Introduction

This text arose from the author's drive to understand martingales other than the Wiener process (Einsteinean Brownian motion), to find nontrivial examples of martingales, and to understand the dynamics of nontrivial martingales with finite variance. This takes us to Ito processes and Fokker–Planck partial differential equations, ruling out Levy processes.

Because martingales provide our main focus, Ito calculus and the Fokker–Planck partial differential equation are developed parallel as methods of analysis. A drift-free Ito process describes a martingale, for example Ito processes can be used to solve some interesting partial differential equations via martingale constructions, and martingales are fundamental in finance theory. Mathematicians use Ito calculus while physicists use the Fokker–Planck equations. We marry the two and provide a bridge between the two different dialects of the same language, Ito calculus transforms 1–1 with Fokker–Planck under coordinate transformations. We combine both methods to achieve new and more powerful results that would have been difficult had we been confined to one description or the other.

Mathematics books on stochastic calculus and differential equations prove powerful theorems about Ito processes and martingales but provide no nontrivial examples. Nontrivial examples from finance and trivial ones from statistical physics are provided in this text. Physics texts on stochastic processes generally do not develop Ito calculus and thereby provide inadequate preparation for understanding finance theory (see Hänggi and Thomas, 1982, for a very rare exception). Indeed, it was Harry Thomas who made me realize that I had to learn to understand and apply Ito calculus in 2002. Once I learned it, I found it so simple, direct, and useful that I would never choose to do without it.

Another problem of communication is the different notations used in physics and mathematics. We use physics notation and provide next a translation to the mathematicians' notation. We also provide the reader with a list of definitions where the notation is defined.

I use Dirac notation $\langle x(t)\rangle$ to denote the ensemble average of a random variable x at time t, whereas mathematicians would write E(x). A conditional expectation, conditioned on having observed x_o at time t_o, is denoted in our notation by $\langle x(t)\rangle_c$, whereas mathematicians typically write $E(x|x_o)$. Conditional expectations, along with martingales, are introduced in Chapter 2 and are used throughout the text. In defining certain quantities in advanced stochastic integration (see Chapter 10), mathematicians use a notation that looks like our bracket notation for averages but isn't. A guide to our and their notation for stochastic integration is provided in Chapter 10. Indeed, the main obstacle initially preventing me from understanding Durrett's generalization of stochastic integration was his notation, which I had to decipher from his text because it was introduced without a word of explanation. Once I understood the notation, I found Durrett's two books (Durrett, 1984; 1996) to be invaluable. A topic of strong interest to mathematicians in the study of martingales is the stopping time for a process. Stopping times are introduced and used in Chapters 3 and 10 in connection with martingales but are called "first passage times," which is the common phrase in physics and also in the older literature on stochastic processes. I hope that these words will help mathematicians in following the book. They're the ones who need it the most in order to understand why their reliance on formal theorems, combined with ignorance of the underlying conditions needed for the application of a stochastic process, created the atmosphere that led to the 2007–8 (and as yet unresolved) finance crisis through the unrestricted and unthinking creation of what Warren Buffett has rightfully and correctly labeled "toxic waste".

The log returns variable in finance plays the same role mathematically as does the space variable x in physics. Only continuous time and "space" stochastic processes are considered here. Stochastic calculus and differential equations are developed in Chapters 2–4, and examples of scaling processes are studied in Chapters 5 and 6. Scaling and correlations are introduced in Chapter 1. Correlations are central to our theme, whereas scaling is not. The reason for this is that scaling has nothing to do with long-time correlations, long-time correlations follow from stationarity of the increments combined with a nonlinear time dependence of the process variance. Furthermore, scaling is not indicated by finance data, the claims of many econophysicists and financial economists to the contrary (interday FX returns can be fitted only by scaling over a few very small time intervals).

The dynamics of martingales in particular, and Ito processes in general, is developed in more detail in Chapters 7–9. In Chapter 10 we follow Durrett in generalizing stochastic calculus from the Wiener process to arbitrary local martingales. There, we also prove and discuss Girsanov's theorem, but mainly we provide examples of solving a partial differential equation "by running an Ito process". First passage time problems (also called stopping times) are formulated and solved. Toward that

end, an elementary discussion of ordinary differential equations linked with partial differential equations is provided in Chapter 1. In Chapter 8 the reader will find new results: the theory and examples of Ito processes that are not Markovian but instead have memory of an initial condition. All time series in the statistical ensemble must pass through that initial condition. Chapter 11 discusses the different historic developments of statistical physics and finance theory and their qualitative and quantitative similarities and differences. Chapters 12–14 have been adapted from my *Dynamics of Markets* (McCauley, 2009) and updated and are used to illustrate the application of martingales to finance. The reasons for the worldwide finance crisis, 2007–present (2012), are discussed, as they were used more extensively in my 2009 book, *Dynamics of Markets*.

We show that a martingale (with finite variance) is a drift free process with uncorrelated noise increments and is therefore diffusive but not necessarily Markovian. Ito stochastic differential equations and their corresponding Fokker–Planck and backward time Kolmogorov pdes are based on martingale noise plus drift and are not restricted to be Markovian. As Doob and Feller stated (see Snell, 1997), but is ignored in the literature, the Chapman–Kolmogorov equation does not imply a Markov process. We add meat to the bare bones by providing several analytically solvable examples of non-Markov Ito processes with memory of an initial state where the Fokker–Planck pde and the Chapman–Kolmogorov equation are satisfied.

A Markov process is defined following the 1960s era Russian literature (see also Wang and Uhlenbeck in Wax, 1954). A Markov process is a process with no memory whatsoever. This means that the drift and diffusion coefficients depend on the present state and on no other state and on no other parameters. This definition prevents us from falling into the trap of labeling an Ito process with memory of one initial condition as a "nonlinear Markov process" (Chapter 8). In a related vein, we also provide an example of superficially nonlinear diffusion in Chapter 6. We explain why 1-point diffusion pdes and 1-point densities cannot be used to imply any particular stochastic process or class of stochastic process. As an example, we show in Chapter 6 that a self-similar Gaussian Markov process has exactly the same 1-point density f(x, t) as fractional Brownian motion. In Chapter 8 we show that examples of processes referred to in the recent literature as "nonlinear Markov" or "nonlinear Fokker–Planck" are linear Ito processes and are either Markovian or else are non-Markovian with memory of an initial condition.

Instead of the abstract probability spaces emphasized in mathematics, we emphasize statistical ensembles, which are necessary for data analysis. We point out that 1-point densities cannot be used to identify a stochastic process, hence, the emphasis in the text on pair correlations. Instead of defining probability spaces, statistical ensembles for nonstationary processes are defined and emphasized. The other side

of the coin is that we explain why time averages should not be used in data analysis. The application of approximate ensembles to time series in finance, a method of analysis constructed by the author and his colleagues Gemunu Gunaratne and Kevin Bassler, is developed in Chapter 13. In Chapter 15 I explain the formal notions of filtrations, adaptations, and probability space measures in terms of time series, n-point densities, and statistical ensembles, allowing the mathematician to understand our language and helping the physicist to understand the abstract mathematical language.

The author happily acknowledges critical and very helpful discussions over the past several years (mainly via email) with Harry Thomas, who unfortunately died in the summer of 2010, with Enrico Scalas, and also with Giulio Bottazzi. The author's two Houston colleagues, Gemunu Gunaratne and Kevin Bassler, contributed much to my understanding and directly to the work presented here. Duncan Foley commented usefully on an earlier version of Chapter 11. Most of the material in the text was taught in my econophysics course at the University of Houston 2004–09. An early preliminary version of the first ten chapters appeared as the author's Pisa Lecture Notes based on a week of lectures at the Sant' Anna School of Economics in 2006; many thanks to my hosts and friends Giulio Bottazzi and Giovani Dosi. Preliminary results from Chapter 13 were not presented then as we had not submitted them for publication at that time. Our sons were 10 and 15 years old then, still boys. After my Pisa lectures, my Frau, two sons, and I spent a very nice week before Easter in a stone house in an olive grove near Vinci. I am extremely grateful to my wife/partner, Cornelia Küffner, for proofreading and criticizing several chapters and the introduction and to Simon Capelin at Cambridge University Press for publishing this text, the author's fifth Cambridge book. Finally, I'm also grateful to our two very good econophysics students at the University of Houston, Lars Seeman and Jia-Chen Hua, for extending the pioneering work of understanding financial time series via a statistical ensemble analysis that Gemunu Gunaratne, Kevin Bassler, and I did in 2007.

1

Random variables and probability distributions

1.1 Particle descriptions of partial differential equations

In this book we will focus mainly on a class of stochastic process that admits descriptions both via a noise-driven differential equation and a corresponding deterministic partial differential equation. We begin with the general question: when does particle motion described by deterministic differential equations admit a corresponding partial differential equation and vice versa. Then, we will add noise and ask the same question.

Particle descriptions of some partial differential equations are possible. Geometrical optics of light rays follows from the characteristic equations $dt = dx/c$ of the linear hyperbolic partial differential equation (pde), the wave equation,

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right)\psi = 0, \tag{1.1}$$

where c is the wave speed. Equation (1.1) can be studied as the consequence of two first-order pdes, each of which has the form of a constant velocity continuity equation. The one pde describes waves traveling along the x-axis, the other describes waves traveling in the opposite direction. This corresponds to the fact that solutions of (1.1) consist of waves traveling in both directions. An oppositely traveling wave arises from a reflection at a boundary.

Let R(x) denote a vector field in an n-dimensional phase space. An autonomous deterministic dynamic system, a phase flow, in particular, $dx/dt = R(x)$ has $n - 1$ (local, not necessarily global) time-independent conservation laws $G(x, t) =$ constant satisfying the first order pde:

$$\frac{dG}{dt} = \frac{\partial G}{\partial t} + R(x)\frac{\partial G}{\partial x} = 0. \tag{1.2}$$

Turning the argument around, this pde has a particle description in terms of the characteristic equations $dt = dx/R$.

The continuity equation from fluid mechanics,

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{\partial(\rho R(x))}{\partial x} = 0, \tag{1.3}$$

where $\rho(x, t)$ is the fluid density and $R(x, t)$ is the fluid velocity field, suggests a particle description. Writing

$$\frac{\partial \rho}{\partial t} + R\frac{\partial \rho}{\partial x} + \rho\frac{\partial R}{\partial x} = 0, \tag{1.4}$$

we have a quasi-linear first order pde whose characteristic equations are given by $dt = dx/R(x, t)$. These differential equations generate the streamlines of the flow locally, but understanding the flow locally requires knowing the three-dimensional velocity field $R(x, t)$. The fluid velocity field is not given a priori. Instead, $R(x, t)$ satisfies the Navier-Stokes pde. Our understanding of fluid turbulence is incomplete, we are not able to solve the Navier-Stokes pde, a quasi-linear second order parabolic pde, adequately approximately analytically for moderate to high Reynolds numbers.

Leaving fluid flow aside and returning to (1.3), where the solution can be given a probability interpretation, here's the main question for us: when does a pde arise from writing

$$dx - R(x)dt = \text{'noise'} \tag{1.5}$$

and vice versa? The answer depends, as we will show in the chapters that follow, entirely on the statistical properties that define the noise. We'll show in Chapter 4 that there are very interesting and useful realizations of this idea. We'll also see in Chapter 6 that there exists noise that can be understood as particle motion but that has *neither* a pde description, *nor* a description in terms of a noise-driven differential equation (1.5).

In order to be precise we must first define what we mean by "noise". In this case, the random differential equation (1.5) cannot generate a trajectory in the deterministic sense. What we would like to regard as "x(t)" can only be defined in some statistical sense depending on the statistics that define the noise. In particular, we will discover that x(t) is wildly nonunique: starting from the same initial condition $x(t_o)$ at time t_o, reruns of the same identical experiment will yield wildly different time series x(t). This leads to the idea of time-series analysis via statistical ensembles. Statistical ensembles were introduced by Gibbs (1960).

Gibbs wrote something similar in his *Elementary Principles of Statistical Mechanics* (1960), but in that case the motion is deterministic chaotic and mixing, not stochastic. The difference is that the trajectories of a deterministic chaotic system diverge wildly for slightly *different* initial conditions and then only at longer times. In a stochastic system the time series diverge wildly for all times t, including for the very shortest time differences $t - t_o$, for exactly the *same* initial condition.

We mean, of course, in principle: "exactly the same initial condition" would require infinite precision, which is physically impossible. Gibbs used a hydrodynamic analogy to make his point, the mixing of a droplet of ink into water when stirred. The same mixing occurs purely dynamically (in Gibbs' sense) if you simply let fall a droplet of ink into the water. A vortex cascade results, and then, after a long-time, diffusion at small scales mixes the ink.

1.2 Random variables and stochastic processes

A random variable x is defined as any variable described by a probability distribution. This implies no randomness of motion, because chaotic differential equations generate pseudo-random time series x(t) and corresponding probability distributions. Here, we're concerned only with stochastic dynamics, "random" *processes* whereby the variable x(t) obeys a noise-driven equation of motion. By a noise process B(t) we mean that there is no change with time in B on the average. That is, there is no deterministic/systematic component in B to produce an average. But how do we define "on the average"? Leaving this ambiguity aside for the moment, we turn to an example that has the germ of generality.

The simplest example of random motion is given by Brownian motion, the motion observed via microscope of a colloidal particle in a fluid. The mathematics of Brownian motion was studied around 1900 by Smoluchowski, Langevin, Einstein, and Bachelier. Markov's earlier work was then generally known. Einstein started with the Langevin equation $dx/dt = \eta(t)$, where $\eta(t)$ is uncorrelated with $\eta(s)$ if $s \neq t$ and with vanishing average. This is called white noise by mathematicians and physicists, but economists mean something entirely different by white noise. Einstein deduced that the probability density f(x, t) for the location of the Brownian particle satisfies the partial differential equation (pde)

$$\frac{\partial f}{\partial t} = \frac{D}{2}\frac{\partial^2 f}{\partial x^2}. \tag{1.6}$$

The generalization to the continuity equation

$$\frac{\partial f}{\partial t} = -\frac{\partial j}{\partial x} \tag{1.7}$$

with

$$j(x,t) = R(x,t)f - \frac{D}{2}\frac{\partial f}{\partial x} \tag{1.8}$$

is called the Smoluchowski pde,

$$\frac{\partial f}{\partial t} = -\frac{\partial (R(x,t)f)}{\partial x} + \frac{D}{2}\frac{\partial^2 f}{\partial x^2}. \tag{1.9}$$

We can argue for the diffusion term in (1.8) in the spirit of irreversible thermodynamics. Or, from the viewpoint of another school, as a Landau-Ginzburg expansion: think of an expansion of j in powers of, and gradients of, f. The coefficient of the gradient term in (1.8) must be positive because particles will tend to diffuse from higher to lower concentrations, $D > 0$.

This is the generalization of (1.3) that includes white noise acting on the particle's velocity. Unlike first order pdes or hyperbolic second order pdes, a parabolic second order pde admits no description of trajectories by a deterministic equation of motion. Equation (1.3) is time reversible but (1.9) is not.

Deterministic motion along a particle's path is always time reversible: one can integrate a system of deterministic differential equations $dx/dt = R(x)$, where R is a vector field, either backward or forward in time. Dissipative motion, deterministic chaos, and complexity are all time reversible. In diffusive dynamics, however, time-reversed motion does not exist. The lack of time reversibility is reflected in the lack of self-adjointness: the pde (Equation 1.6) cannot be made self-adjoint under general boundary or initial conditions. Instead, the pde has an adjoint

$$\frac{\partial f}{\partial t} + \frac{D}{2}\frac{\partial^2 f}{\partial x^2} = 0, \tag{1.10}$$

that describes diffusion *backward* but not forward, in time. Time irreversibility is fundamental for stochastic processes in general and for statistical physics and finance theory in particular. The Black–Scholes (1973) pde for option pricing is, to within a simple transformation, the adjoint of a forward time diffusion equation that describes the distribution of a stock price. We will see that martingales are inherently connected with the idea of diffusion backward in time.

The particle trajectories described by (1.6) are everywhere continuous but are nowhere differentiable, are infinitely jagged, and are, therefore, difficult to describe mathematically. Here's the description of the particle motion generated by (1.6). Suppose that we know the position of a particle at time $t = 0$, let $x(0) = 0$, e.g., where's the particle at time $t > 0$? If we assume that the motion is unbounded, $-\infty < x < \infty$, then the global solution of (1.6) for $x(0) = 0$ is

$$f(x, t) = (2\pi Dt)^{-1/2} e^{-x^2/2Dt}. \tag{1.11}$$

"On the average", the position predicted by (1.11) is zero: $\langle x(t)\rangle = \int x f(x, t)dx = 0$ since $f(x, t) = f(-x, t)$. The probability that the particle lies within the interval $[x_1 - \Delta x, x_1 + \Delta x]$ at time t is then expressed as

$$P(x_1 - \Delta x \le x \le x_1 + \Delta x) = (2\pi Dt)^{-1/2} \int_{x_1-\Delta x}^{x_1+\Delta x} dy e^{-y^2/2Dt}. \tag{1.12}$$

More accurate knowledge of the particle's location than this cannot be predicted in advance. Our ability to predict the particle's future position falls off faster than $t^{-1/2}$, so we systematically and irreversibly lose the knowledge needed to predict the future. Our ability to deduce the particle's earlier history is nil. The most likely position of the particle is the position that maximizes lnf(x, t), yielding $x = 0$, but the scatter about that position grows as t increases. Our inability to predict the particle's position in the long run is in fact unbounded because the process does not approach statistical equilibrium. Equilibrium would require that the probability density f(x, t) approaches a t-independent normalizable limit f(x) as t goes to infinity. We can achieve statistical equilibrium if the particle's confined to a box, $-L \leq x \leq L$, which requires solving (1.6) subject to boundary conditions $f(-L, t) = f(L, t) = 0$. This yields $f(x) = 1/L$ with $-L < \times < L$ for $t \gg 0$.

This is a far simpler state of affairs than for a complex deterministic system, where no probability at all can be predicted. So stochastic processes are simple, not complex, at least for all processes considered this book.

Observe that a particle description of Laplace's pde $\nabla^2 f = 0$ is suggested as in terms of steady-state motions of a Brownian particle. We'll return to this topic in Chapter 10, where the method used to construct solutions relies on martingales.

There are two key questions. First, how can probability densities and statistical averages be obtained from *measurements* of the motion of a particle in a random system? This requires a discussion of statistical ensembles. Second, how much information is required mathematically to pin down a specific random dynamical system? We address the second question in the next section because it leads us to the notion of a conditional probability, which is useful for clarifying the first question. The second question is answered discussed further in Chapter 6 and from a practical standpoint in Section 1.7.

1.3 The n-point probability distributions

According to Kolmogorov, we can define a specific stochastic process precisely iff. we can specify the complete, infinite hierarchy of joint probability distributions. From an empirical standpoint we can at best obtain finitely many histograms from measurement of N different runs of an experiment for n different times, using the frequency definition of probability. So we can never hope to specify a process mathematically uniquely via observation. The best we can hope for is to specify some *class* of process.

Let $P_1(x)$ denote the probability to find a value $X < x$ at time t. This is the 1-point distribution of x. Then $P_2(x, t; x', t')$ denotes the probability to find both $X < x$ at time t and $X' < x'$ at time t', and so on, up to $P_n(x_1, t_1; \ldots; x_n, t_n)$. Clearly, both the number N of runs and the number n of times that we strobe/observe the

system must be large in order have any hope of getting good statistics (meaning reliable histograms). Nonuniqueness in modeling is inherent in the fact that we can *at best* discover the lowest few distributions P_n from empirical data. The class of models that we are able to deduce from data may be unique at the level of P_2 but then will be nonunique at the level of P_3 and beyond. But, in that case, what cannot be discovered from measurement should not be interpreted as license for invention from mathematical imagination. In Section 1.7 we'll discuss the question of whether we can even hope to measure P_1.

We can introduce the hierarchy of probability densities f_n via

$$dP_n(x_1, t_1; \ldots; x_n, t_n) = f_n(x_1, t_1; \ldots; x_n, t_n)dx_1 \ldots dx_n. \tag{1.13}$$

Consider a time series x(t) representing one run of a stochastic process. Empirically, we can only strobe the system a finite number of times, so measurements of x(t) take the form of $x(t_k)$, $k = 1, \ldots, n$ where n is the number of measurements made. If we can extract good-enough histograms from the data, then we can construct the hierarchy of probability densities $f_1(x, t)$, $f_2(x_1, t; x_2, t_2), \ldots, f_k(x_1, t_1; \ldots; x_k, t_k)$, where $k \ll n$ (the 1-point density f_1 reflects a specific choice of initial condition in data analysis). To get decent histograms for f_n, one would then need a much longer time series.

We note that

$$f_{n-1}(x_1,t_1; \ldots; x_{k-1},t_{k-1}; x_{k+1},t_{k+1}; \ldots; x_n,t_n) = \int dx_k f_n(x_1,t_1; \ldots; x_n,t_n) \tag{1.14}$$

so

$$f_1(x, t) = \int dy f_2(y, s; x, t). \tag{1.15}$$

Densities of all orders are normalized to unity if $\int dx x f_1(x, t) = 1$.

We turn next to the question of scaling, which is a property of the 1-point density whenever scaling occurs.

1.4 Simple averages and scaling

Given a dynamical variable A(x), to calculate simple averages we only need the 1-point density,

$$\langle A(x)\rangle = \int A(x)f_1(x, t)dx. \tag{1.16}$$

For example, moments are predicted by

$$\langle x^n\rangle = \int x^n f_1(x, t)dx. \tag{1.17}$$