausdorff measure properties of Z have been studied in great detail. The zero-set Z has a fractal structure, which it inherits from that of W under Brownian scaling. The natural machinery for studying the fine detail of the structure of fractals is, as above, that of Hausdorff measures.

Note. What Constitutes Pathological Behaviour? Weierstrass, and several other analysts of the 19th C., constructed examples of functions which were continuous but nowhere differentiable. These were long regarded as interesting but pathological. Similarly for the paths of Brownian motion. This used to be regarded as very interesting mathematically, but of limited relevance to modelling the real world. Then – following the work of B. B. Mandelbrot (plus computer graphics, etc.) – *fractals* attracted huge attention. It was then realized that such properties were typical of fractals, and so – as we now see fractals everywhere (to quote the title of Barnsley's book) – ubiquitous rather than pathological.

The situation with Lévy paths of infinite activity is somewhat analogous. Because one cannot draw them (or even visualise them, perhaps), they used to be regarded as mathematically interesting but clearly idealised so far as modelling of the real world goes. The above economic/financial interpretation has changed all this. 'Lévy finance' is very much alive at the moment. Moral: one never quite knows when this sort of thing is going to happen in mathematics!

5. Stability [L24] (ref.: b&K p.180-181, 5.5.1).

Suppose we now restrict to *identical distribution* as well as independence in SD above. That is, we seek the class of limit laws of *random walks* $S_n = \sum_1^n X_k$ with (X_n) iid – after an affine transformation (centering and scaling) – that is, for all limit laws of $(S_n - a_n)/b_n$. It turns out that the class of limit laws so obtained is the same as the class of laws for which S_n has the same *type* as X_1 – i.e. the same law to within an affine transformation, or a change of location and scale. Thus the type is 'stable' (invariant, unchanged) under addition of independent copies, whence such laws are called *stable.* They form the class S:

$$S \subset SD \subset I.$$

It turns out that this class of stable laws can be described explicitly by parameters – four in all, of which two (location and scale, specifying the law within the type) are of minor importance, leaving two essential parameters, called the *index* α ($\alpha \in (0, 2]$) and the *skewness parameter* β ($\beta \in [-1, 1]$). To within type, the Lévy exponent is

$$\Psi(u) = |u|^{\alpha}(1 - i\beta sgn(u) \tan \frac{1}{2}\pi\alpha)$$

for $\alpha \neq 1$ ($0 < \alpha < 1$ or $1 < \alpha \leq 2$) and

$$\Psi(u) = |u|(1 + i\beta sgn(u) \log |u|)$$

if $\alpha = 1$. The Lévy measure is absolutely continuous, with density of the form

$$\mu(dx) = \begin{cases} c_+ dx/x^{1+\alpha} & x > 0, \\ c_- dx/|x|^{1+\alpha} & x < 0, \end{cases}$$

with $c_+, c_- \geq 0$ and

$$\beta = (c_+ - c_-)/(c_+ + c_-).$$

The case $\alpha = 2$ (for which β drops out) gives the normal/Gaussian case, already familiar. The case $\alpha = 1$ and $\beta = 0$ gives the (symmetric) Cauchy law above. The case $\alpha = 1$, $\beta \neq 0$ gives the asymmetric Cauchy case, which is awkward, and we shall not pursue it.

From the form of the Lévy exponents of the remaining stable CFs (where the argument u appears only in $|u|^{\alpha}$ and $sgn(u)$), we see that, if $S_n = X_1 + \ldots + X_n$ with X_i independent copies,

$$S_n/n^{1/\alpha} = X_1 \text{ in distribution } (n = 1, 2, \ldots).$$

This is called the *scaling property* of the stable laws; those (all except the asymmetric Cauchy) that possess it are called *strictly stable.*

The stable densities do not have explicit closed forms in general, only series expansions. The normal and (symmetric) Cauchy densities are known (above), as is one further important special case:

Lévy's density. Here $\alpha = 1/2, \beta = +1$. One can check that for each a,

$$f(x) = \frac{a}{\sqrt{2\pi x^3}} \exp\left\{-\frac{1}{2}a^2/x\right\} = \frac{a}{x^{3/2}}\phi(a/\sqrt{x}) \qquad (x > 0)$$

has Laplace transform $\exp\{-a\sqrt{2s}\}$ ($s \geq 0$). This is the density of the first-passage time of Brownian motion over a level $a > 0$.

Holtsmark density. The other remarkable case is that of $\alpha = 3/2$, $\beta = 0$, studied by the Danish astronomer J. Holtsmark in 1919 in connection with the gravitational field of stars – this before Lévy's work on stability. The power 3/2 comes from 3 dimensions and the inverse square law of gravity.

6. Martingale Transforms [Problems 8, after Lecture 24] (ref.: B&K 3.4).

Now think of a gambling game, or series of speculative investments, in discrete time. There is no play at time 0; there are plays at times $n = 1, 2, \ldots,$ and

$$\Delta X_n := X_n - X_{n-1}$$

represents our net winnings per unit stake at play n. Thus if X_n is a martingale, the game is 'fair on average'.

Call a process $C = (C_n)_{n=1}^{\infty}$ *predictable* if C_n is $\mathcal{F}_{\setminus-\infty}$-measurable for all $n \geq 1$. Think of C_n as your stake on play n (C_0 is not defined, as there is no play at time 0). Predictability says that you have to decide how much to stake on play n based on the history *before* time n (i.e., up to and including play $n-1$). Your winnings on game n are $C_n \Delta X_n = C_n(X_n - X_{n-1})$. Your total (net) winnings up to time n are

$$Y_n = \sum_{k=1}^{n} C_k \Delta X_k = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

We write

$$Y = C \bullet X, \qquad Y_n = (C \bullet X)_n, \qquad \Delta Y_n = C_n \Delta X_n$$

($(C \bullet X)_0 = 0$ as $\sum_{k=1}^{0}$ is empty), and call $C \bullet X$ the *martingale transform* of X by C.

Theorem (Martingale Transform Theorem).
(i) If C is a bounded non-negative predictable process and X is a supermartingale, $C \bullet X$ is a supermartingale null at zero.
(ii) If C is bounded and predictable and X is a martingale, $C \bullet X$ is a martingale null at zero.

Proof. $Y = C \bullet X$ is integrable, since C is bounded and X integrable. Now

$$\begin{aligned} E[Y_n - Y_{n-1} | \mathcal{F}_{n-1}] &= E[C_n (X_n - X_{n-1}) | \mathcal{F}_{n-1}] \\ &= C_n E[(X_n - X_{n-1}) | \mathcal{F}_{n-1}] \end{aligned}$$

(as C_n is bounded, so integrable, and $\mathcal{F}_{n-1}$-measurable, so can be taken out)

$$\leq 0$$

in case (i), as $C \geq 0$ and X is a supermartingale,

$$= 0$$

in case (ii), as X is a martingale. //

Interpretation. You can't beat the system! In the martingale case, predictability of C means we can't foresee the future (which is realistic and fair). So we expect to gain nothing – as we should.
Note. Martingale transforms are the discrete analogues of stochastic integrals. They dominate the mathematical theory of finance in discrete time, just as stochastic integrals (Ch. IV) dominate the theory in continuous time.

The old-fashioned term for ordinary calculus is infinitesimal calculus. Martingale transforms belong to the probabilistic elaboration of the discrete analogue of this, the calculus of finite differences. The passage from discrete to continuous is written formally as

$$\Delta Y_n = C_n \Delta X_n \to dY_t = C_t dX_t,$$

where the dY, dX on the right are stochastic differentials.

7. SDE for Geometric Brownian Motion (GBM) [28] (ref.: B&K 5.6.3).

Suppose we wish to model the time evolution of a stock price $S(t)$ (as we will, in Black-Scholes theory). Consider how S will change in some small time-interval from the present time t to a time $t + dt$ in the near future. Writing $dS(t)$ for the change $S(t + dt) - S(t)$ in S, the *return* on S in this interval is $dS(t)/S(t)$. It is economically reasonable to expect this return to decompose into two components, a *systematic* part and a *random* part. The systematic part could plausibly be modelled by μdt, where μ is some parameter representing the mean rate of return of the stock. The random part could plausibly be modelled by $\sigma dW(t)$, where $dW(t)$ represents the noise term driving the stock price dynamics, and σ is a second parameter describing how much effect this noise has – how much the stock price fluctuates. Thus σ governs how volatile the price is, and is called the *volatility* of the stock. The role of the driving noise term is to represent the random buffeting effect of the multiplicity of factors at work in the economic environment in which the stock price is determined by supply and demand.

The most successful single branch of mathematical or scientific knowledge we have is the calculus, dating from Newton and Leibniz in the 17th century, and the resulting theories of differential equations, ordinary and partial (ODEs and PDEs). With any differential equation, the two most basic questions are those of existence and uniqueness of solutions – and to formulate such questions precisely, one has to specify what one means by a solution. For example, for PDEs, 19th century work required solutions in terms of ordinary functions – the only concept available at that time. More modern work has available the concept of generalised functions or distributions (in the sense of Laurent Schwartz (1915–2002)). It has been found that a much cleaner and more coherent theory of PDEs can be obtained if one is willing to admit such generalised functions as solutions. Furthermore, to obtain existence and uniqueness results, one has to impose reasonable regularity conditions on the coefficients occurring in the differential equation.

Numerics. Most differential equations (ordinary or partial) do not have solutions in closed form, and so have to be solved numerically in practice. This is all the more true in the more complicated setting of stochastic differential equations, where it is the exception rather than the rule for there to be an explicit solution. One is thus reliant in practice on numerical methods of solution, and here a great deal is known. We must refer elsewhere for details.

Note. Stochastic calculus with 'anticipating' integrands, 'backward' stochastic integrals, etc., have been developed, and are useful (e.g., in more advanced areas such as the *Malliavin calculus*. But let us learn to walk before we learn to run.

NHB, 15.12.2010.