Contents

Preface

This book originated with several courses given at the University of Texas. The audience consisted of graduate students of mathematics, physics, electrical engineering, and finance. Most had met some stochastic analysis during work in their field; the course was meant to provide the mathematical underpinning. To satisfy the economists, driving processes other than Wiener process had to be treated; to give the mathematicians a chance to connect with the literature and discrete-time martingales, I chose to include driving terms with jumps. This plus a predilection for generality for simplicity's sake led directly to the most general stochastic Lebesgue–Stieltjes integral.

The spirit of the exposition is as follows: just as having finite variation and being right-continuous identifies the useful Lebesgue–Stieltjes distribution functions among all functions on the line, are there criteria for processes to be useful as "random distribution functions." They turn out to be straightforward generalizations of those on the line. A process that meets these criteria is called an *integrator*, and its integration theory is just as easy as that of a deterministic distribution function on the line – provided Daniell's method is used. (This proviso has to do with the lack of convexity in some of the target spaces of the stochastic integral.)

For the purpose of error estimates in approximations both to the stochastic integral and to solutions of stochastic differential equations we define various *numerical sizes of an integrator* Z and analyze rather carefully how they propagate through many operations done on and with Z, for instance, solving a stochastic differential equation driven by Z. These size-measurements arise as generalizations to integrators of the famed Burkholder–Davis–Gundy inequalities for martingales. The present exposition differs in the ubiquitous use of numerical estimates from the many fine books on the market, where convergence arguments are usually done in probability or every once in a while in Hilbert space L^2. For reasons that unfold with the story we employ the L^p-norms in the whole range $0 \le p < \infty$. An effort is made to furnish reasonable estimates for the universal constants that occur in this context.

Such attention to estimates, unusual as it may be for a book on this subject, pays handsomely with some new results that may be edifying even to the expert. For instance, it turns out that every integrator Z can be controlled

by an increasing previsible process much like a Wiener process is controlled by time t; and if not with respect to the given probability, then at least with respect to an equivalent one that lets one view the given integrator as a map into Hilbert space, where computation is comparatively facile. This *previsible controller* obviates prelocal arguments [92] and can be used to construct Picard norms for the solution of stochastic differential equations driven by Z that allow growth estimates, easy treatment of stability theory, and even *pathwise algorithms* for the solution. These schemes extend without ado to *random measures*, including the previsible control and its application to stochastic differential equations driven by them.

All this would seem to lead necessarily to an enormous number of technicalities. A strenuous effort is made to keep them to a minimum, by these devices: everything not directly needed in stochastic integration theory and its application to the solution of stochastic differential equations is either omitted or relegated to the Supplements or to the Appendices. A short survey of the beautiful "General Theory of Processes" developed by the French school can be found there.

A warning concerning the usual conditions is appropriate at this point. They have been replaced throughout with what I call the *natural conditions*. This will no doubt arouse the ire of experts who think one should not "tamper with a mature field." However, many fine books contain erroneous statements of the important Girsanov theorem – in fact, it is hard to find a correct statement in unbounded time – and this is traceable directly to the employ of the usual conditions (see example 3.9.14 on page 164 and 3.9.20). In mathematics, correctness trumps conformity. The natural conditions confer the same benefits as do the usual ones: path regularity (section 2.3), section theorems (page 437 ff.), and an ample supply of stopping times (*ibidem*), without setting a trap in Girsanov's theorem.

The students were expected to know the basics of point set topology up to Tychonoff's theorem, general integration theory, and enough functional analysis to recognize the Hahn–Banach theorem. If a fact fancier than that is needed, it is provided in appendix A, or at least a reference is given.

The exercises are sprinkled throughout the text and form an integral part. They have the following appearance:

Exercise 4.3.2 This is an exercise. It is set in a smaller font. It requires no novel argument to solve it, only arguments and results that have appeared earlier. Answers to some of the exercises can be found in appendix B. Answers to most of them can be found in appendix C, which is available on the web *via* `http://www.ma.utexas.edu/users/cup/Answers`.

I made an effort to index every technical term that appears (page 489), and to make an index of notation that gives a short explanation of every symbol and lists the page where it is defined in full (page 483). Both indexes appear in expanded form at `http://www.ma.utexas.edu/users/cup/Indexes`.

`http://www.ma.utexas.edu/users/cup/Errata` contains the errata. I plead with the gentle reader to send me the errors he/she found via email to `kbi@math.utexas.edu`, so that I may include them, with proper credit of course, in these errata.

At this point I recommend reading the conventions on page 363.

1

Introduction

1.1 Motivation: Stochastic Differential Equations

Stochastic Integration and Stochastic Differential Equations (SDEs) appear in analysis in various guises. An example from physics will perhaps best illuminate the need for this field and give an inkling of its particularities. Consider a physical system whose state at time t is described by a vector X_t in $\mathbb{R}^n$. In fact, for concreteness' sake imagine that the system is a space probe on the way to the moon. The pertinent quantities are its location and momentum. If x_t is its location at time t and p_t its momentum at that instant, then X_t is the 6-vector (x_t, p_t) in the phase space $\mathbb{R}^6$. In an ideal world the evolution of the state is governed by a differential equation:

$$\frac{dX_t}{dt} = \begin{pmatrix} dx_t/dt \\ dp_t/dt \end{pmatrix} = \begin{pmatrix} p_t/m \\ F(x_t, p_t) \end{pmatrix} .$$

Here m is the mass of the probe. The first line is merely the definition of p: momentum = mass $\times$ velocity. The second line is Newton's second law: the rate of change of the momentum is the force F. For simplicity of reading we rewrite this in the form

$$dX_t = a(X_t)\, dt \,, \tag{1.1.1}$$

which expresses the idea that the change of X_t during the time-interval dt is proportional to the time dt elapsed, with a proportionality constant or *coupling coefficient* a that depends on the state of the system and is provided by a model for the forces acting. In the present case $a(X)$ is the 6-vector $(p/m, F(X))$. Given the initial state X_0, there will be a unique solution to (1.1.1). The usual way to show the existence of this solution is Picard's iterative scheme: first one observes that (1.1.1) can be rewritten in the form of an *integral equation*:

$$X_t = X_0 + \int_0^t a(X_s)\, ds \,. \tag{1.1.2}$$

Then one starts Picard's scheme with $X_t^0 = X_0$ or a better guess and defines the iterates inductively by

$$X_t^{n+1} = X^0 + \int_0^t a(X_s^n)\, ds \,.$$

If the coupling coefficient a is a Lipschitz function of its argument, then the *Picard iterates* X^n will converge uniformly on every bounded time-interval and the limit X^∞ is a solution of (1.1.2), and thus of (1.1.1), and the only one. The reader who has forgotten how this works can find details on pages 274–281. Even if the solution of (1.1.1) cannot be written as an analytical expression in t, there exist extremely fast numerical methods that compute it to very high accuracy. Things look rosy.

In the less-than-ideal real world our system is subject to unknown forces, *noise*. Our rocket will travel through gullies in the gravitational field that are due to unknown inhomogeneities in the mass distribution of the earth; it will meet gusts of wind that cannot be foreseen; it might even run into a gaggle of geese that deflect it. The evolution of the system is better modeled by an equation

$$dX_t = a(X_t)\,dt + dG_t\ , \tag{1.1.3}$$

where G_t is a noise that contributes its differential dG_t to the change dX_t of X_t during the interval dt. To accommodate the idea that the noise comes from without the system one assumes that there is a *background noise* Z_t – consisting of gravitational gullies, gusts, and geese in our example – and that its effect on the state during the time-interval dt is proportional to the difference dZ_t of the *cumulative noise* Z_t during the time-interval dt, with a proportionality constant or *coupling coefficient* b that depends on the state of the system:

$$dG_t = b(X_t)\,dZ_t\ .$$

For instance, if our probe is at time t halfway to the moon, then the effect of the gaggle of geese at that instant should be considered negligible, and the effect of the gravitational gullies is small. Equation (1.1.3) turns into

$$dX_t = a(X_t)\,dt + b(X_t)\,dZ_t\ , \tag{1.1.4}$$

$$\text{in integrated form}\quad X_t = X_t^0 + \int_0^t a(X_s)\,ds + \int_0^t b(X_s)\,dZ_s\ . \tag{1.1.5}$$

What is the meaning of this equation in practical terms? Since the background noise Z_t is not known one cannot solve (1.1.5), and nothing seems to be gained. Let us not give up too easily, though. Physical intuition tells us that the rocket, though deflected by gullies, gusts, and geese, will probably not turn all the way around but will rather still head somewhere in the vicinity of the moon. In fact, for all we know the various noises might just cancel each other and permit a perfect landing.

What are the chances of this happening? They seem remote, perhaps, yet it is obviously important to find out how likely it is that our vehicle will at least hit the moon or, better, hit it reasonably closely to the intended landing site. The smaller the noise dZ_t, or at least its *effect* $b(X_t)\,dZ_t$, the better we feel the chances will be. In other words, our intuition tells us to look for

a statistical inference: from some reasonable or measurable assumptions on the background noise Z or its effect $b(X)dZ$ we hope to conclude about the likelihood of a successful landing.

This is all a bit vague. We must cast the preceding contemplations in a mathematical framework in order to talk about them with precision and, if possible, to obtain quantitative answers. To this end let us introduce the set Ω of all possible *evolutions of the world*. The idea is this: at the beginning $t = 0$ of the reckoning of time we may or may not know the *state-of-the-world* ω_0, but thereafter the course that the history $\omega : t \mapsto \omega_t$ of the world actually will take has the vast collection Ω of evolutions to choose from. For any two possible courses-of-history[1] $\omega : t \mapsto \omega_t$ and $\omega' : t \mapsto \omega'_t$ the state-of-the-world might take there will generally correspond different cumulative background noises $t \mapsto Z_t(\omega)$ and $t \mapsto Z_t(\omega')$. We stipulate further that there is a function $\mathbb{P}$ that assigns to certain subsets E of Ω, the *events*, a *probability* $\mathbb{P}[E]$ that they will occur, i.e., that the actual evolution lies in E. It is known that no reasonable probability $\mathbb{P}$ can be defined on *all* subsets of Ω. We assume therefore that the collection of all events that can ever be observed or are ever pertinent form a σ-algebra $\mathcal{F}$ of subsets of Ω and that the function $\mathbb{P}$ is a probability measure on $\mathcal{F}$. It is not altogether easy to defend these assumptions. Why should the observable events form a σ-algebra? Why should $\mathbb{P}$ be σ-additive? We content ourselves with this answer: there is a well-developed theory of such triples $(\Omega, \mathcal{F}, \mathbb{P})$; it comprises a rich calculus, and we want to make use of it. Kolmogorov [58] has a better answer:

Project 1.1.1 *Make a mathematical model for the analysis of random phenomena that does not require σ-additivity at the outset but furnishes it instead.*

So, for every possible course-of-history[1] $\omega \in \Omega$ there is a *background noise* $Z_\bullet : t \mapsto Z_t(\omega)$, and with it comes the *effective noise* $b(X_t)\, dZ_t(\omega)$ that our system is subject to during dt. Evidently the state X_t of the system depends on ω as well. The obvious thing to do here is to compute, for every $\omega \in \Omega$, the solution of equation (1.1.5), to wit,

$$X_t(\omega) = X_t^0 + \int_0^t a(X_s(\omega))\, ds + \int_0^t b(X_s(\omega))\, dZ_s(\omega)\ , \qquad (1.1.6)$$

as the limit of the Picard iterates $X_t^0 \stackrel{\text{def}}{=} X_0$,

$$X_t^{n+1}(\omega) \stackrel{\text{def}}{=} X_t^0 + \int_0^t a(X_s^n(\omega))\, ds + \int_0^t b(X_s^n(\omega))\, dZ_s(\omega)\ . \qquad (1.1.7)$$

Let T be the time when the probe hits the moon. This depends on chance, of course: $T = T(\omega)$. Recall that x_t are the three spatial components of X_t.

[1] The redundancy in these words is for emphasis. [Note how repeated references to a footnote like this one are handled. Also read the last line of the chapter on page 41 to see how to find a repeated footnote.]

Our interest is in the function $\omega \mapsto x_T(\omega) = x_{T(\omega)}(\omega)$, the location of the probe at the time T. Suppose we consider a landing successful if our probe lands within F feet of the ideal landing site s at the time T it does land. We are then most interested in the probability

$$p_F \stackrel{\text{def}}{=} \mathbb{P}\big(\{\omega \in \Omega : \big\| x_T(\omega) - s \big\| < F\}\big)$$

of a successful landing – its value should influence strongly our decision to launch. Now x_T is just a function on Ω, albeit defined in a circuitous way. We should be able to compute the set $\{\omega \in \Omega : \| x_T(\omega) - s \| < F\}$, and if we have enough information about $\mathbb{P}$, we should be able to compute its probability p_F and to make a decision. This is all classical ordinary differential equations (ODE), complicated by the presence of a parameter ω: straightforward in principle, if possibly hard in execution.

The Obstacle

As long as the *paths* $Z_\bullet(\omega) : s \mapsto Z_s(\omega)$ of the background noise are right-continuous and have finite variation, the integrals $\int \cdots_s dZ_s$ appearing in equations (1.1.6) and (1.1.7) have a perfectly clear classical meaning as Lebesgue–Stieltjes integrals, and Picard's scheme works as usual, under the assumption that the coupling coefficients a, b are Lipschitz functions (see pages 274–281).

Now, since we do not know the background noise Z precisely, we must make a model about its statistical behavior. And here a formidable obstacle rears its head: the simplest and most plausible statistical assumptions about Z force it to be so irregular that the integrals of (1.1.6) and (1.1.7) cannot be interpreted in terms of the usual integration theory. The moment we stipulate some symmetry that merely expresses the idea that we don't know it all, obstacles arise that cause the paths of Z to have infinite variation and thus prevent the use of the Lebesgue–Stieltjes integral in giving a meaning to expressions like $\int X_s \, dZ_s(\omega)$.

Here are two assumptions on the random *driving term* Z that are eminently plausible:

(a) The expectation of the increment $dZ_t \approx Z_{t+h} - Z_t$ should be zero; otherwise there is a *drift* part to the noise, which should be subsumed in the first driving term $\int \cdot \, ds$ of equation (1.1.6). We may want to assume a bit more, namely, that if everything of interest, including the noise $Z_\bullet(\omega)$, was actually observed up to time t, then the future increment $Z_{t+h} - Z_t$ still averages to zero. Again, if this is not so, then a part of Z can be shifted into a driving term of finite variation so that the remainder satisfies this condition – see theorem 4.3.1 on page 221 and proposition 4.4.1 on page 233. The mathematical formulation of this idea is as follows: let $\mathcal{F}_t$ be the σ-algebra generated by the collection of all observations that can be made before and at

time t; $\mathcal{F}_t$ is commonly and with intuitive appeal called the *history* or *past* at time t. In these terms our assumption is that the *conditional expectation*

$$\mathbb{E}\left[Z_{t+h} - Z_t \middle| \mathcal{F}_t\right]$$

of the future differential noise given the past vanishes. This makes Z a *martingale on the filtration* $\mathcal{F}_\bullet = \{\mathcal{F}_t\}_{0\le t<\infty}$ – these notions are discussed in detail in sections 1.3 and 2.5.

(b) We may want to assume further that Z does not change too wildly with time, say, that the paths $s \mapsto Z_s(\omega)$ are continuous. In the example of our space probe this reflects the idea that it will not blow up or be hit by lightning; these would be huge and sudden disturbances that we avoid by careful engineering and by not launching during a thunderstorm.

A background noise Z satisfying (a) and (b) has the property that *almost none of its paths* $Z_\bullet(\omega)$ *is differentiable at any instant* – see exercise 3.8.13 on page 152. By a well-known theorem of real analysis,[2] the path $s \mapsto Z_s(\omega)$ does not have finite variation on any time-interval; and this irregularity happens for almost every $\omega \in \Omega$!

We are stumped: since $s \mapsto Z_s$ does not have finite variation, the integrals $\int \cdots dZ_s$ appearing in equations (1.1.6) and (1.1.7) do not make sense in any way we know, and then neither do the equations themselves.

Historically, the situation stalled at this juncture for quite a while. Wiener made an attempt to define the integrals in question in the sense of distribution theory, but the resulting *Wiener integral* is unsuitable for the iteration scheme (1.1.7), for lack of decent limit theorems.

Itô's Way Out of the Quandary

The problem is evidently to give a meaning to the integrals appearing in (1.1.6) and (1.1.7). Not only that, any prospective integral must have rather good properties: to show that the iterates X^n of (1.1.7) form a Cauchy sequence and thus converge there must be estimates available; to show that their limit is the solution of (1.1.6) there must be a limit theorem that permits the interchange of limit and integral, to wit,

$$\int_0^t \lim_n b\big(X_s^n\big)\, dZ_s = \lim_n \int_0^t b\big(X_s^n\big)\, dZ_s\ .$$

In other words, what is needed is an integral satisfying the Dominated Convergence Theorem, say. Convinced that an integral with this property cannot be defined *pathwise*, i.e., ω for ω, the Japanese mathematician Itô decided to try for an integral in the sense of the L^2-mean. His idea was this: while the sums

$$S_{\mathcal{P}}(\omega) \stackrel{\text{def}}{=} \sum_{k=1}^{K} b\big(X_{\sigma_k}(\omega)\big)\left(Z_{s_{k+1}}(\omega) - Z_{s_k}(\omega)\right)\ ,\ s_k \le \sigma_k \le s_{k+1}\,, \qquad (1.1.8)$$

[2] See for example [97, pages 94–100] or [9, page 157 ff.].

which appear in the usual definition of the integral, do not converge for any $\omega \in \Omega$, there may obtain *convergence in mean* as the partition $\mathcal{P} = \{s_0 < s_1 < \ldots < s_{K+1}\}$ is refined. In other words, there may be a random variable I such that

$$\| S_{\mathcal{P}} - I \|_{L^2} \to 0 \quad \text{as} \quad \text{mesh}[\mathcal{P}] \to 0 \,.$$

And if $S_{\mathcal{P}}$ should not converge in L^2-mean, it may converge in L^p-mean for some other $p \in (0, \infty)$, or at least in measure $(p = 0)$.

In fact, this approach succeeds, but not without another observation that Itô made: for the purpose of Picard's scheme it is not necessary to integrate all processes.[3] *An integral defined for non-anticipating integrands suffices.* In order to describe this notion with a modicum of precision, we must refer again to the σ-algebras $\mathcal{F}_t$ comprising the history known at time t. The integrals $\int_0^t a(X_0)\, ds = a(X_0) \cdot t$ and $\int_0^t b(X_0)\, dZ_s(\omega) = b(X_0) \cdot \big(Z_t(\omega) - Z_0(\omega)\big)$ are at any time measurable on $\mathcal{F}_t$ because Z_t is; then so is the first Picard iterate X_t^1. Suppose it is true that the iterate X^n of Picard's scheme is at all times t measurable on $\mathcal{F}_t$; then so are $a(X_t^n)$ and $b(X_t^n)$. Their integrals, being limits of sums as in (1.1.8), will again be measurable on $\mathcal{F}_t$ at all instants t; then so will be the next Picard iterate X_t^{n+1} and with it $a(X_t^{n+1})$ and $b(X_t^{n+1})$, and so on. In other words, the integrands that have to be dealt with *do not anticipate the future*; rather, they are at any instant t measurable on the past $\mathcal{F}_t$. If this is to hold for the approximation of (1.1.8) as well, we are forced to choose for the point σ_i at which $b(X)$ is evaluated the left endpoint s_{i-1}. We shall see in theorem 2.5.24 that the choice $\sigma_i = s_{i-1}$ permits martingale[4] drivers Z – recall that it is the martingales that are causing the problems.

Since our object is to obtain statistical information, evaluating integrals and solving stochastic differential equations in the sense of a mean would pose no philosophical obstacle. It is, however, now not quite clear what it is that equation (1.1.5) models, if the integral is understood in the sense of the mean. Namely, what is the mechanism by which the random variable dZ_t affects the change dX_t in mean but not through its actual realization $dZ_t(\omega)$? Do the possible but not actually realized courses-of-history[1] somehow influence the behavior of our system? We shall return to this question in remarks 3.7.27 on page 141 and give a rather satisfactory answer in section 5.4 on page 310.

Summary: The Task Ahead

It is now clear what has to be done. First, the stochastic integral in the L^p-mean sense for non-anticipating integrands has to be developed. This

[3] A process is simply a function $Y : (s, \omega) \mapsto Y_s(\omega)$ on $\mathbb{R}_+ \times \Omega$. Think of $Y_s(\omega) = b\big(X_s(\omega)\big)$.
[4] See page 5 and section 2.5, where this notion is discussed in detail.

is surprisingly easy. As in the case of integrals on the line, the integral is defined first in a non-controversial way on a collection $\mathcal{E}$ of *elementary integrands*. These are the analogs of the familiar step functions. Then that *elementary integral* is extended to a large class of processes in such a way that it features the Dominated Convergence Theorem. This is not possible for arbitrary driving terms Z, just as not every function z on the line is the distribution function of a σ-additive measure – to earn that distinction z must be right-continuous and have finite variation. The stochastic driving terms Z for which an extension with the desired properties has a chance to exist are identified by conditions completely analogous to these two and are called *integrators*.

For the extension proper we employ Daniell's method. The arguments are so similar to the usual ones that it would suffice to state the theorems, were it not for the deplorable fact that Daniell's procedure is generally not too well known, is even being resisted. Its efficacy is unsurpassed, in particular in the stochastic case.

Then it has to be shown that the integral found can, in fact, be used to solve the stochastic differential equation (1.1.5). Again, the arguments are straightforward adaptations of the classical ones outlined in the beginning of section 5.1, jazzed up a bit in the manner well known from the theory of ordinary differential equations in Banach spaces (e.g., [22, page 279 ff.] – the reader need not be familiar with it, as the details are developed in chapter 5). A pleasant surprise waits in the wings. Although the integrals appearing in (1.1.6) cannot be understood pathwise in the ordinary sense, there is an algorithm that solves (1.1.6) pathwise, i.e., ω–by–ω. This answers satisfactorily the question raised above concerning the meaning of solving a stochastic differential equation "in mean."

Indeed, why not let the cat out of the bag: the algorithm is simply the method of Euler–Peano. Recall how this works in the case of the deterministic differential equation $dX_t = a(X_t)\,dt$. One gives oneself a threshold δ and defines inductively an approximate solution $X_t^{(\delta)}$ at the points $t_k \stackrel{\text{def}}{=} k\delta$, $k \in \mathbb{N}$, as follows: if $X_{t_k}^{(\delta)}$ is constructed, wait until the driving term t has changed by δ, and let $t_{k+1} \stackrel{\text{def}}{=} t_k + \delta$ and

$$X_{t_{k+1}}^{(\delta)} = X_{t_k}^{(\delta)} + a(X_{t_k}^{(\delta)}) \times (t_{k+1} - t_k)\ ;$$

between t_k and t_{k+1} define $X_t^{(\delta)}$ by linearity. The compactness criterion A.2.38 of Ascoli–Arzelà allows the conclusion that the polygonal paths $X^{(\delta)}$ have a limit point as $\delta \to 0$, which is a solution. This scheme actually expresses more intuitively the meaning of the equation $dX_t = a(X_t)\,dt$ than does Picard's. If one can show that it converges, one should be satisfied that the limit is for all intents and purposes a solution of the differential equation.

In fact, the adaptive version of this scheme, where one waits until the *effect of the driving term* $a(X_{t_k}^{(\delta)}) \times (t - t_k)$ is sufficiently large to define t_{k+1}

and $X^{(\delta)}_{t_{k+1}}$, does converge for almost all $\omega \in \Omega$ in the stochastic case, when the deterministic driving term $t \mapsto t$ is replaced by the stochastic driver $t \mapsto Z_t(\omega)$ (see section 5.4).

So now the reader might well ask why we should go through all the labor of stochastic integration: integrals do not even appear in this scheme! And the question of what it means to solve a stochastic differential equation "in mean" does not arise. The answer is that there seems to be no way to prove the almost sure convergence of the Euler–Peano scheme directly, due to the absence of compactness. One has to show[5] that the Picard scheme works before the Euler–Peano scheme can be proved to converge.

So here is a new perspective: what we mean by a solution of equation (1.1.4),

$$dX_t(\omega) = a(X_t(\omega))\,dt + b(X_t(\omega))\,dZ_t(\omega)\ ,$$

is a limit to the Euler–Peano scheme. Much of the labor in these notes is expended just to establish via stochastic integration and Picard's method that this scheme does, in fact, converge almost surely.

Two further points. First, even if the model for the background noise Z is simple, say, is a Wiener process, the stochastic integration theory must be developed for integrators more general than that. The reason is that the solution of a stochastic differential equation is itself an integrator, and in this capacity it can best be analyzed. Moreover, in mathematical finance and in filtering and control theory, the solution of one stochastic differential equation is often used to drive another.

Next, in most applications the state of the system will have many components and there will be several background noises; the stochastic differential equation (1.1.5) then becomes[6]

$$X^\nu_t = C^\nu_t + \sum_{1\le\eta\le d}\int_0^t F^\nu_\eta[X^1,\dots,X^n]\,dZ^\eta\ , \qquad \nu = 1,\dots,n\ .$$

The state of the system is a vector $X = (X^\nu)^{\nu=1\dots n}$ in $\mathbb{R}^n$ whose evolution is driven by a collection $\left\{Z^\eta : 1 \le \eta \le d\right\}$ of scalar integrators. The d vector fields $F_\eta = (F^\nu_\eta)^{\nu=1\dots n}$ are the *coupling coefficients*, which describe the effect of the background noises Z^η on the change of X. $C_t = (C^\nu_t)^{\nu=1\dots n}$ is the *initial condition* – it is convenient to abandon the idea that it be constant. It eases the reading to rewrite the previous equation in vector notation as[7]

$$X_t = C_t + \int_0^t F_\eta[X]\,dZ^\eta\ . \tag{1.1.9}$$

[5] So far – here is a challenge for the reader!

[6] See equation (5.2.1) on page 282 for a more precise discussion.

[7] We shall use the ***Einstein convention*** throughout: summation over repeated indices *in opposite positions* (the η in (1.1.9)) is implied.

The form (1.1.9) offers an intuitive way of reading the stochastic differential equation: the noise Z^η *drives* the state X in the direction $F_\eta[X]$. In our example we had four driving terms: $Z^1_t = t$ is time and F_1 is the systemic force; Z^2 describes the gravitational gullies and F_2 their effect; and Z^3 and Z^4 describe the gusts of wind and the gaggle of geese, respectively. The need for several noises will occasionally call for estimates involving whole slews $\{Z^1, ..., Z^d\}$ of integrators.

1.2 Wiener Process

Wiener process[8] is the model most frequently used for a background noise. It can perhaps best be motivated by looking at Brownian motion, for which it was an early model. Brownian motion is an example not far removed from our space probe, in that it concerns the motion of a particle moving under the influence of noise. It is simple enough to allow a good stab at the background noise.

Example 1.2.1 (Brownian Motion) Soon after the invention of the microscope in the 17th century it was observed that pollen immersed in a fluid of its own specific weight does not stay calmly suspended but rather moves about in a highly irregular fashion, and never stops. The English physicist Brown studied this phenomenon extensively in the early part of the 19th century and found some systematic behavior: the motion is the more pronounced the smaller the pollen and the higher the temperature; the pollen does not aim for any goal – rather, during any time-interval its path appears much the same as it does during any other interval of like duration, and it also looks the same if the direction of time is reversed. There was speculation that the pollen, being live matter, is propelling itself through the fluid. This, however, runs into the objection that it must have infinite energy to do so (jars of fluid with pollen in it were stored for up to 20 years in dark, cool places, after which the pollen was observed to jitter about with undiminished enthusiasm); worse, ground-up granite instead of pollen showed the same behavior.

In 1905 Einstein wrote three Nobel-prize–worthy papers. One offered the Special Theory of Relativity, another explained the Photoeffect (for this he got the Nobel prize), and the third gave an explanation of Brownian motion. It rests on the idea that the pollen is kicked by the much smaller fluid molecules, which are in constant thermal motion. The idea is not, as one might think at first, that the little jittery movements one observes are due to kicks received from particularly energetic molecules; estimates of the distribution of the kinetic energy of the fluid molecules rule this out. Rather, it is this: the pollen suffers an enormous number of collisions with the molecules of the surrounding fluid, each trying to propel it in a different direction, but mostly canceling each other; *the motion observed is due to*

[8] "Wiener process" is sometimes used without an article, in the way "Hilbert space" is.

statistical fluctuations. Formulating this in mathematical terms leads to a stochastic differential equation[9]

$$\begin{pmatrix} dx_t \\ dp_t \end{pmatrix} = \begin{pmatrix} p_t/m\ dt \\ -\alpha\, p_t\ dt\ +\ d\boldsymbol{W}_t \end{pmatrix} \tag{1.2.1}$$

for the location (x, p) of the pollen *in its phase space.* The first line expresses merely the definition of the momentum p; namely, the rate of change of the location x in $\mathbb{R}^3$ is the velocity $v = p/m$, m being the mass of the pollen. The second line attributes the change of p during dt to two causes: $-\alpha p\, dt$ describes the resistance to motion due to the viscosity of the fluid, and $d\boldsymbol{W}_t$ is the sum of the very small momenta that the enormous number of collisions impart to the pollen during dt. The random driving term is denoted $\boldsymbol{W}$ here rather than Z as in section 1.1, since the model for it will be a Wiener process.

This explanation leads to a plausible model for the background noise $\boldsymbol{W}$: $d\boldsymbol{W}_t = \boldsymbol{W}_{t+dt} - \boldsymbol{W}_t$ is the sum of a huge number of exceedingly small momenta, so by the Central Limit Theorem A.4.4 we expect $d\boldsymbol{W}_t$ to have a normal law. (For the notion of a law or distribution see section A.3 on page 391. We won't discuss here Lindeberg's or other conditions that would make this argument more rigorous; let us just assume that whatever condition on the distribution of the momenta of the molecules needed for the CLT is satisfied. We are, after all, doing heuristics here.)

We do not see any reason why kicks in one direction should, on the average, be more likely than in any other, so this normal law should have expectation zero and a multiple of the identity for its covariance matrix. In other words, it is plausible to stipulate that $d\boldsymbol{W}$ be a 3-vector of identically distributed independent normal random variables. It suffices to analyze one of its three scalar components; let us denote it by dW.

Next, there is no reason to believe that the total momenta imparted during non-overlapping time-intervals should have anything to do with one another. In terms of W this means that for consecutive instants $0 = t_0 < t_1 < t_2 < \ldots < t_K$ the corresponding family of consecutive ***increments***

$$\Big\{ W_{t_1} - W_{t_0}\,,\, W_{t_2} - W_{t_1}\,,\, \ldots\,,\, W_{t_K} - W_{t_{K-1}} \Big\}$$

should be independent. In self-explanatory terminology: we stipulate that W have ***independent increments***.

The background noise that we visualize does not change its character with time (except when the temperature changes). Therefore the law of $W_t - W_s$ should not depend on the times s, t individually but only on their difference, the elapsed time $t - s$. In self-explanatory terminology: we stipulate that W be ***stationary***.

[9] Edward Nelson's book, *Dynamical Theories of Brownian Motion* [83], offers a most enjoyable and thorough treatment and opens vistas to higher things.

Subtracting W_0 does not change the differential noises dW_t, so we simplify the situation further by stipulating that $W_0 = 0$.

Let $\delta = var(W_1) = \mathbb{E}[W_1^2]$. The variances of $W_{(k+1)/n} - W_{k/n}$ then must be δ/n, since they are all equal by stationarity and add up to δ by the independence of the increments. Thus the variance of W_q is δq for a rational $q = k/n$. By continuity the variance of W_t is δt, and the stationarity forces the variance of $W_t - W_s$ to be $\delta(t-s)$. ∎

Our heuristics about the cause of the Brownian jitter have led us to a stochastic differential equation, (1.2.1), including a model for the driving term W with rather specific properties: it should have stationary independent increments dW_t distributed as $N(0, \delta \cdot dt)$ and have $W_0 = 0$.

Does such a background noise exist? Yes; see theorem 1.2.2 below. If so, what further properties does it have? Volumes; see, e.g., [48]. How many such noises are there? Essentially one for every *diffusion coefficient* δ (see lemma 1.2.7 on page 16 and exercise 1.2.14 on page 19). They are called *Wiener processes.*

Existence of Wiener Process

What is meant by "Wiener process[8] exists"? It means that there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there lives a family $\{W_t : t \geq 0\}$ of random variables with the properties specified above. The quadruple $\big(\Omega, \mathcal{F}, \mathbb{P}, \{W_t : t \geq 0\}\big)$ is a ***mathematical model*** for the noise envisaged. The case $\delta = 1$ is representative (exercise 1.2.14), so we concentrate on it:

Theorem 1.2.2 (Existence and Continuity of Wiener Process) *(i) There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and on it a family $\{W_t : 0 \leq t < \infty\}$ of random variables that has stationary independent increments, and such that $W_0 = 0$ and the law of the increment $W_t - W_s$ is $N(0, t-s)$.*

*(ii) Given such a family, one may change every W_t on a negligible set in such a way that for every $\omega \in \Omega$ the **path** $t \mapsto W_t(\omega)$ is a continuous function.*

Definition 1.2.3 *Any family $\big\{W_t : t \in [0, \infty)\big\}$ of random variables (defined on some probability space) that has continuous paths and stationary independent increments $W_t - W_s$ with law $N(0, t-s)$, and that is normalized to $W_0 = 0$, is called a **standard Wiener process**.*

A standard Wiener process can be characterized more simply as a continuous martingale W scaled by $W_0 = 0$ and $\mathbb{E}[W_t^2] = t$ (see corollary 3.9.5). In view of the discussion on page 4 it is thus not surprising that it serves as a background noise in the majority of stochastic models for physical, genetic, economic, and other phenomena and plays an important role in harmonic analysis and other branches of mathematics. For example, three-dimensional Wiener process[8] "knows" the zeroes of the ζ-function, and thus

the distribution of the prime numbers – alas, so far it is reluctant to part with this knowledge. Wiener process is frequently called Brownian motion in the literature. We prefer to reserve the name "***Brownian motion***" for the physical phenomenon described in example 1.2.1 and capable of being described to various degrees of accuracy by different mathematical models [83].

Proof of Theorem 1.2.2 (i). To get an idea how we might construct the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the W_t, consider dW as a map that associates with any interval $(s,t]$ the random variable $W_t - W_s$ on Ω, i.e., as a measure on $[0,\infty)$ with values in $L^2(\mathbb{P})$. It is after all in this capacity that the noise W will be used in a stochastic differential equation (see page 5). Eventually we shall need to integrate functions with dW, so we are tempted to extend this measure by linearity to a map $\int \cdot\, dW$ from step functions

$$\phi = \sum_k r_k \cdot 1_{(t_k, t_{k+1}]}$$

on the half-line to random variables in $L^2(\mathbb{P})$ via

$$\int \phi\, dW = \sum_k r_k \cdot (W_{t_{k+1}} - W_{t_k})\ .$$

Suppose that the family $\{W_t : 0 \le t < \infty\}$ has the properties listed in (i). It is then rather easy to check that $\int \cdot\, dW$ extends to a linear isometry U from $L^2[0,\infty)$ to $L^2(\mathbb{P})$ with the property that $U(\phi)$ has a normal law $N(0,\sigma^2)$ with mean zero and variance $\sigma^2 = \int_0^\infty \phi^2(x)\, dx$, and so that functions perpendicular in $L^2[0,\infty)$ have independent images in $L^2(\mathbb{P})$. If we apply U to a basis of $L^2[0,\infty)$, we shall get a sequence (ξ_n) of independent $N(0,1)$ random variables. The verification of these claims is left as an exercise.

We now stand these heuristics on their head and arrive at the

Construction of Wiener Process Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that admits a sequence (ξ_n) of independent identically distributed random variables, each having law $N(0,1)$. This can be had by the following simple construction: prepare countably many copies of $(\mathbb{R}, \mathcal{B}^\bullet(\mathbb{R}), \gamma_1)$ [10] and let $(\Omega, \mathcal{F}, \mathbb{P})$ be their product; for ξ_n take the n^{th} coordinate function. Now pick any orthonormal basis (ϕ_n) of the Hilbert space $L^2[0,\infty)$. Any element f of $L^2[0,\infty)$ can be written uniquely in the form

$$f = \textstyle\sum_{n=1}^\infty a_n \phi_n\ ,$$

with $\|f\|^2_{L^2[0,\infty)} = \sum_{n=1}^\infty a_n^2 < \infty$. So we may define a map Φ by

$$\Phi(f) = \textstyle\sum_{n=1}^\infty a_n \xi_n\ .$$

[10] $\mathcal{B}^\bullet(\mathbb{R})$ is the σ-algebra of Borel sets on the line, and $\gamma_1(dx) = (1/\sqrt{2\pi}) \cdot e^{-x^2/2}\, dx$ is the normalized Gaussian measure, see page 419. For the infinite product see lemma A.3.20.

Φ evidently associates with every class in $L^2[0,\infty)$ an *equivalence class* of square integrable functions in $L^2(\mathbb{P}) = L^2(\Omega, \mathcal{F}, \mathbb{P})$. Recall the argument: the finite sums $\sum_{n=1}^{N} a_n \xi_n$ form a Cauchy sequence in the space $L^2(\mathbb{P})$, because

$$\mathbb{E}\Big[\Big(\sum_{n=M}^{N} a_n \xi_n\Big)^2\Big] = \sum_{n=M}^{N} a_n^2 \leq \sum_{n=M}^{\infty} a_n^2 \xrightarrow[M\to\infty]{} 0\,.$$

Since the space $L^2(\mathbb{P})$ is complete there is a limit in 2-mean; since $L^2(\mathbb{P})$, the space of equivalence classes, is Hausdorff, this limit is unique. Φ is clearly a linear isometry from $L^2[0,\infty)$ *into* $L^2(\mathbb{P})$. It is worth noting here that our recipe Φ does not produce a function but merely an equivalence class modulo $\mathbb{P}$-negligible functions. It is necessary to make some hard estimates to pick a suitable representative from each class, so as to obtain actual random variables (see lemma A.2.37).

Let us establish next that the law of $\Phi(f)$ is $N(0, \|f\|^2_{L^2[0,\infty)})$. To this end note that $f = \sum_n a_n \phi_n$ has the same norm as $\Phi(f)$:

$$\int_0^\infty f^2(t)\,dt = \sum a_n^2 = \mathbb{E}[(\Phi(f))^2]\,.$$

The simple computation

$$\mathbb{E}\Big[e^{i\alpha\Phi(f)}\Big] = \mathbb{E}\Big[e^{i\alpha\sum_n a_n\xi_n}\Big] = \prod_n \mathbb{E}\Big[e^{i\alpha a_n\xi_n}\Big] = e^{-\alpha^2\sum_n a_n^2/2}$$

shows that the characteristic function of $\Phi(f)$ is that of a $N(0, \sum a_n^2)$ random variable (see exercise A.3.45 on page 419). Since the characteristic function determines the law (page 410), the claim follows.

A similar argument shows that if $f_1, f_2, \ldots$ are orthogonal in $L^2[0,\infty)$, then $\Phi(f_1), \Phi(f_2), \ldots$ are not only also orthogonal in $L^2(\mathbb{P})$ but are actually independent:

$$\text{clearly} \quad \big\|\textstyle\sum_k \alpha_k f_k\big\|^2_{L^2[0,\infty)} = \sum_k \alpha_k^2 \cdot \|f_k\|^2_{L^2[0,\infty)}\,,$$

$$\text{whence} \quad \mathbb{E}\Big[e^{i\sum_k \alpha_k \Phi(f_k)}\Big] = \mathbb{E}\Big[e^{i\Phi\left(\sum_k \alpha_k f_k\right)}\Big] = e^{-\left\|\sum_k \alpha_k f_k\right\|^2/2}$$

$$= \prod\nolimits_k e^{-\alpha_k^2\cdot\|f_k\|^2/2} = \prod\nolimits_k \mathbb{E}\Big[e^{i\alpha_k\Phi(f_k)}\Big]\,.$$

This says that the joint characteristic function is the product of the marginal characteristic functions, so the random variables are independent (see exercise A.3.36).

For any $t \geq 0$ let $\dot{W}_t$ be the class $\Phi\big(1_{[0,t]}\big)$ and simply pick a member W_t of $\dot{W}_t$. If $0 \leq s < t$, then $\dot{W}_t - \dot{W}_s = \Phi\big(1_{(s,t]}\big)$ is distributed $N(0, t-s)$ and our family $\{W_t\}$ is stationary. With disjoint intervals being orthogonal functions of $L^2[0,\infty)$, our family has independent increments. ∎

Proof of Theorem 1.2.2 (ii). We start with the following observation: due to exercise A.3.47, the curve $t \mapsto \dot{W}_t$ is continuous from $\mathbb{R}_+$ to the space $L^p(\mathbb{P})$, for any $p < \infty$. In particular, for $p = 4$

$$\mathbb{E}\big[\,|W_t - W_s|^4\,\big] = 4 \cdot |t-s|^2 \,. \tag{1.2.2}$$

Next, in order to have the parameter domain open let us extend the process $\dot{W}_t$ constructed in part (i) of the proof to negative times by $\dot{W}_{-t} = \dot{W}_t$ for $t > 0$. Equality (1.2.2) is valid for any family $\{W_t : t \geq 0\}$ as in theorem 1.2.2 (i). Lemma A.2.37 applies, with $(E,\rho) = (\mathbb{R}, |\ |)$, $p = 4$, $\beta = 1$, $C = 4$: there is a selection $W_t \in \dot{W}_t$ such that the path $t \to W_t(\omega)$ is continuous for all $\omega \in \Omega$. We modify this by setting $W_\bullet(\omega) \equiv 0$ in the negligible set of those points ω where $W_0(\omega) \neq 0$ and then forget about negative times. ∎

Uniqueness of Wiener Measure

A standard Wiener process is, of course, not unique: given the one we constructed above, we paint every element of Ω purple and get a new Wiener process that differs from the old one simply because its domain Ω is different. Less facetious examples are given in exercises 1.2.14 and 1.2.16. What is unique about a Wiener process is its *law* or *distribution*.

Recall – or consult section A.3 for – the notion of the law of a real-valued random variable $f : \Omega \to \mathbb{R}$. It is the measure $f[\mathbb{P}]$ on the codomain of f, $\mathbb{R}$ in this case, that is given by $f[\mathbb{P}](B) \stackrel{\text{def}}{=} \mathbb{P}[f^{-1}(B)]$ on Borels $B \in \mathcal{B}^\bullet(\mathbb{R})$. Now any standard Wiener process $W_\bullet$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can be identified in a natural way with a random variable $\underline{W}$ that has values in the space $\mathscr{C} = C[0,\infty)$ of continuous real-valued functions on the half-line. Namely, $\underline{W}$ is the map that associates with every $\omega \in \Omega$ the function or *path* $w = \underline{W}(\omega)$ whose value at t is $w_t = \overline{W}_t(w) \stackrel{\text{def}}{=} W_t(\omega)$, $t \geq 0$. We also call $\underline{W}$ a *representation* of $W_\bullet$ on path space.[11] It is determined by the equation

$$\overline{W}_t \circ \underline{W}(\omega) = W_t(\omega) \,, \qquad\qquad t \geq 0 \,,\ \omega \in \Omega \,.$$

Wiener measure is the law or distribution of this $\mathscr{C}$-valued random variable $\underline{W}$, and this will turn out to be unique.

Before we can talk about this law, we have to identify the equivalent of the Borel sets $B \subset \mathbb{R}$ above. To do this a little analysis of ***path space*** $\mathscr{C} = C[0,\infty)$ is required. $\mathscr{C}$ has a natural topology, to wit, the topology of uniform convergence on compact sets. It can be described by a metric, for instance,[12]

$$d(w,w') = \sum_{n\in\mathbb{N}} \sup_{0\leq s\leq n} \big|\, w_s - w'_s \,\big| \wedge 2^{-n} \qquad \text{for } w, w' \in \mathscr{C} \,. \tag{1.2.3}$$

[11] "Path space," like "frequency space" or "outer space," may be used without an article.
[12] $a \vee b$ ($a \wedge b$) is the larger (smaller) of a and b.

Exercise 1.2.4 (i) A sequence $(w^{(n)})$ in $\mathscr{C}$ converges uniformly on compact sets to $w \in \mathscr{C}$ if and only if $d(w^{(n)}, w) \to 0$. $\mathscr{C}$ is complete under the metric d.

(ii) $\mathscr{C}$ is Hausdorff, and is separable, i.e., it contains a countable dense subset.

(iii) Let $\{w^{(1)}, w^{(2)}, \ldots\}$ be a countable dense subset of $\mathscr{C}$. Every open subset of $\mathscr{C}$ is the union of balls in the countable collection

$$B_q(w^{(n)}) \stackrel{\text{def}}{=} \left\{ w : d(w, w^{(n)}) < q \right\}, \qquad n \in \mathbb{N}\,,\ 0 < q \in \mathbb{Q}\,.$$

Being separable and complete under a metric that defines the topology makes $\mathscr{C}$ a ***polish space***. The Borel σ-algebra $\mathcal{B}^\bullet(\mathscr{C})$ on $\mathscr{C}$ is, of course, the σ-algebra generated by this topology (see section A.3 on page 391). As to our standard Wiener process W, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and identified with a $\mathscr{C}$-valued map $\underline{W}$ on Ω, it is not altogether obvious that inverse images $\underline{W}^{-1}(B)$ of Borel sets $B \subset \mathscr{C}$ belong to $\mathcal{F}$; yet this is precisely what is needed if the law $\underline{W}[\mathbb{P}]$ of $\underline{W}$ is to be defined, in analogy with the real-valued case, by

$$\underline{W}[\mathbb{P}](B) \stackrel{\text{def}}{=} \mathbb{P}[\underline{W}^{-1}(B)]\,, \qquad B \in \mathcal{B}^\bullet(\mathscr{C})\,.$$

Let us show that they do. To this end denote by $\mathcal{F}^0_\infty[\mathscr{C}]$ the σ-algebra on $\mathscr{C}$ generated by the real-valued functions $\overline{W}_t : w \mapsto w_t$, $t \in [0, \infty)$, the ***evaluation maps***. Since $\overline{W}_t \circ \underline{W} = W_t$ is measurable on $\mathcal{F}_t$, clearly

$$\underline{W}^{-1}(E) \in \mathcal{F}\,, \qquad \forall\, E \in \mathcal{F}^0_\infty[\mathscr{C}]\,. \tag{1.2.4}$$

Let us show next that every ball $B_r(w^{(0)}) \stackrel{\text{def}}{=} \left\{ w :\ d(w, w^{(0)}) < r \right\}$ belongs to $\mathcal{F}^0_\infty[\mathscr{C}]$. To prove this it evidently suffices to show that for fixed $w^{(0)} \in \mathscr{C}$ the map $w \mapsto d(w, w^{(0)})$ is measurable on $\mathcal{F}^0_\infty[\mathscr{C}]$. A glance at equation (1.2.3) reveals that this will be true if for every $n \in \mathbb{N}$ the map $w \mapsto \sup_{0 \le s \le n} |w_s - w^{(0)}_s|$ is measurable on $\mathcal{F}^0_\infty[\mathscr{C}]$. This, however, is clear, since the previous supremum equals the countable supremum of the functions

$$w \mapsto \left| w_q - w^{(0)}_q \right|, \qquad q \in \mathbb{Q}\,,\ q \le n\,,$$

each of which is measurable on $\mathcal{F}^0_\infty[\mathscr{C}]$. We conclude with exercise 1.2.4 (iii) that every open set belongs to $\mathcal{F}^0_\infty[\mathscr{C}]$, and that therefore

$$\mathcal{F}^0_\infty[\mathscr{C}] = \mathcal{B}^\bullet\big(\mathscr{C}\big)\,. \tag{1.2.5}$$

In view of equation (1.2.4) we now know that the inverse image under $\underline{W} : \Omega \to \mathscr{C}$ of a Borel set in $\mathscr{C}$ belongs to $\mathcal{F}$. We are now in position to talk about the image $\underline{W}[\mathbb{P}]$:

$$\underline{W}[\mathbb{P}](B) \stackrel{\text{def}}{=} \mathbb{P}[\underline{W}^{-1}(B)]\,, \qquad B \in \mathcal{B}^\bullet(\mathscr{C})\,.$$

of $\mathbb{P}$ under $\underline{W}$ (see page 405) and to define Wiener measure:

Definition 1.2.5 *The law of a standard Wiener process* $\big(\Omega, \mathcal{F}, \mathbb{P}, W_\bullet\big)$*, that is to say the probability* $\mathbb{W} = \underline{W}[\mathbb{P}]$ *on* $\mathscr{C}$ *given by*

$$\mathbb{W}(B) \stackrel{\text{def}}{=} \underline{W}[\mathbb{P}](B) = \mathbb{P}[\underline{W}^{-1}(B)]\,, \qquad B \in \mathcal{B}^\bullet(\mathscr{C})\,,$$

is called ***Wiener measure****. The topological space* $\mathscr{C}$ *equipped with Wiener measure* $\mathbb{W}$ *on its Borel sets is called* ***Wiener space****. The real-valued random variables on* $\mathscr{C}$ *that map a path* $w \in \mathscr{C}$ *to its value at* t *and that are denoted by* $\overline{W}_t$ *above, and often simply by* w_t*, constitute the* ***canonical Wiener process****.*[8]

Exercise 1.2.6 The name is justified by the observation that the quadruple $(\mathscr{C}, \mathcal{B}^\bullet(\mathscr{C}), \mathbb{W}, \{\overline{W}_t\}_{0\le t<\infty})$ is a standard Wiener process.

Definition 1.2.5 makes sense only if any two standard Wiener processes have the same distribution on $\mathscr{C}$. Indeed they do:

Lemma 1.2.7 *Any two standard Wiener processes have the same law.*

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P}, W_\bullet)$ and $(\Omega', \mathcal{F}', \mathbb{P}', W'_\bullet)$ be two standard Wiener processes and let $\mathbb{W}$ denote the law of $W_\bullet$. Consider a complex-valued function on $\mathscr{C} = C[0,\infty)$ of the form

$$\phi(w) = \exp\Big(i\sum_{k=1}^{K} r_k(w_{t_k} - w_{t_{k-1}})\Big) = \exp\Big(i\sum_{k=1}^{K} r_k(\overline{W}_{t_k}(w) - \overline{W}_{t_{k-1}}(w))\Big)\,, \tag{1.2.6}$$

$r_k \in \mathbb{R}$, $0 = t_0 < t_1 < \ldots < t_K$. Its $\mathbb{W}$-integral can be computed:

$$\int \phi(w)\,\mathbb{W}(dw) = \int \exp\Big(i\sum_{k=1}^{K} r_k(\overline{W}_{t_k}\circ\underline{W} - \overline{W}_{t_{k-1}}\circ\underline{W})\Big)\,d\mathbb{P}$$

by independence:
$$= \prod_{k=1}^{K}\int \exp\big(ir_k(W_{t_k} - W_{t_{k-1}})\big)\,d\mathbb{P}$$

$$= \prod_{k}^{K}\int_{-\infty}^{\infty} e^{ir_k x}\cdot\frac{e^{-x^2/2(t_k-t_{k-1})}}{\sqrt{2\pi(t_k-t_{k-1})}}\,dx$$

by exercise A.3.45:
$$= \prod_{k}^{K} e^{-r_k^2(t_k-t_{k-1})/2}\,.$$

The same calculation can be done for $\mathbb{P}'$ and shows that its distribution $\mathbb{W}'$ under $\underline{W}'$ coincides with $\mathbb{W}$ on functions of the form (1.2.6). Now note that these functions are bounded, and that their collection $\mathcal{M}$ is closed under multiplication and complex conjugation and generates the same σ-algebra as the collection $\{\overline{W}_t :\ t \ge 0\}$, to wit $\mathcal{F}^0_\infty[\mathscr{C}] = \mathcal{B}^\bullet\big(C[0,\infty)\big)$. An application of the Monotone Class Theorem in the form of exercise A.3.5 finishes the proof.

Namely, the vector space $\mathcal{V}$ of bounded complex-valued functions on $\mathscr{C}$ on which $\mathbb{W}$ and $\mathbb{W}'$ agree is sequentially closed and contains $\mathcal{M}$, so it contains every bounded $\mathcal{B}^\bullet\big(C[0,\infty)\big)$-measurable function. ∎

Non-Differentiability of the Wiener Path

The main point of the introduction was that a novel integration theory is needed because the driving term of stochastic differential equations occurring most frequently, Wiener process, has paths of infinite variation. We show this now. In fact,[2] since a function that has finite variation on some interval is differentiable at almost every point of it, the claim is immediate from the following result:

Theorem 1.2.8 (Wiener) *Let W be a standard Wiener process on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. Except for the points ω in a negligible subset $\mathcal{N}$ of Ω, the path $t\mapsto W_t(\omega)$ is nowhere differentiable.*

Proof [28]. Suppose that $t\mapsto W_t(\omega)$ is differentiable at some instant s. There exists a $K\in\mathbb{N}$ with $s<K-1$. There exist $M,N\in\mathbb{N}$ such that for all $n\geq N$ and all $t\in(s-5/n,s+5/n)$, $|W_t(\omega)-W_s(\omega)|\leq M\cdot|t-s|$. Consider the first three consecutive points of the form j/n, $j\in\mathbb{N}$, in the interval $(s,s+5/n)$. The triangle inequality produces

$$|W_{\frac{j+1}{n}}(\omega)-W_{\frac{j}{n}}(\omega)|\leq|W_{\frac{j+1}{n}}(\omega)-W_s(\omega)|+|W_{\frac{j}{n}}(\omega)-W_s(\omega)|\leq 7M/n$$

for each of them. The point ω therefore lies in the set

$$\mathcal{N}=\bigcup_K\bigcup_M\bigcup_N\bigcap_{n\geq N}\bigcup_{k\leq K\cdot n}\bigcap_{j=k}^{k+2}\Big[|W_{\frac{j+1}{n}}-W_{\frac{j}{n}}|\leq 7M/n\Big]\ .$$

To prove that $\mathcal{N}$ is negligible it suffices to show that the quantity

$$\begin{aligned}Q&\stackrel{\text{def}}{=}\mathbb{P}\Big[\bigcap_{n\geq N}\bigcup_{k\leq K\cdot n}\bigcap_{j=k}^{k+2}\Big[|W_{\frac{j+1}{n}}-W_{\frac{j}{n}}|\leq 7M/n\Big]\Big]\\&\leq\liminf_{n\to\infty}\mathbb{P}\Big[\bigcup_{k\leq K\cdot n}\bigcap_{j=k}^{k+2}\Big[|W_{\frac{j+1}{n}}-W_{\frac{j}{n}}|\leq 7M/n\Big]\Big]\end{aligned}$$

vanishes. To see this note that the events

$$\Big[|W_{\frac{j+1}{n}}-W_{\frac{j}{n}}|\leq 7M/n\Big]\ ,\qquad\qquad j=k,k+1,k+2,$$

are independent and have probability

$$\begin{aligned}\mathbb{P}\Big[|W_{\frac{1}{n}}|\leq 7M/n\Big]&=\frac{1}{\sqrt{2\pi/n}}\int_{-7M/n}^{+7M/n}e^{-x^2n/2}\,dx\\&=\frac{1}{\sqrt{2\pi}}\int_{-7M/\sqrt{n}}^{+7M/\sqrt{n}}e^{-\xi^2/2}\,d\xi\leq\frac{14M}{\sqrt{2\pi n}}\ .\end{aligned}$$

Thus

$$Q\leq\liminf_{n\to\infty}K\cdot n\cdot\left(\frac{const}{\sqrt{n}}\right)^3=0\ .$$

∎

Remark 1.2.9 In the beginning of this section Wiener process[8] was motivated as a driving term for a stochastic differential equation describing physical Brownian motion. One could argue that the non-differentiability of the paths was a result of overly much idealization. Namely, the total momentum imparted to the pollen (in our billiard ball model) during the time-interval $[0, t]$ by collisions with the gas molecules is in reality a function of finite variation in t. In fact, it is constant between kicks and jumps at a kick by the momentum imparted; it is, in particular, not continuous. If the interval dt is small enough, there will not be any kicks at all. So the assumption that the differential of the driving term is distributed $N(0, dt)$ is just too idealistic. It seems that one should therefore look for a better model for the driver, one that takes the microscopic aspects of the interaction between pollen and gas molecules into account.

Alas, no one has succeeded so far, and there is little hope: first, the total variation of a momentum transfer during $[0, t]$ turns out to be huge, since it does not take into account the cancellation of kicks in opposite directions. This rules out any reasonable estimates for the convergence of any scheme for the solution of the stochastic differential equation driven by a more accurately modeled noise, in terms of this variation. Also, it would be rather cumbersome to keep track of the statistics of such a process of finite variation if its structure between any two of the huge number of kicks is taken into account.

We shall therefore stick to Wiener process as a model for the driver in the model for Brownian motion and show that the statistics of the solution of equation (1.2.1) on page 10 are close to the statistics of the solution of the corresponding equation driven by a finite variation model for the driver, provided the number of kicks is sufficiently large (exercise A.4.14). We shall return to this circle of problems several times, next in example 2.5.26 on page 79.

Supplements and Additional Exercises

Fix a standard Wiener process $W_\bullet$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any s let $\mathcal{F}_s^0[W_\bullet]$ denote the σ-algebra generated by the collection $\{W_r : 0 \le r \le s\}$. That is to say, $\mathcal{F}_s^0[W_\bullet]$ is the smallest σ-algebra on which the $W_r : r \le s$ are all measurable. Intuitively, $\mathcal{F}_t^0[W_\bullet]$ contains all information about the random variables W_s that can be observed up to and including time t. The collection

$$\mathcal{F}_\bullet^0[W_\bullet] = \{\mathcal{F}_s^0[W_\bullet] : 0 \le s < \infty\}$$

of σ-algebras is called the ***basic filtration of the Wiener process*** $W_\bullet$.

Exercise 1.2.10 $\mathcal{F}_s^0[W_\bullet]$ increases with s and is the σ-algebra generated by the increments $\{W_r - W_{r'} : 0 \le r, r' \le s\}$. For $s < t$, $W_t - W_s$ is independent of $\mathcal{F}_s^0[W_\bullet]$. Also, for $0 \le s < t < \infty$,

$$\mathbb{E}[W_t|\mathcal{F}_s^0[W_\bullet]] = W_s \quad \text{and} \quad \mathbb{E}\left[W_t^2 - W_s^2|\mathcal{F}_s^0[W_\bullet]\right] = t - s\,. \tag{1.2.7}$$

Equations (1.2.7) say that both W_t and $W_t^2 - t$ are *martingales*[4] on $\mathcal{F}_\bullet^0[W_\bullet]$. Together with the continuity of the path they determine the law of the whole process W uniquely. This fact, Lévy's characterization of Wiener process,[8] is proven most easily using stochastic integration, so we defer the proof until corollary 3.9.5. In the meantime here is a characterization that is just as useful:

Exercise 1.2.11 Let $X_\bullet = (X_t)_{t\geq 0}$ be a real-valued process with continuous paths and $X_0 = 0$, and denote by $\mathcal{F}_\bullet^0[X_\bullet]$ its basic filtration – $\mathcal{F}_s^0[X_\bullet]$ is the σ-algebra generated by the random variables $\{X_r : 0 \leq r \leq s\}$. Note that it contains the basic filtration $\mathcal{F}_\bullet^0[M_\bullet^z]$ of the process $M_\bullet^z : t \mapsto M_t^z \stackrel{\text{def}}{=} e^{zX_t - z^2 t/2}$ whenever $0 \neq z \in \mathbb{C}$. The following are equivalent:
(i) X is a standard Wiener process; (ii) the M^z are martingales[4] on $\mathcal{F}_\bullet^0[X_\bullet]$; (iii) $M^\alpha : t \mapsto e^{i\alpha X_t + \alpha^2 t/2}$ is an $\mathcal{F}_\bullet^0[M_\bullet^\alpha]$-martingale for every real α.

Exercise 1.2.12 For any bounded Borel function ϕ and $s < t$

$$\mathbb{E}\left[\phi(W_t)|\mathcal{F}_s^0[W_\bullet]\right] = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{+\infty} \phi(y) \cdot e^{-(y-W_s)^2/2(t-s)}\, dy\ . \tag{1.2.8}$$

Exercise 1.2.13 For any bounded Borel function ϕ on $\mathbb{R}$ and $t > 0$ define the function $T_t\phi$ by $T_0\phi = \phi$ if $t = 0$, and for $t > 0$ by

$$(T_t\phi)(x) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} \phi(x+y) e^{-y^2/2t}\, dy\ .$$

Then T_t is a semigroup (i.e., $T_t \circ T_s = T_{t+s}$) of positive (i.e., $\phi \geq 0 \implies T_t\phi \geq 0$) linear operators with $T_0 = I$ and $T_t 1 = 1$, whose restriction to the space $C_0(\mathbb{R})$ of bounded continuous functions that vanish at infinity is continuous in the sup-norm topology. Rewrite equation (1.2.8) as

$$\mathbb{E}\left[\phi(W_t)|\mathcal{F}_s^0[W_\bullet]\right] = (T_{t-s}\phi)(W_s)\ .$$

Exercise 1.2.14 Let $(\Omega, \mathcal{F}, \mathbb{P}, W_\bullet)$ be a standard Wiener process. (i) For every $a > 0$, $\sqrt{a} \cdot W_{t/a}$ is a standard Wiener process. (ii) $t \mapsto t \cdot W_{1/t}$ is a standard Wiener process. (iii) For $\delta > 0$, the family $\{\sqrt{\delta} W_t : t \geq 0\}$ is a background noise as in example 1.2.1, but with diffusion coefficient δ.

Exercise 1.2.15 (*d*-Dimensional Wiener Process) (i) Let $1 \leq n \in \mathbb{N}$. There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family $(\boldsymbol{W}_t : 0 \leq t < \infty)$ of $\mathbb{R}^d$-valued random variables on it with the following properties:

(a) $\boldsymbol{W}_0 = 0$.

(b) $\boldsymbol{W}_\bullet$ has independent increments. That is to say, if $0 = t_0 < t_1 < \ldots < t_K$ are consecutive instants, then the corresponding family of consecutive increments

$$\left\{ \boldsymbol{W}_{t_1} - \boldsymbol{W}_{t_0}\ ,\ \boldsymbol{W}_{t_2} - \boldsymbol{W}_{t_1}\ ,\ \ldots\ ,\ \boldsymbol{W}_{t_K} - \boldsymbol{W}_{t_{K-1}} \right\}$$

is independent.

(c) The increments $\boldsymbol{W}_t - \boldsymbol{W}_s$ are stationary and have normal law with covariance matrix

$$\int (W_t^\eta - W_s^\eta)(W_t^\theta - W_s^\theta)\, d\mathbb{P} = (t-s) \cdot \delta^{\eta\theta}\ .$$

Here $\delta^{\eta\theta} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \eta = \theta \\ 0 & \text{if } \eta \neq \theta \end{cases}$ is the ***Kronecker delta***.

(ii) Given such a family, one may change every $\boldsymbol{W}_t$ on a negligible set in such a way that for every $\omega \in \boldsymbol{W}$ the path $t \mapsto \boldsymbol{W}_t(\omega)$ is a continuous function from $[0, \infty)$

to $\mathbb{R}^d$. Any family $\{\boldsymbol{W}_t : t \in [0,\infty)\}$ of $\mathbb{R}^d$-valued random variables (defined on some probability space) that has the three properties (a)–(c) and also has continuous paths is called a ***standard d-dimensional Wiener process.***

(iii) The law of a standard d-dimensional Wiener process is a measure defined on the Borel subsets of the topological space

$$\mathscr{C}^d = C_{\mathbb{R}^d}[0,\infty)$$

of continuous paths $w : [0,\infty) \to \mathbb{R}^d$ and is unique. It is again called Wiener measure and is also denoted by $\mathbb{W}$.

(iv) An $\mathbb{R}^d$-valued process $\big(\Omega, \mathcal{F}, (Z_t)_{0\le t<\infty}\big)$ with continuous paths whose law is Wiener measure is a standard d-dimensional Wiener process.

(v) Define the basic filtration $\mathcal{F}_s^0[\boldsymbol{W}_\bullet]$ and redo exercises 1.2.10–1.2.13 after proper reformulation.

Exercise 1.2.16 (The Brownian Sheet) A ***random sheet*** is a family $S_{\eta,t}$ of random variables on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ indexed by the points of some domain in $\mathbb{R}^2$, say of $\check{\boldsymbol{H}} \stackrel{\text{def}}{=} \{(\eta, t) : \eta \in \mathbb{R}\ ,\ 0 \le t < \infty\}$. Any two points $z_1 = (\eta_1, t_1)$ and $z_2 = (\eta_2, t_2)$ in $\check{\boldsymbol{H}}$ with $\eta_1 \le \eta_2$ and $0 \le t_1 \le t_2$ determine a rectangle $(z_1, z_2] = (\eta_1, \eta_2] \times (t_1, t_2]$, and with it goes the "increment"

$$dS\big((z_1, z_2]\big) = S_{\eta_2,t_2} - S_{\eta_2,t_1} - S_{\eta_1,t_2} + S_{\eta_1,t_1}\ .$$

A ***Brownian sheet*** or ***Wiener sheet*** on $\check{\boldsymbol{H}}$ is a random sheet with the following properties:

$W_{0,0} = 0$; if $R_1, \dots R_K$ are disjoint rectangles, then the corresponding family

$$\big\{dS(R_1), \dots, dS(R_K)\big\}$$

of random variables is independent; for any rectangle $\check{\boldsymbol{H}}$, the law of $dS(R)$ is $N\big(0, \lambda(R)\big)$, $\lambda(R)$ being the Lebesgue measure of R.

Show: there exists a Brownian sheet; its paths, or better, *sheets*, $(\eta, t) \mapsto S_{\eta,t}(\omega)$ can be chosen to be continuous for every $\omega \in \Omega$; the law of a Brownian sheet is a probability defined on all Borel subsets of the polish space $C(\check{\boldsymbol{H}})$ of continuous functions from $\check{\boldsymbol{H}}$ to the reals and is unique; for fixed η, $t \mapsto \eta^{-1/2} S_{\eta,t}$ is a standard Wiener process.

Exercise 1.2.17 Define the Brownian box and show that it is continuous.

1.3 The General Model

Wiener process is not the only driver for stochastic differential equations, albeit the most frequent one. For instance, the solution of a stochastic differential equation can be used to drive yet another one; even if it is not used for this purpose, it can best be analyzed in its capacity as a driver. We are thus automatically led to consider the class of all *drivers* or *integrators*.

As long as the integrators are Wiener processes or solutions of stochastic differential equations driven by Wiener processes, or are at least continuous, we can take for the underlying probability space Ω the path space $\mathscr{C}$ of the previous section (exercise 1.2.6). Recall how the uniqueness proof of the law of a Wiener process was facilitated greatly by the polish topology on $\mathscr{C}$. Now there are systems that should be modeled by drivers having jumps, for

instance, the signal from a Geiger counter or a stock price. The corresponding space of trajectories does not consist of continuous paths anymore. After some analysis we shall see in section 2.3 that the appropriate path space Ω is the space $\mathscr{D}$ of right-continuous paths with left limits. The probabilistic analysis leads to estimates involving the so-called maximal process, which means that the naturally arising topology on $\mathscr{D}$ is again the topology of uniform convergence on compacta. However, under this topology $\mathscr{D}$ fails to be polish because it is not separable, and the relation between measurability and topology is not so nice and "tight" as in the case $\mathscr{C}$. Skorohod has given a useful polish topology on $\mathscr{D}$, which we shall describe later (section A.7). However, this topology is not compatible with the vector space structure of $\mathscr{D}$ and thus does not permit the use of arguments from Fourier analysis, as in the uniqueness proof of Wiener measure.

These difficulties can, of course, only be sketched here, lest we never reach our goal of solving stochastic differential equations. Identifying them has taken probabilists many years, and they might at this point not be too clear in the reader's mind. So we shall from now on follow the French School and mostly disregard topology. To identify and analyze general integrators we shall distill a general mathematical model directly from the heuristic arguments of section 1.1. It should be noted here that when a specific physical, financial, etc., system is to be modeled by specific assumptions about a driver, a model for the driver has to be constructed (as we did for Wiener process,[8] the driver of Brownian motion) and shown to fit this general mathematical model. We shall give some examples of this later (page 267).

Before starting on the general model it is well to *get acquainted with some notations and conventions* laid out in the beginning of appendix A on page 363 that are fairly but not altogether standard and are designed to cut down on verbiage and to enhance legibility.

Filtrations on Measurable Spaces

Now to the general probabilistic model suggested by the heuristics of section 1.1. First we need a probability space on which the random variables X_t, Z_t, etc., of section 1.1 are realized as functions – so we can apply functional calculus – and a notion of *past* or *history* (see page 6). Accordingly, we stipulate that we are given a ***filtered measurable space*** on which everything of interest lives. This is a pair $(\Omega, \mathcal{F}_\bullet)$ consisting of a set Ω and an increasing family

$$\mathcal{F}_\bullet = \{\mathcal{F}_t\}_{0 \le t < \infty}$$

of σ-algebras on Ω. It is convenient to begin the reckoning of time at $t = 0$; if the starting time is another finite time, a linear scaling will reduce the situation to this case. It is also convenient to end the reckoning of time at ∞. The reader interested in only a finite time-interval $[0, u)$ can use

everything said here simply by reading the symbol ∞ as another name for his ultimate time u of interest.

To say that $\mathcal{F}_\bullet$ is increasing means of course that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $0 \leq s \leq t$. The family $\mathcal{F}_\bullet$ is called a ***filtration*** or ***stochastic basis*** on Ω. The intuitive meaning of it is this: Ω is the set of all evolutions that the world or the system under consideration might take, and $\mathcal{F}_t$ models the collection of "all events that will be observable by the time t," the "history at time t." We close the filtration at ∞ with three objects: first there are

the *algebra* of sets $\qquad \mathcal{A}_\infty \stackrel{\text{def}}{=} \bigcup_{0\leq t<\infty} \mathcal{F}_t$

and the σ-*algebra* $\qquad \mathcal{F}_\infty \stackrel{\text{def}}{=} \bigvee_{0\leq t<\infty} \mathcal{F}_t$

that it generates. Lastly there is the universal completion $\mathcal{F}_\infty^*$ of $\mathcal{F}_\infty$ – see page 407 of appendix A.

A ***random variable*** is simply a universally (i.e., $\mathcal{F}_\infty^*$-) measurable function on Ω.

The filtration $\mathcal{F}_\bullet$ is ***universally complete*** if $\mathcal{F}_t$ is universally complete at any instant $t < \infty$. We shall eventually require that $\mathcal{F}_\bullet$ have this and further properties.

The Base Space

The noises and other processes of interest are functions on the ***base space***

$$\boldsymbol{B} \stackrel{\text{def}}{=} [0,\infty) \times \Omega\ .$$

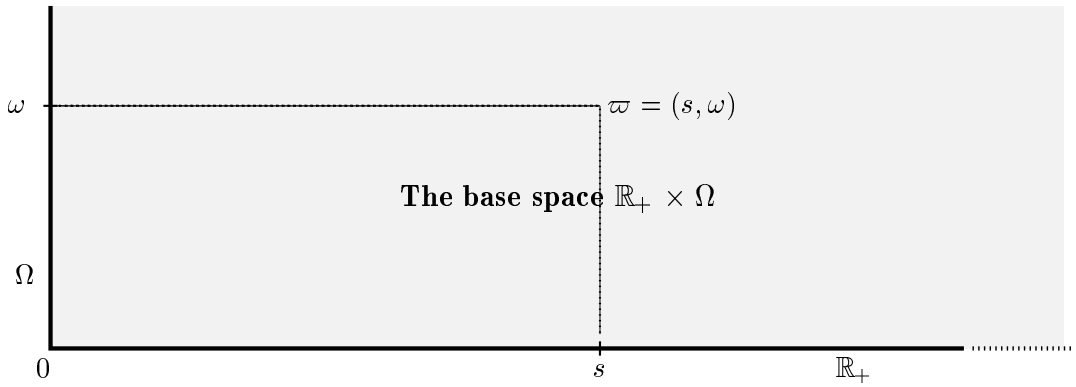

Figure 1.1 The base space

Its typical point is a pair (s,ω), which will frequently be denoted by ϖ. The spirit of this exposition is to reduce stochastic analysis to the analysis of real-valued functions on $\boldsymbol{B}$. The base space has a rather rich structure, being a product whose fibers $\{s\} \times \Omega$ carry finer and finer σ-algebras $\mathcal{F}_s$ as time s increases. This structure gives rise to quite a bit of terminology, which we will be discussing for a while. Fortunately, most notions attached to a filtration are quite intuitive.

Processes

Processes are simply functions[13] on the base space $\boldsymbol{B}$. We are mostly concerned with processes that take their values in the reals $\mathbb{R}$ or in the extended reals $\overline{\mathbb{R}}$ (see item A.1.2 on page 363). So unless the range is explicitly specified differently, ***a process is numerical***. A process is ***measurable*** if it is measurable on the product σ-algebra $\mathcal{B}^\bullet[0,\infty) \otimes \mathcal{F}^*_\infty$ on $\mathbb{R}_+ \times \Omega$.

It is customary to write $Z_s(\omega)$ for the value of the process Z at the point $\varpi = (s,\omega) \in \boldsymbol{B}$, and to denote by Z_s the function $\omega \mapsto Z_s(\omega)$:

$$Z_s : \omega \mapsto Z_s(\omega) , \qquad s \in \mathbb{R}_+ .$$

The process Z is ***adapted*** to the filtration $\mathcal{F}_\bullet$ if at every instant t the random variable Z_t is measurable on $\mathcal{F}_t$; one then writes

$$Z_t \in \mathcal{F}_t .$$

In other words, the symbol $\in$ is not only shorthand for "is an element of" but also for "is measurable on" (see page 391).

The ***path*** or ***trajectory*** of the process Z, on the other hand, is the function

$$Z_\bullet(\omega) : s \mapsto Z_s(\omega) , \qquad 0 \le s < \infty ,$$

one for each $\omega \in \Omega$. A statement such as "Z is ***continuous*** (***left-continuous***, ***right-continuous*** , ***increasing***, ***of finite variation***, etc.)" means that every path of Z has this property.

Stopping a process is a useful and ubiquitous tool. The process Z stopped at time t is the function[12] $(s,\omega) \mapsto Z_{s\wedge t}(\omega)$ and is denoted by Z^t. After time t its path is constant with value $Z_t(\omega)$.

Remark 1.3.1 Frequently the only randomness of interest is the one introduced by some given process Z of interest. Then one appropriate filtration is the ***basic filtration*** $\mathcal{F}^0_\bullet[Z] = \{\mathcal{F}^0_t[Z] : 0 \le t < \infty\}$ of Z. $\mathcal{F}^0_t[Z]$ is the σ-algebra generated by the random variables $\{Z_s : 0 \le s \le t\}$. An instance of this was considered in exercise 1.2.10. We shall see soon (pages 37–40) that there are more convenient filtrations, even in this simple case.

Exercise 1.3.2 The projection on Ω of a measurable subset of the base space is universally measurable. A measurable process has Borel-measurable paths.

Exercise 1.3.3 A process Z is adapted to its basic filtration $\mathcal{F}^0_\bullet[Z]$. Conversely, if Z is adapted to the filtration $\mathcal{F}_\bullet$, then $\mathcal{F}^0_t[Z] \subseteq \mathcal{F}_t$ for all t.

---

[13] The reader has no doubt met before the propensity of probabilists to give new names to everyday mathematical objects – for instance, calling the elements of Ω outcomes, the subsets events, the functions on Ω random variables, etc. This is meant to help intuition but sometimes obscures the distinction between a physical system and its mathematical model.

Wiener Process Revisited On the other hand, it occurs that a Wiener process W is forced to live together with other processes on a filtration $\mathcal{F}_\bullet$ larger than its own basic one (see, e.g., example 5.5.1). A modicum of compatibility is usually required:

Definition 1.3.4 *W is **standard Wiener process on the filtration** $\mathcal{F}_\bullet$ if it is adapted to $\mathcal{F}_\bullet$ and $W_t - W_s$ is independent of $\mathcal{F}_s$ for $0 \le s < t < \infty$.*

See corollary 3.9.5 on page 160 for Lévy's characterization of standard Wiener process on a filtration.

Right- and Left-Continuous Processes Let $\mathscr{D}$ denote the collection of all paths $[0,\infty) \to \mathbb{R}$ that are right-continuous and have finite left limits at all instants $t \in \mathbb{R}_+$, and $\mathscr{L}$ the collection of paths that are left-continuous and have right limits in $\mathbb{R}$ at all instants. A path in $\mathscr{D}$ is also called[14] ***càdlàg*** and a path in $\mathscr{L}$ ***càglàd***. The paths of $\mathscr{D}$ and $\mathscr{L}$ have discontinuities only where they jump; they do not oscillate. Most of the processes that we have occasion to consider are adapted and have paths in one or the other of these classes. They deserve their own symbols: the family of adapted processes whose paths are right-continuous and have left limits is denoted by $\mathfrak{D} = \mathfrak{D}[\mathcal{F}_\bullet]$, and the family of adapted processes whose paths are left-continuous and have right limits is denoted by $\mathfrak{L} = \mathfrak{L}[\mathcal{F}_\bullet]$. Clearly $\mathscr{C} = \mathscr{L} \cap \mathscr{D}$, and $\mathfrak{C} = \mathfrak{L} \cap \mathfrak{D}$ is the collection of continuous adapted processes.

The Left-Continuous Version $X_{\bullet-}$ of a right-continuous process X with left limits has at the instant t the value

$$X_{t-} \stackrel{\text{def}}{=} \begin{cases} 0 & \text{for } t = 0\,, \\ \lim\limits_{t>s\to t} X_s & \text{for } 0 < t \le \infty\,. \end{cases}$$

Clearly $X_{\bullet-} \in \mathfrak{L}$ whenever $X \in \mathfrak{D}$. Note that the left-continuous version is forced to have the value zero at the instant zero. Given an $X \in \mathfrak{L}$ we might – but seldom will – consider its ***right-continuous version*** $X_{\bullet+}$:

$$X_{t+} \stackrel{\text{def}}{=} \lim_{t<u\downarrow t} X_u\,, \qquad 0 \le t < \infty\,.$$

If $X \in \mathfrak{D}$, then taking the right-continuous version of $X_{\bullet-}$ leads back to X. But if $X \in \mathfrak{L}$, then the left-continuous version of $X_{\bullet+}$ differs from X at $t = 0$, unless X happens to vanish at that instant. This slightly unsatisfactory lack of symmetry is outweighed by the simplification of the bookkeeping it affords in Itô's formula and related topics (section 4.2). Here is a mnemonic device: imagine that all processes have the value 0 for strictly negative times; this forces $X_{0-} = 0$.

[14] càdlàg is an acronym for the French "continu à droite, limites à gauche" and càglàd for "continu à gauche, limites à droite." Some authors write "***corlol***" and "***collor***" and others write "***rcll***" and "***lcrl***." "càglàd," though of French origin, is pronounceable.

The Jump Process ΔX of a right-continuous process X with left limits is the difference between itself and its left-continuous version:

$$\Delta X_t \stackrel{\text{def}}{=} \big(X_t - X_{t-}\big) = \begin{cases} X_0 & \text{for } t = 0, \\ X_t - X_{t-} & \text{for } t > 0. \end{cases}$$

$\Delta X_\bullet$ is evidently adapted to the basic filtration $\mathcal{F}^0_\bullet[X]$ of X.

Progressive Measurability The adaptedness of a process Z reflects the idea that at no instant t should a look into the future be necessary to determine the value of the random variable Z_t. It is still thinkable that some property of the whole path of Z up to time t depends on more than the information in $\mathcal{F}_t$. (Such a property would have to involve the values of Z at uncountably many instants, of course.) Progressive measurability rules out such a contingency: Z is ***progressively measurable*** if for every $t \in \mathbb{R}$ the stopped process Z^t is measurable on $\mathcal{B}^\bullet[0,\infty) \otimes \mathcal{F}_t$. This is the the same as saying that the restriction of Z to $[0,t] \times \Omega$ is measurable on $\mathcal{B}^\bullet[0,t] \otimes \mathcal{F}_t$ and means that any measurable information about the whole path up to time t is contained in $\mathcal{F}_t$.

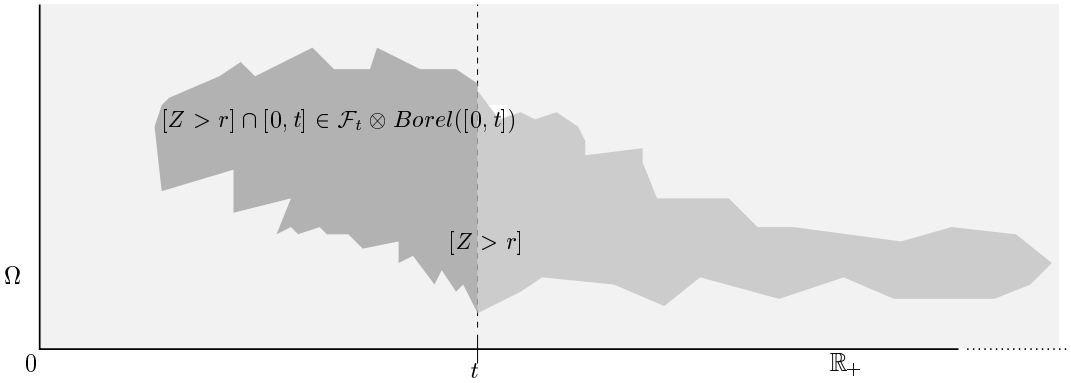

Figure 1.2 Progressive measurability

Proposition 1.3.5 There is some interplay between the notions above:
(i) A progressively measurable process is adapted.
(ii) A left- or right-continuous adapted process is progressively measurable.
(iii) The progressively measurable processes form a sequentially closed family.

Proof. (i): Z_t is the composition of $\omega \mapsto (t,\omega)$ with Z. (ii): If Z is left-continuous and adapted, set

$$Z^{(n)}_s(\omega) = Z_{\frac{k}{n}}(\omega) \quad \text{for} \quad \frac{k}{n} \le s < \frac{k+1}{n}\,.$$

Clearly $Z^{(n)}$ is progressively measurable. Also, $Z^{(n)}(\varpi) \xrightarrow[n\to\infty]{} Z(\varpi)$ at every point $\varpi = (s,\omega) \in \boldsymbol{B}$. To see this, let s_n denote the largest rational of the form k/n less than or equal to s. Clearly $Z^{(n)}(\varpi) = Z_{s_n}(\omega)$ converges to $Z_s(\omega) = Z(\varpi)$.

Suppose now that Z is right-continuous and fix an instant t. The stopped process Z^t is evidently the pointwise limit of the functions

$$Z_s^{(n)}(\omega) = \sum\nolimits_{k=0}^{+\infty} Z_{\frac{k+1}{n}\wedge t}(\omega)\cdot 1_{\left(\frac{k}{n},\frac{k+1}{n}\right]}(s)\ ,$$

which are measurable on $\mathcal{B}^\bullet[0,\infty)\otimes\mathcal{F}_t$. (iii) is left to the reader. ∎

The Maximal Process $Z^\star$ of a process $Z : \boldsymbol{B} \to \overline{\mathbb{R}}$ is defined by

$$Z_t^\star = \sup_{0\le s\le t} |Z_s|\ , \qquad 0 \le t \le \infty\ .$$

This is a supremum over uncountably many indices and so is not in general expected to be measurable. However, when Z is progressively measurable and the filtration is universally complete, then $Z^\star$ is again progressively measurable. This is shown in corollary A.5.13 with the help of some capacity theory. We shall deal mostly with processes Z that are left- or right-continuous, and then we don't need this big cannon. $Z^\star$ is then also left- or right-continuous, respectively, and if Z is adapted, then so is $Z^\star$, inasmuch as it suffices to extend the supremum in the definition of $Z_t^\star$ over instants s in the countable set

$$\mathbb{Q}^t \stackrel{\text{def}}{=} \{q\in\mathbb{Q}:\ 0\le q<t\}\cup\{t\}\ .$$

A path that has finite right and left limits in $\mathbb{R}$ is bounded on every bounded interval, so the maximal process of a process in $\mathfrak{D}$ or in $\mathfrak{L}$ is finite at all instants.

Exercise 1.3.6 The maximal process $W^\star$ of a standard Wiener process almost surely increases without bound.

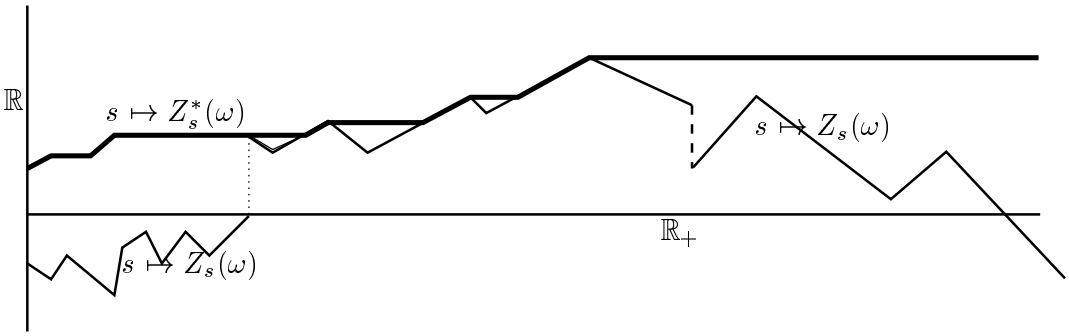

Figure 1.3 A path and its maximal path

The Limit at Infinity of a process Z is the random variable $\lim_{t\to\infty} Z_t(\omega)$, provided this limit exists almost surely. For consistency's sake it should be and is denoted by $Z_{\infty-}$. It is convenient and unambiguous to use also the notation Z_∞:

$$Z_\infty = Z_{\infty-} \stackrel{\text{def}}{=} \lim_{t\to\infty} Z_t\ .$$

The maximal process $Z^\star$ always has a limit $Z^\star_\infty$, possibly equal to $+\infty$ on a large set. If Z is adapted and right-continuous, say, then Z_∞ is evidently measurable on $\mathcal{F}_\infty$.

Stopping Times and Stochastic Intervals

Definition 1.3.7 *A **random time** is a universally measurable function on Ω with values in $[0,\infty]$. A random time T is a **stopping time** if*

$$[T \le t] \in \mathcal{F}_t \qquad \forall\, t \in \mathbb{R}_+ \,. \tag{$*$}$$

*This notion depends on the filtration $\mathcal{F}_\bullet$; and if this dependence must be stressed, then T is called an **$\mathcal{F}_\bullet$-stopping time**. The collection of all stopping times is denoted by $\mathfrak{T}$, or by $\mathfrak{T}[\mathcal{F}_\bullet]$ if the filtration needs to be made explicit.*

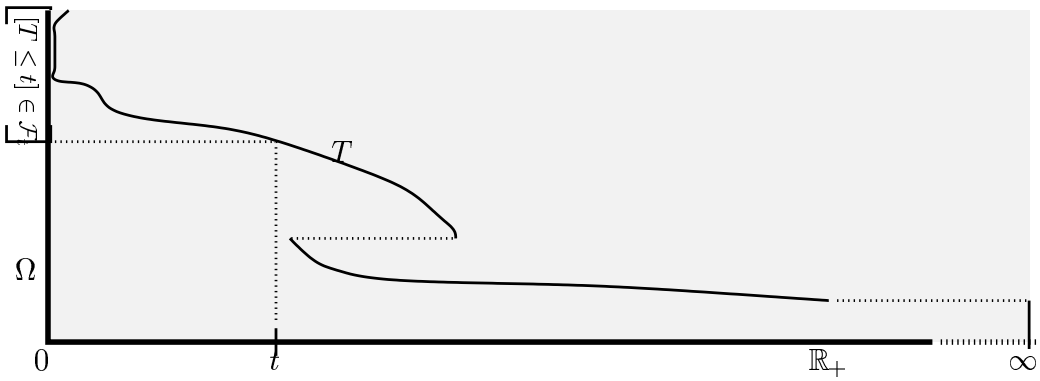

Figure 1.4 Graph of a stopping time

Condition $(*)$ expresses the idea that at the instant t no look into the future is necessary in order to determine whether the time T has arrived. In section 1.1, for example, we were led to consider the first time T our space probe hit the moon. This time evidently depends on chance and is thus a random time. Moreover, the event $[T \le t]$ that the probe hits the moon at or before the instant t can certainly be determined if the history $\mathcal{F}_t$ of the universe up to this instant is known: in our stochastic model, T should turn out to be a stopping time. If the probe never hits the moon, then the time T should be $+\infty$, as $+\infty$ is by general convention the infimum of the void set of numbers. This explains why a stopping time is permitted to have the value $+\infty$. Here are a few natural notions attached to random and stopping times:

Definition 1.3.8 *(i) If T is a stopping time, then the collection*

$$\mathcal{F}_T \stackrel{\text{def}}{=} \Big\{ A \in \mathcal{F}_\infty :\ A \cap [T \le t] \in \mathcal{F}_t \quad \forall\, t \in [0,\infty] \Big\}$$

*is easily seen to be a σ-algebra on Ω. It is called the **past** at time T or the past of T.* To paraphrase: an event A occurs in the past of T if at any instant t at which T has arrived no look into the future is necessary to determine whether the event A has occurred.
(ii) The value of a process Z at a random time T is the random variable

$$Z_T : \omega \mapsto Z_{T(\omega)}(\omega) \ .$$

(iii) Let S, T be two random times. The random interval $[\![S, T]\!]$ is the set

$$[\![S, T]\!] \stackrel{\text{def}}{=} \Big\{(s, \omega) \in \boldsymbol{B} \,:\, S(\omega) < s \le T(\omega),\ s < \infty\Big\} \ ;$$

*and $[\![S, T)\!]$, $(\!]S, T]\!]$, and $(\!]S, T)\!]$ are defined similarly. Note that the point (∞, ω) does not belong to any of these intervals, even if $T(\omega) = \infty$. If both S and T are stopping times, then the random intervals $[\![S, T]\!]$, $[\![S, T)\!]$, $(\!]S, T]\!]$, and $(\!]S, T)\!]$ are called **stochastic intervals**. A stochastic interval is **finite** if its endpoints are finite stopping times, and **bounded** if the endpoints are bounded stopping times.*
*(iv) The **graph** of a random time T is the random interval*

$$[\![T]\!] = [\![T, T]\!] \stackrel{\text{def}}{=} \Big\{(s, \omega) \in \boldsymbol{B} \,:\, T(\omega) = s < \infty\Big\} \ .$$

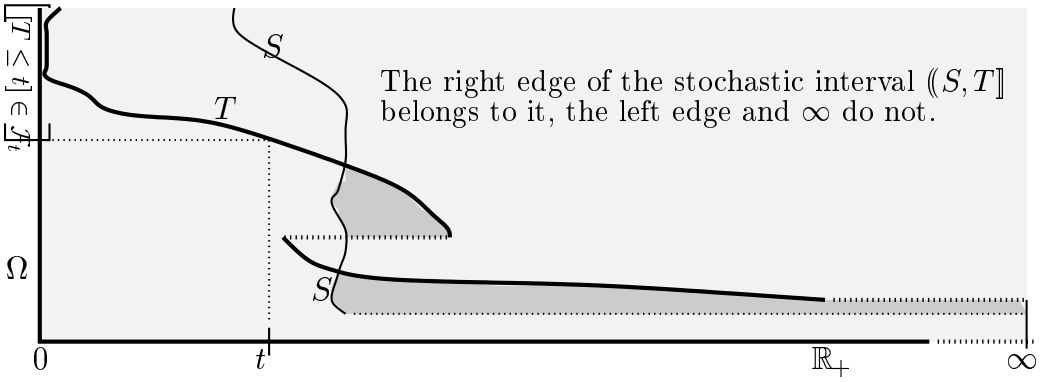

Figure 1.5 A stochastic interval

Proposition 1.3.9 *Suppose that Z is a progressively measurable process and T is a stopping time. The **stopped process** Z^T, defined by $Z^T_s(\omega) = Z_{T\wedge s}(\omega)$, is progressively measurable, and Z_T is measurable on $\mathcal{F}_T$.*

We can paraphrase the last statement by saying "a progressively measurable process is adapted to the 'expanded filtration' $\{\mathcal{F}_T : T \text{ a stopping time}\}$."

Proof. For fixed t let $\mathcal{F}^t = \mathcal{B}^\bullet[0, \infty) \otimes \mathcal{F}_t$. The map[12] that sends (s, ω) to $\big(T(\omega) \wedge t \wedge s, \omega\big)$ from $\boldsymbol{B}$ to itself is easily seen to be $\mathcal{F}^t/\mathcal{F}^t$-measurable.

$(Z^T)^t = Z^{T\wedge t}$ is the composition of Z^t with this map and is therefore $\mathcal{F}^t$-measurable. This holds for all t, so Z^T is progressively measurable. For $a \in \mathbb{R}$, $[Z_T > a] \cap [T \le t] = [Z^T_t > a] \cap [T \le t]$ belongs to $\mathcal{F}_t$ because Z^T is adapted. This holds for all t, so $[Z_T > a] \in \mathcal{F}_T$ for all a: $Z_T \in \mathcal{F}_T$, as claimed. ∎

Exercise 1.3.10 If the process Z is progressively measurable, then so is the process $t \mapsto Z_{T\vee t} - Z_T$.

Let $T_1 \le T_2 \le \ldots \le T_\infty = \infty$ be an increasing sequence of stopping times and X a progressively measurable process. For $r \in \mathbb{R}$ define $K = \inf\{k \in \mathbb{N} : X_{T_k} > r\}$. Then $T_K : \omega \mapsto T_{K(\omega)}(\omega)$ is a stopping time.

Some Examples of Stopping Times

Stopping times occur most frequently as first hitting times – of the moon in our example of section 1.1, or of sets of bad behavior in much of the analysis below. First hitting times are stopping times, provided that the filtration $\mathcal{F}_\bullet$ satisfies some natural conditions – see figure 1.6 on page 40. This is shown with the help of a little capacity theory in appendix A, section A.5. A few elementary results, established with rather simple arguments, will go a long way:

Proposition 1.3.11 *Let I be an adapted process with increasing right-continuous paths and let $\lambda \in \mathbb{R}$. Then*

$$T^\lambda \stackrel{\text{def}}{=} \inf\{t : I_t \ge \lambda\}$$

is a stopping time, and $I_{T^\lambda} \ge \lambda$ on the set $[T^\lambda < \infty]$. Moreover, the functions $\lambda \mapsto T^\lambda(\omega)$ are increasing and left-continuous.

Proof. $T^\lambda(\omega) \le t$ if and only if $I_t(\omega) \ge \lambda$. In other words, $[T^\lambda \le t] = [I_t \ge \lambda] \in \mathcal{F}_t$, so T^λ is a stopping time. If $T^\lambda(\omega) < \infty$, then there is a sequence (t_n) of instants that decreases to $T^\lambda(\omega)$ and has $I_{t_n}(\omega) \ge \lambda$. The right-continuity of I produces $I_{T^\lambda(\omega)}(\omega) \ge \lambda$.

That $T^\lambda \le T^\mu$ when $\lambda \le \mu$ is obvious: $T^\bullet$ is indeed increasing. If $T^\lambda \le t$ for all $\lambda < \mu$, then $I_t \ge \lambda$ for all $\lambda < \mu$, and thus $I_t \ge \mu$ and $T^\mu \le t$. That is to say, $\sup_{\lambda<\mu} T^\lambda = T^\mu$: $\lambda \mapsto T^\lambda$ is left-continuous. ∎

The main application of this near-trivial result is to the maximal process of some process I. Proposition 1.3.11 applied to $(Z - Z^S)^\star$ yields the

Corollary 1.3.12 *Let S be a finite stopping time and $\lambda > 0$. Suppose that Z is adapted and has right-continuous paths. Then the first time the maximal gain of Z after S exceeds λ,*

$$T = \inf\{t > S : \sup_{S<s\le t} |Z_s - Z_S| \ge \lambda\} = \inf\{t : |Z - Z^S|^\star_t \ge \lambda\}\ ,$$

is a stopping time strictly greater than S, and $|Z - Z_S|^\star_T \ge \lambda$ on $[T < \infty]$.

Proposition 1.3.13 *Let Z be an adapted process, T a stopping time, X a random variable measurable on $\mathcal{F}_T$, and $S = \{s_0 < s_1 < \ldots < s_N\} \subset [0,u)$ a finite set of instants. Define*

$$T' = \inf\{s \in S : s > T\ ,\ Z_s \diamond X\} \wedge u\ ,$$

where $\diamond$ stands for any of the relations $>, \geq, =, \leq, <$. Then T' is a stopping time, and $Z_{T'} \in \mathcal{F}_{T'}$ satisfies $Z_{T'} \diamond X$ on $[T' < u]$.

Proof. If $t \geq u$, then $[T' \leq t] = \Omega \in \mathcal{F}_t$. Let then $t < u$. Then

$$[T' \leq t] = \bigcup \big\{ [T < s] \cap [Z_s \diamond X] : S \ni s \leq t \big\}\ .$$

Now $[T < s] \in \mathcal{F}_s$ and so $[T < s]Z_s \in \mathcal{F}_s$. Also, $[X > x] \cap [T < s] \in \mathcal{F}_s$ for all x, so that $[T < s]X \in \mathcal{F}_s$ as well. Hence $[T < s] \cap [Z_s \diamond X] \in \mathcal{F}_s$ for $s \leq t$ and so $[T' \leq t] \in \mathcal{F}_t$. Clearly $Z_{T'}[T' \leq t] = \bigcup_{S \ni s \leq t} Z_s[T'{=}s] \in \mathcal{F}_t$ for all $t \in S$, and so $Z_{T'} \in \mathcal{F}_{T'}$. ∎

Proposition 1.3.14 *Let S be a stopping time, let $c > 0$, and let $X \in \mathfrak{D}$. Then*

$$T = \inf\{t > S :\ |\Delta X_t| \geq c\}$$

is a stopping time that is strictly later than S on the set $[S < \infty]$, and $|\Delta X_T| \geq c$ on $[T < \infty]$.

Proof. Let us prove the last point first. Let $t_n \geq T$ decrease to T and $|\Delta X_{t_n}| \geq c$. Then (t_n) must be ultimately constant. For if it is not, then it can be replaced by a strictly decreasing subsequence, in which case both X_{t_n} and X_{t_n-} converge to the same value, to wit, X_T. This forces $\Delta X_{t_n} \xrightarrow[n\to\infty]{} 0$, which is impossible since $|\Delta X_{t_n}| \geq c > 0$. Thus $T > S$ and $\Delta X_T \geq c$.

Next observe that $T \leq t$ precisely if for every $n \in \mathbb{N}$ there are numbers q, q' in the countable set

$$\mathbb{Q}^t = (\mathbb{Q} \cap [0,t]) \cup \{t\}$$

with $S < q < q'$ and $q' - q < 1/n$, and such that $|X_{q'} - X_q| \geq c - 1/n$. This condition is clearly necessary. To see that it is sufficient note that in its presence there are rationals $S < q_n < q_n' \leq t$ with $q_n' - q_n \to 0$ and $|X_{q_n'} - X_{q_n}| \geq c - 1/n$. Extracting a subsequence we may assume that both (q_n) and (q_n') converge to some point $s \in [S,t]$. (q_n) can clearly not contain a constant subsequence; if (q_n') does, then $|\Delta X_s| \geq c$ and $T \leq t$. If (q_n') has no constant subsequence, it can be replaced by a strictly monotone subsequence. We may thus assume that both (q_n) and (q_n') are strictly monotone. Recalling the first part of the proof we see that this is possible only if (q_n) is increasing and (q_n') decreasing, in which case $T \leq t$ again. The upshot of all this is that

$$[T \leq t] = \bigcap_{n\in\mathbb{N}} \bigcup_{\substack{q,q'\in\mathbb{Q}^t \\ q<q'<q+1/n}} [S < q] \cap \big[|X_{q'} - X_q| \geq c - 1/n\big]\ ,$$

a set that is easily recognizable as belonging to $\mathcal{F}_t$. ∎

Further elementary but ubiquitous facts about stopping times are developed in the next exercises. Most are clear upon inspection, and they are used freely in the sequel.

Exercise 1.3.15 (i) An instant $t \in \mathbb{R}_+$ is a stopping time, and its past equals $\mathcal{F}_t$. (ii) The infimum of a finite number and the supremum of a countable number of stopping times are stopping times.

Exercise 1.3.16 Let S, T be any two stopping times. (i) If $S \le T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$. (ii) In general, the sets $[S < T]$, $[S \le T]$, $[S = T]$ belong to $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$.

Exercise 1.3.17 A random time T is a stopping time precisely if the (indicator function of the) random interval $[\![0, T)\!)$ is an adapted process.[15] If S, T are stopping times, then any stochastic interval with left endpoint S and right endpoint T is an adapted process.

Exercise 1.3.18 Let T be a stopping time and $A \in \mathcal{F}_T$. Setting

$$T_A(\omega) \stackrel{\text{def}}{=} \begin{cases} T(\omega) & \text{if } \omega \in A \\ \infty & \text{if } \omega \notin A \end{cases} \;=\; T + \infty \cdot A^c$$

defines a new stopping time T_A, the ***reduction*** of the stopping time T to A.

Exercise 1.3.19 Let T be a stopping time and $A \in \mathcal{F}_\infty$. Then $A \in \mathcal{F}_T$ if and only if the reduction T_A is a stopping time. A random variable f is measurable on $\mathcal{F}_T$ if and only if $f \cdot [T \le t] \in \mathcal{F}_t$ at all instants t.

Exercise 1.3.20 The following "discrete approximation from above" of a stopping time T will come in handy on several occasions. For $n = 1, 2, \dots$ set

$$T^{(n)} \stackrel{\text{def}}{=} \begin{cases} 0 & \text{on } [T = 0]; \\ \dfrac{k+1}{n} & \text{on } \Big[\dfrac{k}{n} < T \le \dfrac{k+1}{n}\Big], \\ \infty & \text{on } [T = \infty]. \end{cases} \qquad k = 0, 1, \dots;$$

Using convention A.1.5 on page 364, we can rewrite this as

$$T^{(n)} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{k+1}{n} \cdot \Big[\frac{k}{n} < T \le \frac{k+1}{n}\Big] \;+\; \infty \cdot [T = \infty]\,.$$

Then $T^{(n)}$ is a stopping time that takes only countably many values, and T is the pointwise infimum of the decreasing sequence $T^{(n)}$.

Exercise 1.3.21 Let $X \in \mathfrak{D}$. (i) The set $\{s \in \mathbb{R}_+ : |\Delta X_s(\omega)| \ge \epsilon\}$ is discrete (has no accumulation point in $\mathbb{R}_+$) for every $\omega \in \Omega$ and $\epsilon > 0$. (ii) There exists a countable family $\{T_n\}$ of stopping times with bounded disjoint graphs $[\![T_n]\!]$ at which the jumps of X occur:

$$[\Delta X \neq 0] \subseteq \bigcup_n [\![T_n]\!]\,.$$

(iii) Let h be a Borel function on $\mathbb{R}$ and assume that for all $t < \infty$ the sum $J_t \stackrel{\text{def}}{=} \sum_{0\le s\le t} h(\Delta X_s)$ converges absolutely. Then $J_\bullet$ is adapted.

[15] See convention A.1.5 and figure A.14 on page 365.

Probabilities

A probabilistic model of a system requires, of course, a probability measure $\mathbb{P}$ on the pertinent σ-algebra $\mathcal{F}_\infty$, the idea being that a priori assumptions on, or measurements of, $\mathbb{P}$ plus mathematical analysis will lead to estimates of the random variables of interest.

The need to consider a family $\mathfrak{P}$ of pertinent probabilities does arise: first, there is often not enough information to specify one particular probability as the right one, merely enough to narrow the class. Second, in the context of stochastic differential equations and Markov processes, whole slews of probabilities appear that may depend on a starting point or other parameter (see theorem 5.7.3). Third, it is possible and often desirable to replace a given probability by an equivalent one with respect to which the stochastic integral has superior properties (this is done in section 4.1 and is put to frequent use thereafter). Nevertheless, we shall mostly develop the theory for a fixed probability $\mathbb{P}$ and simply apply the results to each $\mathbb{P} \in \mathfrak{P}$ separately. The pair $\big(\mathcal{F}_\bullet, \mathbb{P}\big)$ or $\big(\mathcal{F}_\bullet, \mathfrak{P}\big)$, as the case may be, is termed a ***measured filtration***.

Let $\mathbb{P} \in \mathfrak{P}$. It is customary to denote the integral with respect to $\mathbb{P}$ by $\mathbb{E}^{\mathbb{P}}$ and to call it the ***expectation***; that is to say, for $f : \Omega \to \mathbb{R}$ measurable on $\mathcal{F}_\infty$,

$$\mathbb{E}^{\mathbb{P}}[f] = \int f \, d\mathbb{P} = \int f(\omega) \, \mathbb{P}(d\omega) \,, \qquad \mathbb{P} \in \mathfrak{P} \,.$$

If there is no doubt which probability $\mathbb{P} \in \mathfrak{P}$ is meant, we write simply $\mathbb{E}$.

A subset $N \subset \Omega$ is commonly called $\mathbb{P}$-negligible, or simply ***negligible*** when there is no doubt about the probability, if its outer measure $\mathbb{P}^*[N]$ equals zero. This is the same as saying that it is contained in a set of $\mathcal{F}_\infty$ that has measure zero. A function on Ω is negligible if it vanishes off a negligible set; this is the same as saying that the upper integral[16] of its absolute value vanishes. The functions that differ negligibly, i.e., only in a negligible set, from f constitute the ***equivalence class*** $\dot{f}$. We have seen in the proof of theorem 1.2.2 (ii) that in the present business we sometimes have to make the distinction between a random variable and its class, boring as this is. We write $f \doteq g$ if f and g differ negligibly and also $\dot{f} \doteq \dot{g}$ if f and g belong to the same equivalence class, etc.

A property of the points of Ω is said to hold $\mathbb{P}$-almost surely or simply ***almost surely***, if the set N of points of Ω where it does not hold is negligible. The abbreviation $\mathbb{P}$-a.s. or simply ***a.s.*** is common. The terminology "almost everywhere" and its short form "a.e." will be avoided in context with $\mathbb{P}$ since it is employed with a different meaning in chapter 3.

[16] See page 396.

The Sizes of Random Variables

With every probability $\mathbb{P}$ on $\mathcal{F}_\infty$ there come many different ways of measuring the size of a random variable. We shall review a few that have proved particularly useful in many branches of mathematics and that continue to be so in the context of stochastic integration and of stochastic differential equations.

For a function f measurable on the universal completion $\mathcal{F}_\infty^*$ and $0 < p < \infty$, set

$$\| f \|_p \stackrel{\text{def}}{=} \| f \|_{L^p} \stackrel{\text{def}}{=} \Big(\int |f|^p \, d\mathbb{P} \Big)^{1/p} .$$

If there is need to stress which probability $\mathbb{P} \in \mathfrak{P}$ is meant, we write $\| f \|_{L^p(\mathbb{P})}$. The ***p*-mean** $\| \ \|_p$ is

absolute-homogeneous: $\quad \|r \cdot f\|_p = |r| \cdot \|f\|_p$

and ***subadditive***: $\quad \|f + g\|_p \le \|f\|_p + \|g\|_p$

in the range $1 \le p < \infty$, but not for $0 < p < 1$. Since it is often more convenient to have subadditivity at one's disposal rather than homogeneity, we shall mostly employ the subadditive versions

$$\llbracket f \rrbracket_p \stackrel{\text{def}}{=} \begin{cases} \|f\|_{L^p(\mathbb{P})} = \Big(\int |f|^p \, d\mathbb{P} \Big)^{1/p} & \text{for } 1 \le p < \infty, \\ \|f\|^p_{L^p(\mathbb{P})} = \int |f|^p \, d\mathbb{P} & \text{for } 0 < p \le 1 \ . \end{cases} \tag{1.3.1}$$

L^p or $L^p(\mathbb{P})$ denotes the collection of measurable functions f with $\llbracket f \rrbracket_p < \infty$, the p-integrable functions. The collection of $\mathcal{F}_t$-measurable functions in L^p is $L^p(\mathcal{F}_t)$ or $L^p(\mathcal{F}_t, \mathbb{P})$. It is well known that L^p is a complete pseudometric space under the distance $dist_p(f, g) = \llbracket f - g \rrbracket_p$ – it is to make $dist_p$ a metric that we generally prefer the subadditive size measurement $\llbracket \ \rrbracket_p$ over its homogeneous cousin $\| \ \|_p$.

Two random variables in the same class have the same p-means, so we shall also talk about $\llbracket \dot{f} \rrbracket_p$, etc.

The prominence of the p-means $\| \ \|_p$ and $\llbracket \ \rrbracket_p$ among other size measurements that one might think up is due to Hölder's inequality A.8.4, which provides a partial alleviation of the fact that L^1 is not an algebra, and to the method of interpolation (see proposition A.8.24). Section A.8 contains further information about the p-means and the L^p-spaces. A process Z is called ***p*-integrable** if the random variables Z_t are all p-integrable, and ***L^p*-bounded** if

$$\sup_t \| Z_t \|_p < \infty \ , \qquad\qquad 0 < p \le \infty.$$

The largest class of useful random variables is that of measurable a.s. finite ones. It is denoted by L^0, $L^0(\mathbb{P})$, or $L^0(\mathcal{F}_t, \mathbb{P})$, as the context requires. It

extends the slew of the L^p-spaces at $p = 0$. It plays a major role in stochastic analysis due to the fact that it forms an algebra and does not change when $\mathbb{P}$ is replaced by an equivalent probability (exercise A.8.11). There are several ways to attach a numerical size to a function $f \in L^0$, the most common[17] being

$$\llbracket f \rrbracket_0 = \llbracket f \rrbracket_{0;\mathbb{P}} \stackrel{\text{def}}{=} \inf\Big\{\lambda : \mathbb{P}\big[|f| > \lambda\big] \leq \lambda\Big\} .$$

It measures ***convergence in probability***, also called ***convergence in measure***; namely, $f_n \to f$ in probability if

$$dist_0(f_n, f) \stackrel{\text{def}}{=} \llbracket f_n - f \rrbracket_0 \xrightarrow[n\to\infty]{} 0 .$$

$\llbracket\ \rrbracket_0$ is subadditive but not homogeneous (exercise A.8.1). There is also a whole slew of absolute-homogeneous but non-subadditive functionals, one for every $\alpha \in \mathbb{R}$, that can be used to describe the topology of $L^0(\mathbb{P})$:

$$\|f\|_{[\alpha]} = \|f\|_{[\alpha;\mathbb{P}]} \stackrel{\text{def}}{=} \inf\big\{\lambda > 0 : \mathbb{P}[|f| > \lambda] \leq \alpha\big\} .$$

Further information about these various size measurements and their relation to each other can be found in section A.8. The reader not familiar with L^0 or the basic notions concerning topological vector spaces is strongly advised to peruse that section. In the meantime, here is a little mnemonic device: functionals with "straight sides" $\|\ \|$ are homogeneous, and those with a little "crossbar," like $\llbracket\ \rrbracket$, are subadditive. Of course, if $1 \leq p < \infty$, then $\|\ \|_p = \llbracket\ \rrbracket_p$ has both properties.

Exercise 1.3.22 While some of the functionals $\llbracket\ \rrbracket_p$ are not homogeneous – the $\llbracket\ \rrbracket_p$ for $0 \leq p < 1$ – and some are not subadditive – the $\|\ \|_p$ for $0 < p < 1$ and the $\|\ \|_{[\alpha]}$ – all of them respect the order: $|f| \leq |g| \Longrightarrow \|f\|_\bullet \leq \|g\|_\bullet$. Functionals with this property are termed ***solid***.

Two Notions of Equality for Processes

Modifications Let $\mathbb{P}$ be a probability in the pertinent class $\mathfrak{P}$. Two processes X, Y are $\mathbb{P}$-***modifications*** of each other if, at every instant t, $\mathbb{P}$-almost surely $X_t = Y_t$. We also say "X is a modification of Y," suppressing as usual any mention of $\mathbb{P}$ if it is clear from the context. In fact, it may happen that $[X_t \neq Y_t]$ is so wild that the only set of $\mathcal{F}_t$ containing it is the whole space Ω. It may, in particular, occur that X is adapted but a modification Y of it is not.

Indistinguishability Even if X and a modification Y are adapted, the sets $[X_t \neq Y_t]$ may vary wildly with t. They may even cover all of Ω. In other words, while the values of X and Y might at no finite instant be distinguishable with $\mathbb{P}$, an apparatus rigged to ring the first time X and

[17] It is commonly attributed to Ky–Fan.

Y differ may ring for sure, even immediately. There is evidently a need for a more restrictive notion of equality for processes than merely being almost surely equal at every instant.

To approach this notion let us assume that X, Y are progressively measurable, as respectable processes without ESP are supposed to be. It seems reasonable to say that X, Y are indistinguishable if their entire paths $X_\bullet, Y_\bullet$ agree almost surely, that is to say, if the event

$$N \stackrel{\text{def}}{=} [X_\bullet \neq Y_\bullet] = \bigcup\nolimits_{k\in\mathbb{N}} [|X - Y|^\star_k > 0]$$

has no chance of occurring. Now $N^k = [X^k_\bullet \neq Y^k_\bullet] = [|X - Y|^\star_k > 0]$ is the uncountable union of the sets $[X_s \neq Y_s]$, $s \leq k$, and looks at first sight nonmeasurable. By corollary A.5.13, though, N^k belongs to the universal completion of $\mathcal{F}_k$; there is no problem attaching a measure to it. There is still a little conceptual difficulty in that N^k may not belong to $\mathcal{F}_k$ itself, meaning that it is not observable at time k, but this seems like splitting hairs. Anyway, the filtration will soon be enlarged so as to become regular; this implies its universal completeness, and our little trouble goes away. We see that no apparatus will ever be able to detect any difference between X and Y if they differ only on a set like $\bigcup_{k\in\mathbb{N}} N^k$, which is the countable union of negligible sets in $\mathcal{A}_\infty$. Such sets should be declared inconsequential. Letting $\mathcal{A}_{\infty\sigma}$ denote the collection of sets that are countable unions of sets in $\mathcal{A}_\infty$, we are led to the following definition of indistinguishability. Note that it makes sense without any measurability assumption on X, Y.

Definition 1.3.23 *(i) A subset of Ω is **nearly empty** if it is contained in a negligible set of $\mathcal{A}_{\infty\sigma}$. A random variable is nearly zero if it vanishes outside a nearly empty set.*

*(ii) A property P of the points $\omega \in \Omega$ is said to hold **nearly** if the set N of points of Ω where it does not hold is nearly empty. Writing $f = g$ for two random variables f, g generally means that f and g nearly agree.*

*(iii) Two processes X, Y are **indistinguishable** if $[X_\bullet \neq Y_\bullet]$ is nearly empty. A process or subset of the base space $\boldsymbol{B}$ that cannot be distinguished from zero is called **evanescent**. When we write $X = Y$ for two processes X, Y we mean generally that X and Y are indistinguishable.*

When the probability $\mathbb{P} \in \mathfrak{P}$ must be specified, then we talk about $\mathbb{P}$-nearly empty sets or $\mathbb{P}$-nearly vanishing random variables, properties holding $\mathbb{P}$-nearly, processes indistinguishable with $\mathbb{P}$ or $\mathbb{P}$-indistinguishable, and $\mathbb{P}$-evanescent processes.

A set N is nearly empty if someone with a finite if possibly very long life span t can measure it ($N \in \mathcal{F}_t$) and find it to be negligible, or if it is the countable union of such sets. If he and his offspring must wait past the expiration of time (check whether $N \in \mathcal{F}_\infty$) to ascertain that N is negligible – in other words, if this must be left to God – then N is not nearly empty even

though it be negligible. Think of nearly empty sets as sets whose negligibility can be detected before the expiration of time.

There is an apologia for the introduction of this class of sets in warnings 1.3.39 on page 39 and 3.9.20 on page 167.

Example 1.3.24 Take for Ω the unit interval $[0,1]$. For $n \in \mathbb{N}$ let $\mathcal{F}_n$ be the σ-algebra generated by the closed intervals $[k2^{-n}, (k+1)2^{-n}]$, $0 \le k \le 2^n$. To obtain a filtration indexed by $[0,\infty)$ set $\mathcal{F}_t = \mathcal{F}_n$ for $n \le t < n+1$. For $\mathbb{P}$ take Lebesgue measure λ. The negligible sets in $\mathcal{F}_n$ are the sets of ***dyadic rationals*** of the form $k2^{-n}$, $0 < k < 2^n$. In this case $\mathcal{A}_\infty$ is the algebra of finite unions of intervals with dyadic-rational endpoints, and its span $\mathcal{F}_\infty$ is the σ-algebra of all Borel sets on $[0,1]$. A set is nearly empty if and only if it is a subset of the dyadic rationals in $(0,1)$. There are many more negligible sets than these. Here is a striking phenomenon: consider a countable set $\mathcal{I}$ of irrational numbers dense in $[0,1]$. It is Lebesgue negligible but has outer measure 1 for any of the measured triples $\big([0,1], \mathcal{F}_t, \mathbb{P}_{|\mathcal{F}_t}\big)$. The upshot: the notion of a nearly empty set is rather more restrictive than that of a negligible set. In the present example there are $2^{\aleph_0}$ of the former and $\ge 2^{\aleph_1}$ of the latter. For more on this see example 1.3.32.

Exercise 1.3.25 $N \subset \Omega$ is negligible if and only if there is, for every $\epsilon > 0$, a set of $\mathcal{A}_{\infty\sigma}$ that has measure less than ϵ and contains N. It is nearly empty if and only if there exists a set of $\mathcal{A}_{\infty\sigma}$ that has measure equal to zero and contains N. $N \subset \Omega$ is nearly empty if and only if there exist instants $t_n < \infty$ and negligible sets $N_n \in \mathcal{F}_{t_n}$ whose union contains N.

Exercise 1.3.26 A subset of a nearly empty set is nearly empty; so is the countable union of nearly empty sets. A subset of an evanescent set is evanescent; so is the countable union of evanescent sets.

Near-emptyness and evanescence are "solid" notions: if f, g are random variables, g nearly zero and $|f| \le |g|$, then f is nearly zero; if X, Y are processes, Y evanescent and $|X| \le |Y|$, then X is evanescent. The pointwise limit of a sequence of nearly zero random variables is nearly zero. The pointwise limit of a sequence of evanescent processes is evanescent.

A process X is evanescent if and only if the projection $\pi_\Omega[X \neq 0]$ is nearly empty.

Exercise 1.3.27 (i) Two stopping times that agree almost surely agree nearly.

(ii) If T is a nearly finite stopping time and $N \in \mathcal{F}_T$ is negligible, then it is nearly empty.

Exercise 1.3.28 Indistinguishable processes are modifications of each other. Two adapted left- or right-continuous processes that are modifications of each other are indistinguishable.

The Natural Conditions

We shall now enlarge the given filtration slightly, and carefully. The purpose of this is to gain regularity results for paths of integrators (theorem 2.3.4) and to increase the supply of stopping times (exercise 1.3.30 and appendix A, pages 436–438).

Right-Continuity of a Filtration Many arguments are simplified or possible only when the filtration $\mathcal{F}_\bullet$ is right-continuous:

Definition 1.3.29 *The* ***right-continuous version*** $\mathcal{F}_{\bullet+}$ *of a filtration* $\mathcal{F}_\bullet$ *is defined by*

$$\mathcal{F}_{t+} \stackrel{\text{def}}{=} \bigcap_{u>t} \mathcal{F}_u \qquad \forall\, t \geq 0\,.$$

The given filtration $\mathcal{F}_\bullet$ *is termed* ***right-continuous*** *if* $\mathcal{F}_\bullet = \mathcal{F}_{\bullet+}$.

The following exercise develops some of the benefits of having the filtration right-continuous. We shall see soon (proposition 2.2.11) that it costs nothing to replace any given filtration by its right-continuous version, so that we can easily avail ourselves of these benefits.

Exercise 1.3.30 The right-continuity of the filtration implies all of this: (i) A random time T is a stopping time if and only if $[T < t] \in \mathcal{F}_t$ for all $t > 0$. This is often easier to check than that $[T \leq t] \in \mathcal{F}_t$ for all t. For instance (compare with proposition 1.3.11): (ii) If Z is an adapted process with right- or left-continuous paths, then for any $\lambda \in \mathbb{R}$

$$T^{\lambda+} \stackrel{\text{def}}{=} \inf\{t : Z_t > \lambda\}$$

is a stopping time. Moreover, the functions $\lambda \mapsto T^{\lambda+}(\omega)$ are increasing and right-continuous. (iii) If T is a stopping time, then $A \in \mathcal{F}_T$ iff $A \cap [T < t] \in \mathcal{F}_t \quad \forall\, t$.

(iv) The infimum T of a countable collection $\{T_n\}$ of stopping times is a stopping time, and its past is $\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n}$ (cf. exercise 1.3.15).

(v) $\mathcal{F}_\bullet$ and $\mathcal{F}_{\bullet+}$ have the same adapted left-continuous processes. A process adapted to the filtration $\mathcal{F}_\bullet$ and progressively measurable on $\mathcal{F}_{\bullet+}$ is progressively measurable on $\mathcal{F}_\bullet$.

Regularity of a Measured Filtration It is still possible that there exist measurable indistinguishable processes X, Y of which one is adapted, the other not. This unsatisfactory state of affairs is ruled out if the filtration is regular. For motivation consider a subset $N \subset \Omega$ that is not measurable on $\mathcal{F}_t$ (too wild to be observable now, at time t) but that is measurable on $\mathcal{F}_u$ for some $u > t$ (observable then) and turns out to have probability $\mathbb{P}[N] = 0$ of occurring. Or N might merely be a subset of such a set. The class of such N and their countable unions is precisely the class of nearly empty sets. It does no harm but confers great technical advantage to declare such an event N to be both observable and impossible now. Precisely:

Definition 1.3.31 *(i) Given a measured filtration* $\big(\mathcal{F}_\bullet, \mathfrak{P}\big)$ *on* Ω *and a probability* $\mathbb{P}$ *in the pertinent class* $\mathfrak{P}$*, set*

$$\mathcal{F}_t^{\mathbb{P}} \stackrel{\text{def}}{=} \big\{A \subset \Omega\,:\, \exists A_{\mathbb{P}} \in \mathcal{F}_t \;\; \textit{so that} \;\; |A - A_{\mathbb{P}}| \;\; \textit{is } \mathbb{P}\textit{-nearly empty}\big\}\,.$$

Here $|A - A_{\mathbb{P}}|$ *is the symmetric difference* $\big(A \setminus A_{\mathbb{P}}\big) \cup \big(A_{\mathbb{P}} \setminus A\big)$ *(see convention A.1.5).* $\mathcal{F}_t^{\mathbb{P}}$ *is easily seen to be a* σ*-algebra; in fact, it is the* σ*-algebra generated by* $\mathcal{F}_t$ *and the* $\mathbb{P}$*-nearly empty sets. The collection*

$$\mathcal{F}_\bullet^{\mathbb{P}} \stackrel{\text{def}}{=} \big\{\mathcal{F}_t^{\mathbb{P}}\big\}_{0 \leq t \leq \infty}$$

is the $\mathbb{P}$***-regularization*** *of* $\mathcal{F}_\bullet$*. The filtration* $\mathcal{F}^{\mathfrak{P}}$ *composed of the* σ*-algebras*

$$\mathcal{F}_t^{\mathfrak{P}} \stackrel{\text{def}}{=} \bigcap_{\mathbb{P}\in\mathfrak{P}} \mathcal{F}_t^{\mathbb{P}}\,, \qquad t \geq 0\,,$$

is the $\mathfrak{P}$***-regularization****, or simply the* ***regularization****, when* $\mathfrak{P}$ *is clear from the context.*

(ii) The measured filtration $\big(\mathcal{F}_\bullet, \mathfrak{P}\big)$ *is* ***regular*** *if* $\mathcal{F}_\bullet = \mathcal{F}_\bullet^{\mathfrak{P}}$*. We then also write "*$\mathcal{F}_\bullet$ *is* $\mathfrak{P}$***-regular****," or simply "*$\mathcal{F}_\bullet$ *is* ***regular****" when* $\mathfrak{P}$ *is understood.*

Let us paraphrase the regularity of a filtration in intuitive terms: "an event that proves in the long run to be indistinguishable, whatever the probability in the admissible class $\mathfrak{P}$, from some event observable now is considered to be observable now."

$\mathcal{F}_t^{\mathbb{P}}$ contains the completion of $\mathcal{F}_t$ under the restriction $\mathbb{P}_{|\mathcal{F}_t}$, which in turn contains the universal completion. The regularization of $\mathcal{F}$ is thus universally complete. If $\mathcal{F}$ is regular, then the maximal process of a progressively measurable process is again progressively measurable (corollary A.5.13). This is nice. The main point of regularity is, though, that it allows us to prove the path regularity of integrators (section 2.3 and definition 3.7.6). The following exercises show how much – or rather how little – is changed by such a replacement and develop some of the benefits of having the filtration regular. We shall see soon (proposition 2.2.11) that it costs nothing to replace a given filtration by its regularization, so that we can easily avail ourselves of these benefits.

Example 1.3.32 In the right-continuous measured filtration $(\Omega = [0,1], \mathcal{F}, \mathbb{P} = \lambda)$ of example 1.3.24 the $\mathcal{F}_t$ are all universally complete, and the couples $(\mathcal{F}_t, \mathbb{P})$ are complete. Nevertheless, the regularization differs from $\mathcal{F}$: $\mathcal{F}_t^{\mathbb{P}}$ is the σ-algebra generated by $\mathcal{F}_t$ and the dyadic-rational points in $(0,1)$. For more on this see example 1.3.45.

Exercise 1.3.33 A random variable f is measurable on $\mathcal{F}_t^{\mathbb{P}}$ if and only if there exists an $\mathcal{F}_t$-measurable random variable $f_{\mathbb{P}}$ that $\mathbb{P}$-nearly equals f.

Exercise 1.3.34 (i) $\mathcal{F}_\bullet^{\mathfrak{P}}$ is regular. (ii) A random variable f is measurable on $\mathcal{F}_t^{\mathfrak{P}}$ if and only if for every $\mathbb{P} \in \mathfrak{P}$ there is an $\mathcal{F}_t$-measurable random variable $\mathbb{P}$-nearly equal to f.

Exercise 1.3.35 Assume that $\mathcal{F}_\bullet$ is right-continuous and let $\mathbb{P}$ be a probability on $\mathcal{F}_\infty$.

(i) Let X be a right-continuous process adapted to $\mathcal{F}_\bullet^{\mathbb{P}}$. There exists a process X' that is $\mathbb{P}$-nearly right-continuous and adapted to $\mathcal{F}_\bullet$ and cannot be distinguished from X with $\mathbb{P}$. If X is a set, then X' can be chosen to be a set; and if X is increasing, then X' can be chosen increasing and right-continuous everywhere.

(ii) A random time T is a stopping time on $\mathcal{F}_\bullet^{\mathbb{P}}$ if and only if there exists an $\mathcal{F}_\bullet$-stopping time $T_{\mathbb{P}}$ that nearly equals T. A set A belongs to $\mathcal{F}_T$ if and only if there exists a set $A_{\mathbb{P}} \in \mathcal{F}_{T_{\mathbb{P}}}$ that is nearly equal to A.

Exercise 1.3.36 (i) The right-continuous version of the regularization equals the regularization of the right-continuous version; if $\mathcal{F}_\bullet$ is regular, then so is $\mathcal{F}_{\bullet+}$.

(ii) Substituting $\mathcal{F}_{\bullet+}$ for $\mathcal{F}_\bullet$ will increase the supply of adapted and of progressively measurable processes, and of stopping times, and will sometimes enlarge the spaces $\mathrm{L}^p[\mathcal{F}_t, \mathbb{P}]$ of equivalence classes (sometimes it will not – see exercise 1.3.47.).

Exercise 1.3.37 $\mathcal{F}_t^{\mathfrak{P}}$ contains the σ-algebra generated by $\mathcal{F}_t$ and the nearly empty sets, and coincides with that σ-algebra if there happens to exist a probability with respect to which every probability in $\mathfrak{P}$ is absolutely continuous.

Definition 1.3.38 (The Natural Conditions) *Let $(\mathcal{F}_\bullet, \mathfrak{P})$ be a measured filtration. The **natural enlargement** of $\mathcal{F}_\bullet$ is the filtration $\mathcal{F}_{\bullet+}^{\mathfrak{P}}$ obtained by regularizing the right-continuous version of $\mathcal{F}_\bullet$ (or, equivalently, by taking the right-continuous version of the regularization — see exercise 1.3.36).*

*Suppose that Z is a process and the pertinent class $\mathfrak{P}$ of probabilities is understood; then the natural enlargement of the basic filtration $\mathcal{F}_\bullet^0[Z]$ is called the **natural filtration** of Z and is denoted by $\mathcal{F}_\bullet[Z]$. If $\mathfrak{P}$ must be mentioned, we write $\mathcal{F}_\bullet^{\mathfrak{P}}[Z]$.*

*A measured filtration is said to satisfy the **natural conditions** if it equals its natural enlargement.*

Warning 1.3.39 The reader will find the term ***usual conditions*** at this juncture in most textbooks, instead of "natural conditions." The usual conditions require that $\mathcal{F}_\bullet$ equal its ***usual enlargement***, which is effected by replacing $\mathcal{F}_\bullet$ with its right-continuous version and throwing into every $\mathcal{F}_{t+}$, $t < \infty$, ***all*** $\mathbb{P}$-negligible sets of $\mathcal{F}_\infty$ and their subsets, i.e., all sets that are negligible for the outer measure $\mathbb{P}^*$ constructed from $(\mathcal{F}_\infty, \mathbb{P}^*)$. The latter class is generally cardinalities bigger than the class of nearly empty sets (see example 1.3.24). Doing the regularization (frequently called ***completion***) of the filtration this way evidently has the consequence that a probability absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}_0$ is already absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}_\infty$. Failure to observe this has occasionally led to vacuous investigations of the local equivalence of probabilities and to erroneous statements of Girsanov's theorem (see example 3.9.14 on page 164 and warning 3.9.20 on page 167). The term "usual conditions" was coined by the French School and is now in universal use.

We shall see in due course that definition 1.3.38 of the enlargement furnishes the advantages one expects: path regularity of integrators and a plentiful supply of stopping times, without incurring some of the disadvantages that come with too liberal an enlargement. Here is a mnemonic device: the natural conditions are obtained by adjoining the nearly empty (instead of the negligible) sets to the right-continuous version of the filtration; and they are nearly the usual conditions, but not quite: ***The natural enlargement does not in general contain every negligible set of $\mathcal{F}_\infty$!*** ▬▮

The natural conditions can of course be had by the simple expedient of replacing the given filtration with its natural enlargement – and, according

to proposition 2.2.11, doing this costs nothing so far as the stochastic integral is concerned. Here is one pretty consequence of doing such a replacement. Consider a progressively measurable subset B of the base space $\boldsymbol{B}$. The ***debut*** of B is the time (see figure 1.6)

$$D_B(\omega) \stackrel{\text{def}}{=} \inf\{t : (t,\omega) \in B\} .$$

It is shown in corollary A.5.12 that under the natural conditions D_B is a stopping time. The proof uses some capacity theory, which can be found in appendix A. Our elementary analysis of integrators won't need to employ this big result, but we shall make use of the larger supply of stopping times provided by the regularity and right-continuity and established in exercises 1.3.35 and 1.3.30.

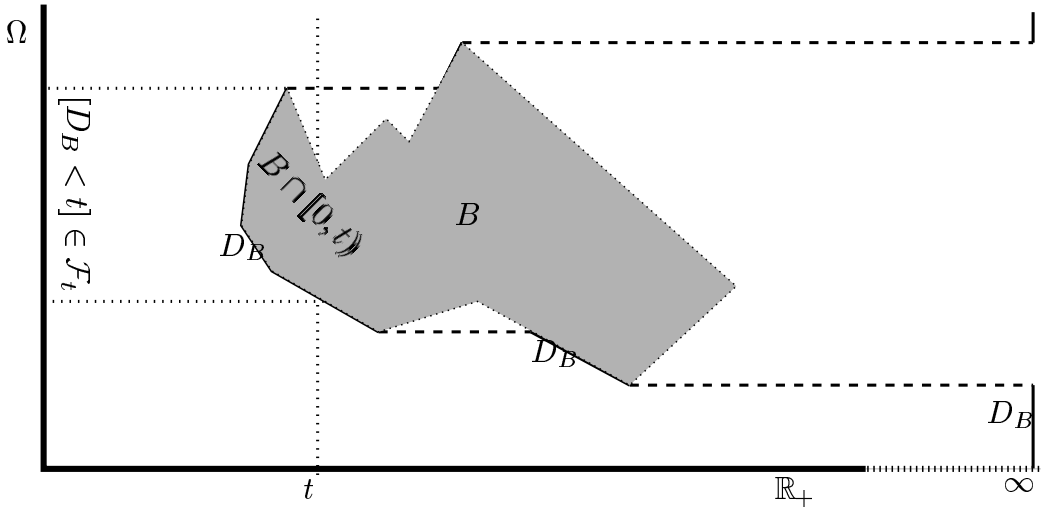

Figure 1.6 The debut of A

Exercise 1.3.40 The natural enlargement has the same nearly empty sets and evanescent processes as the original measured filtration.

Local Absolute Continuity A probability $\mathbb{P}'$ on $\mathcal{F}_\infty$ is ***locally absolutely continuous*** with respect to $\mathbb{P}$ if for all finite $t < \infty$ its restriction $\mathbb{P}'_t$ to $\mathcal{F}_t$ is absolutely continuous with respect to the restriction $\mathbb{P}_t$ of $\mathbb{P}$ to $\mathcal{F}_t$. This is evidently the same as saying that a $\mathbb{P}$-nearly empty set is $\mathbb{P}'$-nearly empty and is written $\mathbb{P}' \ll_{\bullet} \mathbb{P}$. This can very well happen without $\mathbb{P}'$ being absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}_\infty$! If both $\mathbb{P}' \ll_{\bullet} \mathbb{P}$ and $\mathbb{P} \ll_{\bullet} \mathbb{P}'$, we say that $\mathbb{P}$ and $\mathbb{P}'$ are ***locally equivalent*** and write $\mathbb{P} \approx_{\bullet} \mathbb{P}'$; it simply means that $\mathbb{P}$ and $\mathbb{P}'$ have the same nearly empty sets. For more on the subject see pages 162–167.

Exercise 1.3.41 Let $\mathbb{P}' \ll_{\bullet} \mathbb{P}$. (i) A $\mathbb{P}$-evanescent process is $\mathbb{P}'$-evanescent. (ii) $(\mathcal{F}^{\mathbb{P}}_\bullet, \mathbb{P}')$ is $\mathbb{P}'$-regular. (iii) If T is a $\mathbb{P}$-nearly finite stopping time, then a $\mathbb{P}$-negligible set in $\mathcal{F}_T$ is $\mathbb{P}'$-negligible.

Exercise 1.3.42 In order to see that the definition of local absolute continuity conforms with our usual use of the word "local" (page 51), show that $\mathbb{P}' \ll_\bullet \mathbb{P}$ if and only if there are arbitrarily large finite stopping times T so that $\mathbb{P}' \ll \mathbb{P}$ on $\mathcal{F}_T$. If $\mathfrak{P}$ is enlarged by adding every measure locally absolutely continuous with respect to some probability in $\mathfrak{P}$, then the regularization does not change. In particular, if there exists a probability $\mathbb{P}$ in $\mathfrak{P}$ with respect to which all of the others are locally absolutely continuous, then $\mathcal{F}_\bullet^{\mathfrak{P}} = \mathcal{F}_\bullet^{\mathbb{P}}$.

Exercise 1.3.43 Replacing $\mathcal{F}_t$ by $\mathcal{F}_t^{\mathfrak{P}}$ is harmless in this sense: it will increase the supply of adapted and of progressively measurable processes, but it will not change the spaces $\mathrm{L}^p[\mathcal{F}_t, \mathbb{P}]$ of equivalence classes $0 \le p \le \infty, 0 \le t \le \infty, \mathbb{P} \in \mathfrak{P}$.

Exercise 1.3.44 Construct a measured filtration $(\mathcal{F}_\bullet, \mathbb{P})$ that is not regular yet has the property that the pairs $(\mathcal{F}_t, \mathbb{P})$ all are complete measure spaces.

Example 1.3.45 Recall the measured filtration $(\Omega = [0,1], \mathcal{F}_\bullet^{\mathbb{P}}, \mathbb{P} = \lambda)$ of example 1.3.32. It is right-continuous and regular. On $\mathcal{F}_\infty = \mathcal{B}^\bullet[0,1]$ let $\mathbb{P}'$ be Dirac measure at 0.[18] Its restriction to $\mathcal{F}_t^{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$; in fact, for $n \le t < n+1$ a Radon–Nikodym derivative is $M_t \stackrel{\text{def}}{=} 2^n \cdot [0, 2^{-n}]$. So $\mathbb{P}'$ is locally absolutely continuous with respect to $\mathbb{P}$, evidently without being absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}_\infty$. For another example see theorem 3.9.19 on page 167.

Exercise 1.3.46 The set $\mathcal{N}$ of theorem 1.2.8 where the Wiener path is differentiable at at least one instant was actually nearly empty.

Exercise 1.3.47 (A Zero-One Law) Let W be a standard Wiener process on $(\Omega, \mathbb{P})$. (i) The $\mathbb{P}$-regularization of the basic filtration $\mathcal{F}_\bullet^0[W]$ is right-continuous. (ii) Set $T^\pm \stackrel{\text{def}}{=} \inf\{t > 0 : W_t \gtrless 0\}$. Then $\mathbb{P}[T^+ = 0] = \mathbb{P}[T^- = 0] = 1$; to paraphrase "$W$ starts off by oscillating about 0."

Exercise 1.3.48 A standard Wiener process[8] is ***recurrent***. That is to say, for every $s \in [0, \infty)$ and $x \in \mathbb{R}$ and almost all $\omega \in \Omega$ there is an instant $t > s$ at which $W_t(\omega) = x$.

Repeated Footnotes: 3[1] 5[2] 6[4] 8[7] 9[8] 14[12]

[18] Dirac measure at ω is the measure $A \mapsto A(\omega)$ – see convention A.1.5.