Chapter 17

Diffusions and the Wiener Process

Section 17.1 introduces the ideas which will occupy us for the next few lectures, the continuous Markov processes known as diffusions, and their description in terms of stochastic calculus.

Section 17.2 collects some useful properties of the most important diffusion, the Wiener process.

Section 17.3 shows, first heuristically and then more rigorously, that almost all sample paths of the Wiener process don't have derivatives.

17.1 Diffusions and Stochastic Calculus

So far, we have looked at Markov processes in general, and then paid particular attention to Feller processes, because the Feller properties are very natural continuity assumptions to make about stochastic models and have very important consequences, especially the strong Markov property and cadlag sample paths. The natural next step is to go to Markov processes with continuous sample paths. The most important case, overwhelmingly dominating the literature, is that of *diffusions*.

Definition 177 (Diffusion) *A stochastic process X adapted to a filtration $\mathcal{F}$ is a* diffusion *when it is a strong Markov process with respect to $\mathcal{F}$, homogeneous in time, and has continuous sample paths.*[1]

Diffusions matter to us for several reasons. First, they are very natural models of many important systems — the motion of physical particles (the

[1]Having said that, I should confess that some authors don't insist that diffusions be homogeneous, and some even don't insist that they be *strong* Markov processes. But this is the general sense in which the term is used.

source of the term "diffusion"), fluid flows, noise in communication systems, financial time series, etc. Probabilistic and statistical studies of time-series data thus need to understand diffusions. Second, many discrete Markov models have large-scale limits which are diffusion processes: these are important in physics and chemistry, population genetics, queueing and network theory, certain aspects of learning theory[2], etc. These limits are often more tractable than more exact finite-size models. (We saw a hint of this in Section 15.3.) Third, many statistical-inferential problems can be described in terms of diffusions, most prominently ones which concern goodness of fit, the convergence of empirical distributions to true probabilities, and nonparametric estimation problems of many kinds.

The easiest way to get at diffusions is to through the theory of stochastic differential equations; the most important diffusions can be thought of as, roughly speaking, the result of adding a noise term to the right-hand side of a differential equation. A more exact statement is that, just as an autonomous ordinary differential equation

$$\frac{dx}{dt} = f(x), \; x(t_0) = x_0 \tag{17.1}$$

has the solution

$$x(t) = \int_{t_0}^{t} f(x)ds + x_0 \tag{17.2}$$

a stochastic differential equation

$$\frac{dX}{dt} = f(X) + g(X)\frac{dY}{dt}, \; X(t_0) = x_0 \text{ a.s.} \tag{17.3}$$

where X and Y are stochastic processes, is solved by

$$X(t) = \int f(X)ds + \int g(X)dY + x_0 \tag{17.4}$$

where $\int g(X,t)dY$ is a *stochastic integral*. It turns out that, properly constructed, this sort of integral, and so this sort of stochastic differential equation, makes sense even when dY/dt does not make sense as any sort of ordinary derivative, so that the more usual way of writing an SDE is

$$dX = f(X)dt + g(X)dY, \; X(t_0) = x_0 \text{ a.s.} \tag{17.5}$$

even though this seems to invoke infinitessimals, which don't exist.[3]

[2]Specifically, discrete-time reinforcement learning converges to the continuous-time replicator equation of evolutionary theory.

[3]Some people, like Ethier and Kurtz (1986), prefer to talk about stochastic *integral* equations, rather than stochastic *differential* equations, because things like 17.5 are really shorthands for "find an X such that Eq. 17.4 holds", and objects like dX don't really make much sense on their own. There's a certain logic to this, but custom is overwhelmingly against them.

The fully general theory of stochastic calculus considers integration with respect to a very broad range of stochastic processes, but the original case, which is still the most important, is integration with respect to the Wiener process, which corresponds to driving a system with white noise. In addition to its many applications in all the areas which use diffusions, the theory of integration against the Wiener process occupies a central place in modern probability theory; I simply would not be doing my job if this course did not cover it. We therefore begin our study of diffusions and stochastic calculus by reviewing some of the properties of the Wiener process — which is also the most important diffusion process.

17.2 Once More with the Wiener Process and Its Properties

To review, the standard Wiener process $W(t)$ is defined by (i) $W(0) = 0$, (ii) centered Gaussian increments with linearly-growing variance, $\mathcal{L}(W(t_2) - W(t_1)) = \mathcal{N}(0, t_2 - t_1)$, (iii) independent increments and (iv) continuity of sample paths. We have seen that it is a homogeneous Markov process (Section 11.1), and in fact (Section 16.1) a Feller process (and therefore a strong Markov process), whose generator is $\frac{1}{2}\nabla^2$. By Definition 177, W is a diffusion.

This section proves a few more useful properties.

Proposition 178 *The Wiener process is a martingale with respect to its natural filtration.*

PROOF: This follows directly from the Gaussian increment property:

$$\begin{aligned}
\mathbf{E}\left[W(t+h)|\mathcal{F}_t^X\right] &= \mathbf{E}\left[W(t+h)|W(t)\right] && (17.6)\\
&= \mathbf{E}\left[W(t+h) - W(t) + W(t)|W(t)\right] && (17.7)\\
&= \mathbf{E}\left[W(t+h) - W(t)|W(t)\right] + W(t) && (17.8)\\
&= 0 + W(t) = W(t) && (17.9)
\end{aligned}$$

where the first line uses the Markov property of W, and the last line the Gaussian increments property. □

Definition 179 *If $W(t,\omega)$ is adapted to a filtration $\mathcal{F}$ and is an $\mathcal{F}$-filtration, it is an $\mathcal{F}$* Wiener process *or $\mathcal{F}$* Brownian motion*.*

It seems natural to speak of the Wiener process as a Gaussian process. This motivates the following definition.

Definition 180 (Gaussian Process) *A real-valued stochastic process is Gaussian when all its finite-dimensional distributions are multivariate Gaussian distributions.*

Proposition 181 *The Wiener process is a Gaussian process.*

PROOF: Pick any k times $t_1 < t_2 < \ldots < t_k$. Then the increments $W(t_1) - W(0)$, $W(t_2) - W(t_1)$, $W(t_3) - W(t_2)$, $\ldots$ $W(t_k) - W(t_{k-1})$ are independent Gaussian random variables. If X and Y are independent Gaussians, then $X, X+Y$ is a multivariate Gausssian, so (recursively) $W(t_1) - W(0), W(t_2) - W(0), \ldots W(t_k) - W(0)$ has a multivariate Gaussian distribution. Since $W(0) = 0$, the Gaussian distribution property follows. Since $t_1, \ldots t_k$ were arbitrary, as was k, all the finite-dimensional distributions are Gaussian. □

Just as the distribution of a Gaussian random variable is determined by its mean and covariance, the distribution of a Gaussian process is determined by its mean over time, $\mathbf{E}[X(t)]$, and its covariance function, $\text{cov}(X(s), X(t))$. (You might find it instructive to prove this *without* looking at Lemma 13.1 in Kallenberg.) Clearly, $\mathbf{E}[W(t)] = 0$, and, taking $s \leq t$ without loss of generality,

$$\begin{aligned}
\text{cov}(W(s), W(t)) &= \mathbf{E}[W(s)W(t)] - \mathbf{E}[W(s)]\,\mathbf{E}[W(t)] && (17.10)\\
&= \mathbf{E}[(W(t) - W(s) + W(s))W(s)] && (17.11)\\
&= \mathbf{E}[(W(t) - W(s))W(s)] + \mathbf{E}[W(s)W(s)] && (17.12)\\
&= \mathbf{E}[W(t) - W(s)]\,\mathbf{E}[W(s)] + s && (17.13)\\
&= s && (17.14)
\end{aligned}$$

17.3 Wiener Measure; Most Continuous Curves Are Not Differentiable

We can regard the Wiener process as establishing a measure on the space $\mathbf{C}(\mathbb{R}^+)$ of continuous real-valued functions; this is one of the considerations which led Wiener to it (Wiener, 1958)[4]. This will be important when we want to do statistical inference for stochastic processes. All Bayesian methods, and most frequentist ones, will require us to have a likelihood for the model θ given data x, $f_\theta(x)$, but likelihoods are really Radon-Nikodym derivatives, $f_\theta(x) = \frac{d\nu_\theta}{d\mu}(x)$ with respect to some reference measure μ. When our sample space is $\mathbb{R}^d$, we generally use Lebesgue measure as our reference measure, since its support is the whole space, it treats all points uniformly, and it's reasonably normalizable. Wiener measure will turn out to play a similar role when our sample space is $\mathbf{C}$.

A mathematically important question, which will also turn out to matter to us very greatly when we try to set up stochastic differential equations, is whether, under this *Wiener measure*, most curves are differentiable. If, say, almost all curves were differentiable, then it would be easy to define dW/dt. Unfortunately, this is not the case; almost all curves are nowhere differentiable.

There is an easy heuristic argument to this conclusion. $W(t)$ is a Gaussian,

[4]The early chapters of this book form a wonderfully clear introduction to Wiener measure, starting from prescriptions on the measures of finite-dimensional cylinders and building from there, deriving the incremental properties we've started with as consequences.

whose variance is t. If we look at the ratio in a derivative

$$\frac{W(t+h) - W(t)}{(t+h) - t}$$

the numerator has variance h and the denominator is the constant h, so the ratio has variance $1/h$, which goes to infinity as $h \to 0$. In other words, as we look at the curve of $W(t)$ on smaller and smaller scales, it becomes more and more erratic, and the slope finally blows up into a completely unpredictable quantity. This is basically the shape of the more rigorous argument as well.

Theorem 182 *With probability 1, $W(t)$ is nowhere-differentiable.*

PROOF: Assume, by way of contradiction, that $W(t)$ is differentiable at t_0. Then

$$\lim_{t \to t_0} \frac{W(t,\omega) - W(t_0,\omega)}{t - t_0} \tag{17.15}$$

must exist, for some set of ω of positive measure. Call its supposed value $W'(t_0,\omega)$. That is, for every $\epsilon > 0$, we must have some δ such that $t_0 - \delta \leq t \leq t_0 + \delta$ implies

$$\left|\frac{W(t,\omega) - W(t_0,\omega)}{t - t_0} - W'(t_0,\omega)\right| \leq \epsilon \tag{17.16}$$

Without loss of generality, take $t > t_0$. Then $W(t,\omega) - W(t_0,\omega)$ is independent of $W(t_0,\omega)$ and has a Gaussian distribution with mean zero and variance $t - t_0$. Therefore the differential ratio is $\mathcal{N}(0, \frac{1}{t-t_0})$. The quantity inside the absolute value sign in Eq. 17.16 is thus Gaussian with distribution $\mathcal{N}(-W'(t_0), \frac{1}{t-t_0})$. The probability that it exceeds any ϵ is therefore always positive, and in fact can be made arbitrarily large by taking t sufficiently close to t_0. Hence, with probability 1, there is no point of differentiability. □

Continuous curves which are nowhere differentiable are odd-looking beasts, but we've just established that such "pathological" cases are in fact typical, and non-pathological ones vanishingly rare in **C**. What's worse, in the functional central limit theorem (174), we obtained W as the limit of piecewise constant, and so piecewise differentiable, random functions. We could even have linearly interpolated between the points of the random walk, and those random functions would also have converged in distribution on W. The continuous, almost-everywhere-differentiable curves form a subset of **C**, and now we have a sequence of measures which give them probability 1, converging on Wiener measure, which gives them probability 0. This sounds like trouble, especially if we want to use Wiener measure as a reference measure in likelihoods, because it sounds like lots of interesting measures, which *do* produce differentiable curves, will not be absolutely continuous...

The trick here is to consider carefully our σ-algebra. Wiener measure is a probability measure on $\mathbf{R}^{\mathbf{R}^+}, \mathcal{B}^{\mathbf{R}^+} \cap \mathbf{C}(\mathbb{R}^+)$. The differentiability of a function in the vicinity of a point depends on its value at uncountably many coordinates. Hence (Exercise 1.1) it is not a member of the σ-field.