An Introduction to Stochastic Differential Equations Version 1.2

Lawrence C. Evans
Department of Mathematics
UC Berkeley

PREFACE

These notes survey, without too many precise details, the basic theory of probability, random differential equations and some applications.

Stochastic differential equations is usually, and justly, regarded as a graduate level subject. A really careful treatment assumes the students' familiarity with probability theory, measure theory, ordinary differential equations, and partial differential equations as well.

But as an experiment I tried to design these lectures so that starting graduate students (and maybe really strong undergraduates) can follow most of the theory, at the cost of some omission of detail and precision. I for instance downplayed most measure theoretic issues, but did emphasize the intuitive idea of σ–algebras as "containing information". Similarly, I "prove" many formulas by confirming them in easy cases (for simple random variables or for step functions), and then just stating that by approximation these rules hold in general. I also did not reproduce in class some of the more complicated proofs provided in these notes, although I did try to explain the guiding ideas.

My thanks especially to Lisa Goldberg, who several years ago presented my class with several lectures on financial applications, and to Fraydoun Rezakhanlou, who has taught from these notes and added several improvements.

I am also grateful to Jonathan Weare for several computer simulations illustrating the text. Thanks also to many readers who have found errors, especially Robert Piche, who provided me with an extensive list of typos and suggestions that I have incorporated into this latest version of the notes.

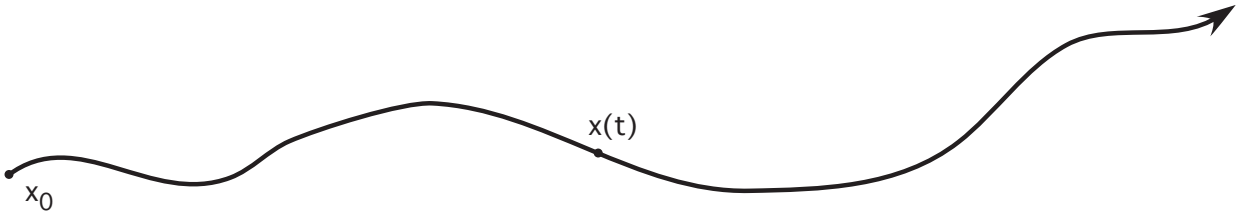
CHAPTER 1: INTRODUCTION

A. MOTIVATION

Fix a point $x_0 \in \mathbb{R}^n$ and consider then the ordinary differential equation:

$$(ODE) \qquad \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{b}(\mathbf{x}(t)) \quad (t > 0) \\ \mathbf{x}(0) = x_0, \end{cases}$$

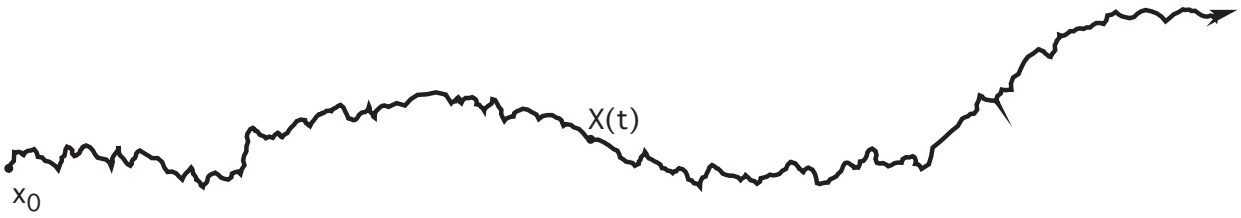
where $\mathbf{b} : \mathbb{R}^n \to \mathbb{R}^n$ is a given, smooth vector field and the solution is the trajectory $\mathbf{x}(\cdot) : [0, \infty) \to \mathbb{R}^n$.



Trajectory of the differential equation

Notation. $\mathbf{x}(t)$ is the *state of the system* at time $t \geq 0$, $\dot{\mathbf{x}}(t) := \frac{d}{dt}\mathbf{x}(t)$. □

In many applications, however, the experimentally measured trajectories of systems modeled by (ODE) do not in fact behave as predicted:



Sample path of the stochastic differential equation

Hence it seems reasonable to modify (ODE), somehow to include the possibility of random effects disturbing the system. A *formal* way to do so is to write:

$$(1) \qquad \begin{cases} \dot{\mathbf{X}}(t) = \mathbf{b}(\mathbf{X}(t)) + \mathbf{B}(\mathbf{X}(t))\boldsymbol{\xi}(t) \quad (t > 0) \\ \mathbf{X}(0) = x_0, \end{cases}$$

where $\mathbf{B} : \mathbb{R}^n \to \mathbb{M}^{n \times m}$ (= space of $n \times m$ matrices) and

$$\boldsymbol{\xi}(\cdot) := m\text{-dimensional "white noise".}$$

This approach presents us with these **mathematical problems:**

- Define the "white noise" $\boldsymbol{\xi}(\cdot)$ in a rigorous way.
- Define what it means for $\mathbf{X}(\cdot)$ to solve (1).
- Show (1) has a solution, discuss uniqueness, asymptotic behavior, dependence upon x_0, $\mathbf{b}$, $\mathbf{B}$, etc.

B. SOME HEURISTICS

Let us first study (1) in the case $m = n$, $x_0 = 0$, $\mathbf{b} \equiv 0$, and $\mathbf{B} \equiv I$. The solution of (1) in this setting turns out to be the n-dimensional *Wiener process*, or *Brownian motion*, denoted $\mathbf{W}(\cdot)$. Thus we may symbolically write

$$\dot{\mathbf{W}}(\cdot) = \boldsymbol{\xi}(\cdot),$$

thereby asserting that *"white noise" is the time derivative of the Wiener process.*

Now return to the general case of the equation (1), write $\frac{d}{dt}$ instead of the dot:

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{b}(\mathbf{X}(t)) + \mathbf{B}(\mathbf{X}(t))\frac{d\mathbf{W}(t)}{dt},$$

and finally multiply by "dt":

$$(SDE) \qquad \begin{cases} d\mathbf{X}(t) = \mathbf{b}(\mathbf{X}(t))dt + \mathbf{B}(\mathbf{X}(t))d\mathbf{W}(t) \\ \mathbf{X}(0) = x_0. \end{cases}$$

This expression, properly interpreted, is a *stochastic differential equation.* We say that $\mathbf{X}(\cdot)$ *solves* (SDE) provided

$$(2) \qquad \mathbf{X}(t) = x_0 + \int_0^t \mathbf{b}(\mathbf{X}(s))\, ds + \int_0^t \mathbf{B}(\mathbf{X}(s))\, d\mathbf{W} \quad \text{for all times } t > 0\ .$$

Now we must:

- Construct $\mathbf{W}(\cdot)$: See Chapter 3.
- Define the *stochastic integral* $\int_0^t \cdots d\mathbf{W}$: See Chapter 4.
- Show (2) has a solution, etc.: See Chapter 5.

And once all this is accomplished, there will still remain these **modeling problems:**

- Does (SDE) truly model the physical situation?
- Is the term $\boldsymbol{\xi}(\cdot)$ in (1) "really" white noise, or is it rather some ensemble of smooth, but highly oscillatory functions? See Chapter 6.

As we will see later these questions are subtle, and different answers can yield completely different solutions of (SDE). Part of the trouble is the strange form of the chain rule in the stochastic calculus:

C. ITÔ'S FORMULA

Assume $n = 1$ and $X(\cdot)$ solves the SDE

$$(3) \qquad dX = b(X)dt + dW.$$

Suppose next that $u : \mathbb{R} \to \mathbb{R}$ is a given smooth function. We ask: what stochastic differential equation does

$$Y(t) := u(X(t)) \quad (t \geq 0)$$

solve? Offhand, we would guess from (3) that

$$dY = u'dX = u'bdt + u'dW,$$

according to the usual chain rule, where $' = \frac{d}{dx}$. *This is wrong, however*! In fact, as we will see,

$$dW \approx (dt)^{1/2} \tag{4}$$

in some sense. Consequently if we compute dY and keep all terms of order dt or $(dt)^{\frac{1}{2}}$, we obtain

$$\begin{aligned} dY &= u'dX + \frac{1}{2}u''(dX)^2 + \dots \\ &= u'(\underbrace{bdt + dW}_{\text{from (3)}}) + \frac{1}{2}u''(bdt + dW)^2 + \dots \\ &= \left(u'b + \frac{1}{2}u''\right)dt + u'dW + \{\text{terms of order } (dt)^{3/2} \text{ and higher}\}. \end{aligned}$$

Here we used the "fact" that $(dW)^2 = dt$, which follows from (4). Hence

$$dY = \left(u'b + \frac{1}{2}u''\right)dt + u'dW,$$

with the extra term "$\frac{1}{2}u''dt$" not present in ordinary calculus.

A major goal of these notes is to provide a rigorous interpretation for calculations like these, involving stochastic differentials.

EXAMPLE 1. According to Itô's formula, the solution of the stochastic differential equation

$$\begin{cases} dY = YdW, \\ Y(0) = 1 \end{cases}$$

is

$$Y(t) := e^{W(t) - \frac{t}{2}},$$

and *not* what might seem the obvious guess, namely $\hat{Y}(t) := e^{W(t)}$. □

EXAMPLE 2. Let $P(t)$ denote the (random) price of a stock at time $t \geq 0$. A standard model assumes that $\frac{dP}{P}$, the relative change of price, evolves according to the SDE

$$\frac{dP}{P} = \mu dt + \sigma dW$$

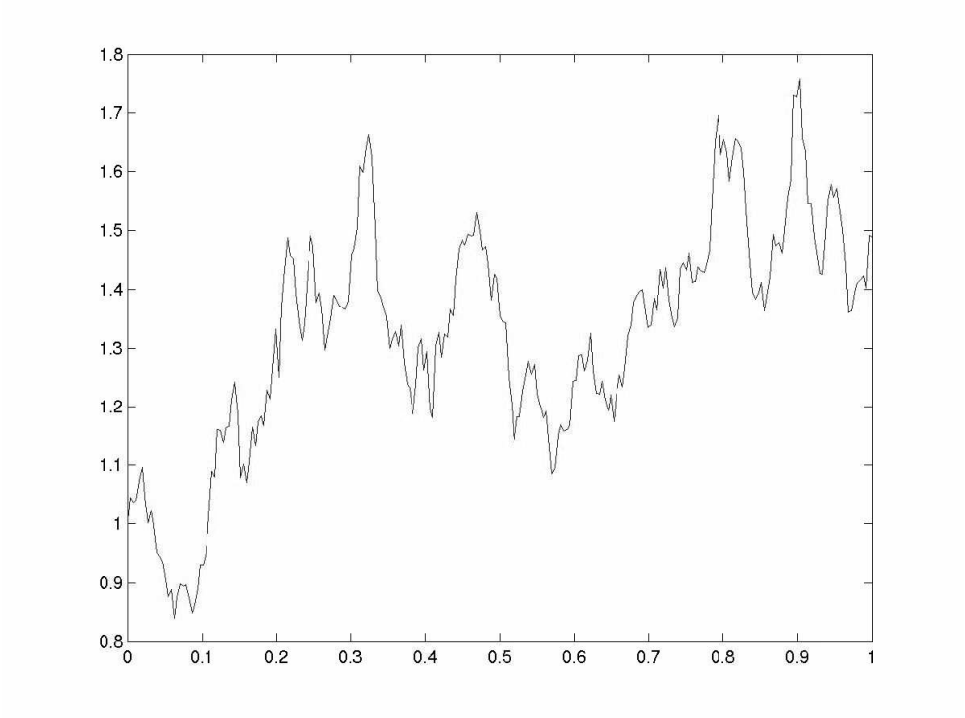
for certain constants $\mu > 0$ and σ, called respectively the *drift* and the *volatility* of the stock. In other words,

$$\begin{cases} dP = \mu P dt + \sigma P dW \\ P(0) = p_0, \end{cases}$$

where p_0 is the starting price. Using once again Itô's formula we can check that the solution is

$$P(t) = p_0 e^{\sigma W(t) + \left(\mu - \frac{\sigma^2}{2}\right)t}.$$

□



A sample path for stock prices

CHAPTER 2: A CRASH COURSE IN BASIC PROBABILITY THEORY

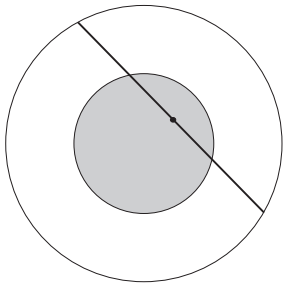
A. Basic definitions
B. Expected value, variance
C. Distribution functions
D. Independence
E. Borel–Cantelli Lemma
F. Characteristic functions
G. Strong Law of Large Numbers, Central Limit Theorem
H. Conditional expectation
I. Martingales

This chapter is a *very* rapid introduction to the measure theoretic foundations of probability theory. More details can be found in any good introductory text, for instance Bremaud [**Br**], Chung [**C**] or Lamperti [**L1**].

A. BASIC DEFINITIONS Let us begin with a puzzle:

Bertrand's paradox. Take a circle of radius 2 inches in the plane and choose a chord of this circle at random. What is the probability this chord intersects the concentric circle of radius 1 inch?

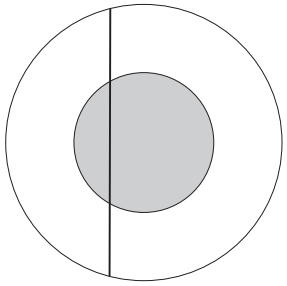
Solution #1 Any such chord (provided it does not hit the center) is uniquely determined by the location of its midpoint.



Thus

$$\text{probability of hitting inner circle} = \frac{\text{area of inner circle}}{\text{area of larger circle}} = \frac{1}{4}.$$

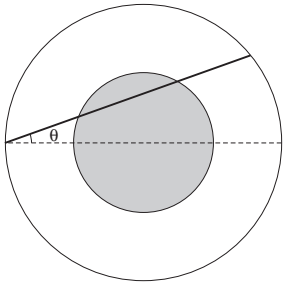
Solution #2 By symmetry under rotation we may assume the chord is vertical. The diameter of the large circle is 4 inches and the chord will hit the small circle if it falls within its 2-inch diameter.



Hence

$$\text{probability of hitting inner circle} = \frac{2 \text{ inches}}{4 \text{ inches}} = \frac{1}{2}.$$

Solution #3 By symmetry we may assume one end of the chord is at the far left point of the larger circle. The angle θ the chord makes with the horizontal lies between $\pm\frac{\pi}{2}$ and the chord hits the inner circle if θ lies between $\pm\frac{\pi}{6}$.



Therefore

$$\text{probability of hitting inner circle} = \frac{\frac{2\pi}{6}}{\frac{2\pi}{2}} = \frac{1}{3}.$$

□

PROBABILITY SPACES. This example shows that we must carefully define what we mean by the term "random". The correct way to do so is by introducing as follows the precise mathematical structure of a *probability space*.

We start with a nonempty set, denoted Ω, certain subsets of which we will in a moment interpret as being "events".

DEFINITION. A *σ-algebra* is a collection $\mathcal{U}$ of subsets of Ω with these properties:
(i) $\emptyset, \Omega \in \mathcal{U}$.
(ii) If $A \in \mathcal{U}$, then $A^c \in \mathcal{U}$.
(iii) If $A_1, A_2, \cdots \in \mathcal{U}$, then

$$\bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{\infty} A_k \in \mathcal{U}.$$

Here $A^c := \Omega - A$ is the complement of A.

DEFINITION. Let $\mathcal{U}$ be a σ-algebra of subsets of Ω. We call $P : \mathcal{U} \to [0, 1]$ a *probability measure* provided:
(i) $P(\emptyset) = 0, P(\Omega) = 1$.
(ii) If $A_1, A_2, \cdots \in \mathcal{U}$, then

$$P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k).$$

(iii) If $A_1, A_2, \ldots$ are *disjoint* sets in $\mathcal{U}$, then

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k).$$

It follows that if $A, B \in \mathcal{U}$, then

$$A \subseteq B \quad \text{implies} \quad P(A) \leq P(B).$$

DEFINITION. A triple $(\Omega, \mathcal{U}, P)$ is called a *probability space* provided Ω is any set, $\mathcal{U}$ is a σ-algebra of subsets of Ω, and P is a probability measure on $\mathcal{U}$.

Terminology. (i) A set $A \in \mathcal{U}$ is called an *event*; points $\omega \in \Omega$ are *sample points*.

(ii) $P(A)$ is the *probability* of the event A.

(iii) A property which is true except for an event of probability zero is said to hold *almost surely* (usually abbreviated "a.s.").

Example 1. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ be a finite set, and suppose we are given numbers $0 \leq p_j \leq 1$ for $j = 1, \ldots, N$, satisfying $\sum p_j = 1$. We take $\mathcal{U}$ to comprise all subsets of Ω. For each set $A = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_m}\} \in \mathcal{U}$, with $1 \leq j_1 < j_2 < \ldots j_m \leq N$, we define $P(A) := p_{j_1} + p_{j_2} + \cdots + p_{j_m}$. □

Example 2. The smallest σ-algebra containing all the open subsets of $\mathbb{R}^n$ is called the Borel σ-algebra, denoted $\mathcal{B}$. Assume that f is a nonnegative, integrable function, such that $\int_{\mathbb{R}^n} f\, dx = 1$. We define

$$P(B) := \int_B f(x)\, dx$$

for each $B \in \mathcal{B}$. Then $(\mathbb{R}^n, \mathcal{B}, P)$ is a probability space. We call f the *density* of the probability measure P. □

Example 3. Suppose instead we fix a point $z \in \mathbb{R}^n$, and now define

$$P(B) := \begin{cases} 1 & \text{if } z \in B \\ 0 & \text{if } z \notin B \end{cases}$$

for sets $B \in \mathcal{B}$. Then $(\mathbb{R}^n, \mathcal{B}, P)$ is a probability space. We call P the *Dirac mass* concentrated at the point z, and write $P = \delta_z$. □

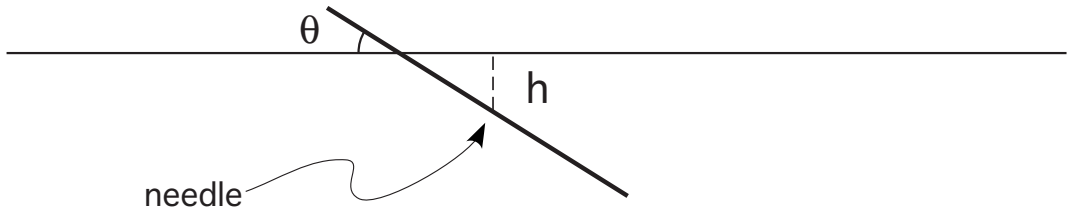
A probability space is the proper setting for mathematical probability theory. This means that we must first of all carefully identify an appropriate $(\Omega, \mathcal{U}, P)$ when we try to solve problems. The reader should convince himself or herself that the three "solutions" to Bertrand's paradox discussed above represent three distinct interpretations of the phrase "at random", that is, to three distinct models of $(\Omega, \mathcal{U}, P)$.

Here is another example.

EXAMPLE 4 (BUFFON'S NEEDLE PROBLEM). The plane is ruled by parallel lines 2 inches apart and a 1-inch long needle is dropped at random on the plane. What is the probability that it hits one of the parallel lines?

The first issue is to find some appropriate probability space $(\Omega, \mathcal{U}, P)$. For this, let

$$\begin{cases} h = \text{distance from the center of needle to nearest line,} \\ \theta = \text{angle } (\leq \frac{\pi}{2}) \text{ that the needle makes with the horizontal.} \end{cases}$$



These fully determine the position of the needle, up to translations and reflection. Let us next take

$$\begin{cases} \Omega = \underbrace{[0, \frac{\pi}{2})}_{\text{values of } \theta} \times \underbrace{[0,1]}_{\text{values of } h}, \ \mathcal{U} = \text{Borel subsets of } \Omega, \\ P(B) = \frac{2 \cdot \text{area of } B}{\pi} \quad \text{for each } B \in \mathcal{U}. \end{cases}$$

We denote by A the event that the needle hits a horizontal line. We can now check that this happens provided $\frac{h}{\sin\theta} \leq \frac{1}{2}$. Consequently $A = \{(\theta, h) \in \Omega \,|\, h \leq \frac{\sin\theta}{2}\}$, and so $P(A) = \frac{2(\text{area of } A)}{\pi} = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{1}{2} \sin\theta \, d\theta = \frac{1}{\pi}$. □

RANDOM VARIABLES. We can think of the probability space as being an essential mathematical construct, which is nevertheless not "directly observable". We are therefore interested in introducing mappings X from Ω to $\mathbb{R}^n$, the values of which we can observe.

Remember from Example 2 above that

$\mathcal{B}$ denotes the collection of *Borel subsets* of $\mathbb{R}^n$, which is the smallest σ-algebra of subsets of $\mathbb{R}^n$ containing all open sets.

We may henceforth informally just think of $\mathcal{B}$ as containing all the "nice, well-behaved" subsets of $\mathbb{R}^n$.

DEFINITION. Let $(\Omega, \mathcal{U}, P)$ be a probability space. A mapping

$$\mathbf{X} : \Omega \to \mathbb{R}^n$$

is called an n-dimensional *random variable* if for each $B \in \mathcal{B}$, we have

$$\mathbf{X}^{-1}(B) \in \mathcal{U}.$$

We equivalently say that $\mathbf{X}$ is $\mathcal{U}$-*measurable*.

Notation, comments. We usually write "$\mathbf{X}$" and *not* "$\mathbf{X}(\omega)$". This follows the custom within probability theory of mostly not displaying the dependence of random variables on the sample point $\omega \in \Omega$. We also denote $P(\mathbf{X}^{-1}(B))$ as "$P(\mathbf{X} \in B)$", the probability that $\mathbf{X}$ is in B.

In these notes we will usually use capital letters to denote random variables. Boldface usually means a vector-valued mapping.

We will also use without further comment various standard facts from measure theory, for instance that sums and products of random variables are random variables. □

Example 1. Let $A \in \mathcal{U}$. Then the *indicator function* of A,

$$\chi_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A, \end{cases}$$

is a random variable.

Example 2. More generally, if $A_1, A_2, \ldots, A_m \in \mathcal{U}$, with $\Omega = \cup_{i=1}^m A_i$, and $a_1, a_2, \ldots, a_m$ are real numbers, then

$$X = \sum_{i=1}^m a_i \chi_{A_i}$$

is a random variable, called a *simple* function. □

LEMMA. *Let $\mathbf{X} : \Omega \to \mathbb{R}^n$ be a random variable. Then*

$$\mathcal{U}(\mathbf{X}) := \{\mathbf{X}^{-1}(B) \,|\, B \in \mathcal{B}\}$$

is a σ-algebra, called the σ-algebra generated by $\mathbf{X}$. This is the smallest sub-σ-algebra of $\mathcal{U}$ with respect to which $\mathbf{X}$ is measurable.

PROOF. Check that $\{\mathbf{X}^{-1}(B) \,|\, B \in \mathcal{B}\}$ is a σ-algebra; clearly it is the smallest σ-algebra with respect to which $\mathbf{X}$ is measurable. □

IMPORTANT REMARK. It is essential to understand that, in probabilistic terms, the σ-algebra $\mathcal{U}(\mathbf{X})$ can be interpreted as "containing all relevant information" about the random variable $\mathbf{X}$.

In particular, *if a random variable Y is a function of $\mathbf{X}$*, that is, if

$$Y = \Phi(\mathbf{X})$$

for some reasonable function Φ, *then Y is $\mathcal{U}(\mathbf{X})$-measurable.*

Conversely, suppose $Y : \Omega \to \mathbb{R}$ is $\mathcal{U}(\mathbf{X})$-measurable. *Then there exists a function Φ such that*

$$Y = \Phi(\mathbf{X}).$$

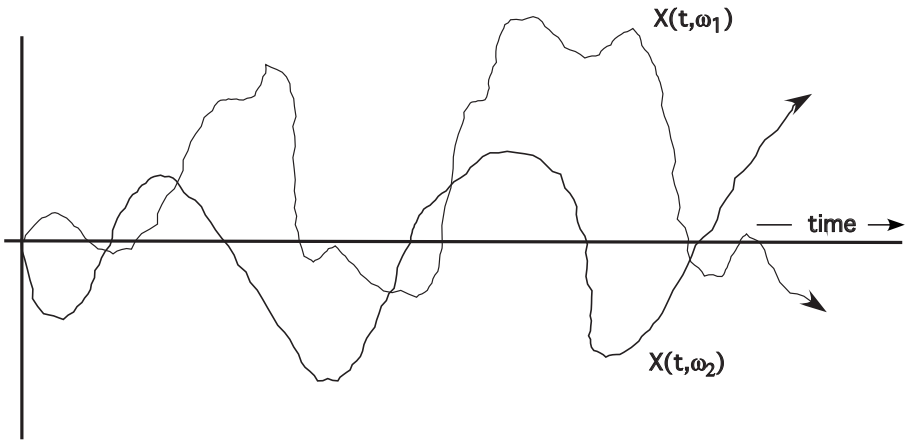
Hence if Y is $\mathcal{U}(\mathbf{X})$-measurable, Y is in fact a function of $\mathbf{X}$. Consequently if we know the value $\mathbf{X}(\omega)$, we in principle know also $Y(\omega) = \Phi(\mathbf{X}(\omega))$, although we may have no practical way to construct Φ. □

STOCHASTIC PROCESSES. We introduce next random variables depending upon time.

DEFINITIONS. (i) A collection $\{\mathbf{X}(t) \,|\, t \geq 0\}$ of random variables is called a *stochastic process.*

(ii) For each point $\omega \in \Omega$, the mapping $t \mapsto \mathbf{X}(t, \omega)$ is the corresponding *sample path.*

The idea is that if we run an experiment and observe the random values of $\mathbf{X}(\cdot)$ as time evolves, we are in fact looking at a sample path $\{\mathbf{X}(t, \omega) \,|\, t \geq 0\}$ for some fixed $\omega \in \Omega$. If we rerun the experiment, we will in general observe a different sample path.



Two sample paths of a stochastic process

B. EXPECTED VALUE, VARIANCE

Integration with respect to a measure. If $(\Omega, \mathcal{U}, P)$ is a probability space and $X = \sum_{i=1}^{k} a_i \chi_{A_i}$ is a real-valued simple random variable, we define the *integral* of X by

$$\int_\Omega X \, dP := \sum_{i=1}^{k} a_i P(A_i).$$

If next X is a *nonnegative* random variable, we define

$$\int_\Omega X \, dP := \sup_{Y \leq X, Y \text{simple}} \int_\Omega Y \, dP.$$

Finally if $X : \Omega \to \mathbb{R}$ is a random variable, we write

$$\int_\Omega X \, dP := \int_\Omega X^+ \, dP - \int_\Omega X^- \, dP,$$

provided at least one of the integrals on the right is finite. Here $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$; so that $X = X^+ - X^-$.

Next, suppose $\mathbf{X} : \Omega \to \mathbb{R}^n$ is a vector-valued random variable, $\mathbf{X} = (X^1, X^2, \ldots, X^n)$. Then we write

$$\int_\Omega \mathbf{X} \, dP = \left(\int_\Omega X^1 \, dP, \int_\Omega X^2 \, dP, \cdots, \int_\Omega X^n \, dP \right).$$

We will assume without further comment the usual rules for these integrals. □

DEFINITION. We call

$$E(\mathbf{X}) := \int_\Omega \mathbf{X}\, dP$$

the *expected value* (or *mean value*) of $\mathbf{X}$.

DEFINITION. We call

$$V(\mathbf{X}) := \int_\Omega |\mathbf{X} - E(\mathbf{X})|^2\, dP$$

the *variance* of $\mathbf{X}$, where $|\cdot|$ denotes the Euclidean norm.

Observe that

$$V(\mathbf{X}) = E(|\mathbf{X} - E(\mathbf{X})|^2) = E(|\mathbf{X}|^2) - |E(\mathbf{X})|^2.$$

LEMMA (Chebyshev's inequality). *If $\mathbf{X}$ is a random variable and $1 \leq p < \infty$, then*

$$P(|\mathbf{X}| \geq \lambda) \leq \frac{1}{\lambda^p} E(|\mathbf{X}|^p) \quad \textit{for all } \lambda > 0.$$

Proof. We have

$$E(|\mathbf{X}|^p) = \int_\Omega |\mathbf{X}|^p\, dP \geq \int_{\{|\mathbf{X}| \geq \lambda\}} |\mathbf{X}|^p\, dP \geq \lambda^p P(|\mathbf{X}| \geq \lambda).$$

□

C. DISTRIBUTION FUNCTIONS

Let $(\Omega, \mathcal{U}, P)$ be a probability space and suppose $\mathbf{X} : \Omega \to \mathbb{R}^n$ is a random variable.

Notation. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Then

$$x \leq y$$

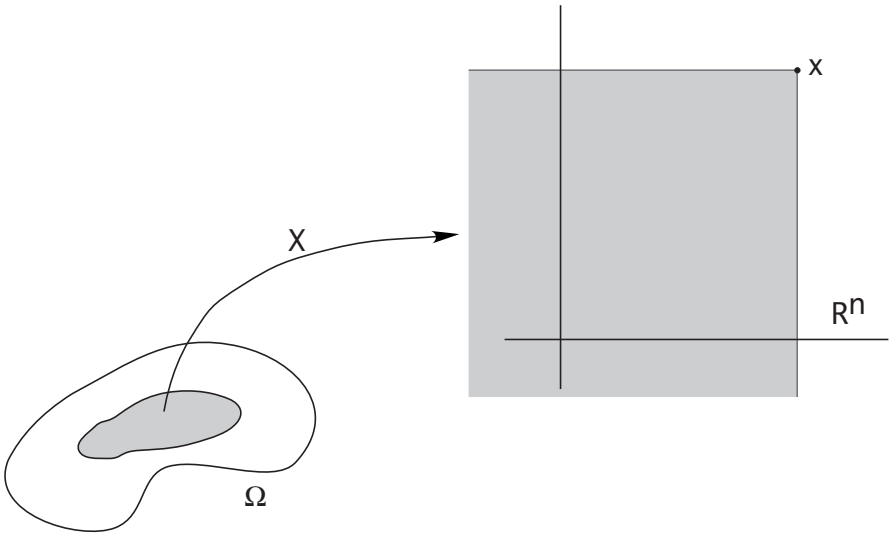
means $x_i \leq y_i$ for $i = 1, \ldots, n$. □

DEFINITIONS. (i) The *distribution function* of $\mathbf{X}$ is the function $F_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$ defined by

$$F_{\mathbf{X}}(x) := P(\mathbf{X} \leq x) \quad \text{for all } x \in \mathbb{R}^n$$

(ii) If $\mathbf{X}_1, \ldots, \mathbf{X}_m : \Omega \to \mathbb{R}^n$ are random variables, their *joint distribution function* is $F_{\mathbf{X}_1, \ldots, \mathbf{X}_m} : (\mathbb{R}^n)^m \to [0, 1]$,

$$F_{\mathbf{X}_1, \ldots, \mathbf{X}_m}(x_1, \ldots, x_m) := P(\mathbf{X}_1 \leq x_1, \ldots, \mathbf{X}_m \leq x_m) \quad \text{for all } x_i \in \mathbb{R}^n,\ i = 1, \ldots, m.$$



DEFINITION. Suppose $\mathbf{X} : \Omega \to \mathbb{R}^n$ is a random variable and $F = F_{\mathbf{X}}$ its distribution function. If there exists a nonnegative, integrable function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$F(x) = F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \ldots, y_n)\, dy_n \ldots dy_1,$$

then f is called the *density* function for $\mathbf{X}$.

It follows then that

$$P(\mathbf{X} \in B) = \int_B f(x)\, dx \quad \text{for all } B \in \mathcal{B} \tag{1}$$

This formula is important as the expression on the right hand side is an ordinary integral, and can often be explicitly calculated.

EXAMPLE 1. If $X : \Omega \to \mathbb{R}$ has density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|x-m|^2}{2\sigma^2}} \quad (x \in \mathbb{R}),$$

we say X has a *Gaussian* (or *normal*) distribution, with mean m and variance σ^2. In this case let us write

$$X \text{ is an } N(m, \sigma^2) \text{ random variable.}$$

EXAMPLE 2. If $\mathbf{X} : \Omega \to \mathbb{R}^n$ has density

$$f(x) = \frac{1}{((2\pi)^n \det C)^{1/2}} e^{-\frac{1}{2}(x-m)\cdot C^{-1}(x-m)} \qquad (x \in \mathbb{R}^n)$$

for some $m \in \mathbb{R}^n$ and some positive definite, symmetric matrix C, we say $\mathbf{X}$ has a *Gaussian* (or *normal*) distribution, with mean m and covariance matrix C. We then write

$$\mathbf{X} \text{ is an } N(m, C) \text{ random variable.}$$

□

LEMMA. *Let $\mathbf{X} : \Omega \to \mathbb{R}^n$ be a random variable, and assume that its distribution function $F = F_{\mathbf{X}}$ has the density f. Suppose $g : \mathbb{R}^n \to \mathbb{R}$, and*

$$Y = g(\mathbf{X})$$

is integrable. Then

$$E(Y) = \int_{\mathbb{R}^n} g(x) f(x)\, dx.$$

In particular,

$$E(\mathbf{X}) = \int_{\mathbb{R}^n} x f(x)\, dx \quad \textit{and} \quad V(\mathbf{X}) = \int_{\mathbb{R}^n} |x - E(\mathbf{X})|^2 f(x)\, dx.$$

IMPORTANT REMARK. Hence we can compute $E(\mathbf{X})$, $V(\mathbf{X})$, etc. in terms of integrals over $\mathbb{R}^n$. This is an important observation, since as mentioned before the probability space $(\Omega, \mathcal{U}, P)$ is "unobservable": all that we "see" are the values $\mathbf{X}$ takes on in $\mathbb{R}^n$. Indeed, *all quantities of interest in probability theory can be computed in $\mathbb{R}^n$ in terms of the density f.* □

PROOF. Suppose first g is a simple function on $\mathbb{R}^n$:

$$g = \sum_{i=1}^{m} b_i \chi_{B_i} \qquad (B_i \in \mathcal{B}).$$

Then

$$E(g(\mathbf{X})) = \sum_{i=1}^{m} b_i \int_{\Omega} \chi_{B_i}(\mathbf{X})\, dP = \sum_{i=1}^{m} b_i P(\mathbf{X} \in B_i).$$

But also

$$\begin{aligned} \int_{\mathbb{R}^n} g(x) f(x)\, dx &= \sum_{i=1}^{m} b_i \int_{B_i} f(x)\, dx \\ &= \sum_{i=1}^{m} b_i P(\mathbf{X} \in B_i) \quad \text{by (1).} \end{aligned}$$

Consequently the formula holds for all simple functions g and, by approximation, it holds therefore for general functions g. □

EXAMPLE. If X is $N(m, \sigma^2)$, then

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = m$$

and

$$V(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2.$$

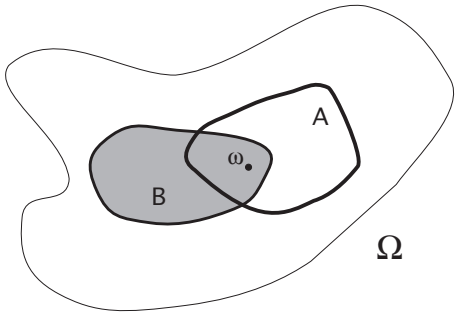
Therefore m is indeed the mean, and σ^2 the variance. □

D. INDEPENDENCE

MOTIVATION. Let $(\Omega, \mathcal{U}, P)$ be a probability space, and let $A, B \in \mathcal{U}$ be two events, with $P(B) > 0$. We want to find a reasonable definition of

$$P(A \,|\, B), \text{ the probability of } A, \text{ given } B.$$

Think this way. Suppose some point $\omega \in \Omega$ is selected "at random" and we are told $\omega \in B$. What then is the probability that $\omega \in A$ also?



Since we know $\omega \in B$, we can regard B as being a new probability space. Therefore we can define $\tilde{\Omega} := B$, $\tilde{\mathcal{U}} := \{C \cap B \,|\, C \in \mathcal{U}\}$ and $\tilde{P} := \frac{P}{P(B)}$; so that $\tilde{P}(\tilde{\Omega}) = 1$. Then the probability that ω lies in A is $\tilde{P}(A \cap B) = \frac{P(A \cap B)}{P(B)}$.

This observation motivates the following

DEFINITION. We write

$$P(A \,|\, B) := \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) > 0.$$

Now what should it mean to say "A and B are *independent*"? This should mean $P(A \,|\, B) = P(A)$, since presumably any information that the event B has occurred is irrelevant in determining the probability that A has occurred. Thus

$$P(A) = P(A \,|\, B) = \frac{P(A \cap B)}{P(B)}$$

and so

$$P(A \cap B) = P(A)P(B)$$

if $P(B) > 0$. We take this for the definition, even if $P(B) = 0$:

DEFINITION. Two events A and B are called *independent* if

$$P(A \cap B) = P(A)P(B).$$

This concept and its ramifications are the hallmarks of probability theory.

To gain some insight, the reader may wish to check that if A and B are independent events, then so are A^c and B. Likewise, A^c and B^c are independent.

DEFINITION. Let $A_1, \ldots, A_n, \ldots$ be events. These events are *independent* if for all choices $1 \leq k_1 < k_2 < \cdots < k_m$, we have

$$P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_m}) = P(A_{k_1})P(A_{k_1}) \cdots P(A_{k_m}).$$

It is important to extend this definition to σ-algebras:

DEFINITION. Let $\mathcal{U}_i \subseteq \mathcal{U}$ be σ-algebras, for $i = 1, \ldots$. We say that $\{\mathcal{U}_i\}_{i=1}^{\infty}$ are *independent* if for all choices of $1 \leq k_1 < k_2 < \cdots < k_m$ and of events $A_{k_i} \in \mathcal{U}_{k_i}$, we have

$$P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_m}) = P(A_{k_1})P(A_{k_2}) \ldots P(A_{k_m}).$$

Lastly, we transfer our definitions to random variables:

DEFINITION. Let $\mathbf{X}_i : \Omega \to \mathbb{R}^n$ be random variables $(i = 1, \ldots)$. We say the random variables $\mathbf{X}_1, \ldots$ are *independent* if for all integers $k \geq 2$ and all choices of Borel sets $B_1, \ldots B_k \subseteq \mathbb{R}^n$:

$$P(\mathbf{X}_1 \in B_1, \mathbf{X}_2 \in B_2, \ldots, \mathbf{X}_k \in B_k) = P(\mathbf{X}_1 \in B_1)P(\mathbf{X}_2 \in B_2) \cdots P(\mathbf{X}_k \in B_k).$$

This is equivalent to saying that the σ-algebras $\{\mathcal{U}(X_i)\}_{i=1}^{\infty}$ are independent.

EXAMPLE. Take $\Omega = [0, 1)$, $\mathcal{U}$ the Borel subsets of $[0, 1)$, and P Lebesgue measure.

Define for $n = 1, 2, \ldots$

$$X_n(\omega) := \begin{cases} 1 & \text{if } \frac{k}{2^n} \leq \omega < \frac{k+1}{2^n},\ k \textit{ even} \\ -1 & \text{if } \frac{k}{2^n} \leq \omega < \frac{k+1}{2^n},\ k \textit{ odd} \end{cases} \qquad (0 \leq \omega < 1).$$

These are the *Rademacher functions*, which we assert are in fact independent random variables. To prove this, it suffices to verify

$$P(\mathbf{X}_1 = e_1, \mathbf{X}_2 = e_2, \ldots, \mathbf{X}_k = e_k) = P(\mathbf{X}_1 = e_1)P(\mathbf{X}_2 = e_2) \cdots P(\mathbf{X}_k = e_k),$$

for all choices of $e_1, \ldots, e_k \in \{-1, 1\}$. This can be checked by showing that both sides are equal to 2^{-k}. □

LEMMA. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_{m+n}$ *be* independent $\mathbb{R}^k$*-valued random variables. Suppose* $f : (\mathbb{R}^k)^n \to \mathbb{R}$ *and* $g : (\mathbb{R}^k)^m \to \mathbb{R}$. *Then*

$$\mathbf{Y} := f(\mathbf{X}_1, \ldots, \mathbf{X}_n) \textit{ and } \mathbf{Z} := g(\mathbf{X}_{n+1}, \ldots, \mathbf{X}_{n+m})$$

are independent.

We omit the proof, which may be found in Breiman [**B**].

THEOREM. *The random variables* $\mathbf{X}_1, \cdots, \mathbf{X}_m : \Omega \to \mathbb{R}^n$ *are independent* if and only if

(2)
$$F_{\mathbf{X}_1,\cdots,\mathbf{X}_m}(x_1,\ldots,x_m) = F_{\mathbf{X}_1}(x_1)\cdots F_{\mathbf{X}_m}(x_m) \quad \textit{for all } x_i \in \mathbb{R}^n,\ i = 1,\ldots,m.$$

If the random variables have densities, (2) *is equivalent to*

(3)
$$f_{\mathbf{X}_1,\cdots,\mathbf{X}_m}(x_1,\ldots,x_m) = f_{\mathbf{X}_1}(x_1)\cdots f_{\mathbf{X}_m}(x_m) \quad \textit{for all } x_i \in \mathbb{R}^n,\ i = 1,\ldots,m,$$

where the functions f *are the appropriate densities.*

PROOF. 1. Assume first that $\{\mathbf{X}_k\}_{k=1}^m$ are independent. Then

$$\begin{aligned} F_{\mathbf{X}_1\cdots\mathbf{X}_m}(x_1,\ldots,x_m) &= P(\mathbf{X}_1 \le x_1,\ldots,\mathbf{X}_m \le x_m) \\ &= P(\mathbf{X}_1 \le x_1)\cdots P(\mathbf{X}_m \le x_m) \\ &= F_{\mathbf{X}_1}(x_1)\cdots F_{\mathbf{X}_m}(x_m). \end{aligned}$$

2. We prove the converse statement for the case that all the random variables have densities. Select $A_i \in \mathcal{U}(\mathbf{X}_i), i = 1,\ldots,m$. Then $A_i = \mathbf{X}_i^{-1}(B_i)$ for some $B_i \in \mathcal{B}$. Hence

$$\begin{aligned} P(A_1 \cap \cdots \cap A_m) &= P(\mathbf{X}_1 \in B_1,\ldots,\mathbf{X}_m \in B_m) \\ &= \int_{B_1\times\ldots\times B_m} f_{\mathbf{X}_1\cdots\mathbf{X}_m}(x_1,\ldots,x_m)\,dx_1\cdots dx_m \\ &= \left(\int_{B_1} f_{\mathbf{X}_1}(x_1)\,dx_1\right)\ldots\left(\int_{B_m} f_{\mathbf{X}_m}(x_m)\,dx_m\right) \text{ by (3)} \\ &= P(\mathbf{X}_1 \in B_1)\cdots P(\mathbf{X}_m \in B_m) \\ &= P(A_1)\cdots P(A_m). \end{aligned}$$

Therefore $\mathcal{U}(\mathbf{X}_1),\cdots,\mathcal{U}(\mathbf{X}_m)$ are independent σ-algebras. □

One of the most important properties of independent random variables is this:

THEOREM. *If* $X_1,\ldots,X_m$ *are independent, real-valued random variables, with*

$$E(|X_i|) < \infty \quad (i = 1,\ldots,m),$$

then $E(|X_1\cdots X_m|) < \infty$ *and*

$$E(X_1\cdots X_m) = E(X_1)\cdots E(X_m).$$

PROOF. Suppose that each X_i is bounded and has a density. Then

$$\begin{aligned} E(X_1\cdots X_m) &= \int_{\mathbb{R}^m} x_1\cdots x_m\, f_{X_1\cdots X_m}(x_1,\ldots,x_m)\,dx_1\ldots x_m \\ &= \left(\int_{\mathbb{R}} x_1\, f_{X_1}(x_1)\,dx_1\right)\cdots\left(\int_{\mathbb{R}} x_m\, f_{X_m}(x_m)\,dx_m\right) \text{ by (3)} \\ &= E(X_1)\cdots E(X_m). \end{aligned}$$

□

THEOREM. *If $X_1, \ldots, X_m$ are independent, real-valued random variables, with*

$$V(X_i) < \infty \quad (i = 1, \ldots, m),$$

then

$$V(X_1 + \cdots + X_m) = V(X_1) + \cdots + V(X_m).$$

PROOF. Use induction, the case $m = 2$ holding as follows. Let $m_1 := EX_1$, $m_2 := E(X_2)$. Then $E(X_1 + X_2) = m_1 + m_2$ and

$$\begin{aligned} V(X_1 + X_2) &= \int_\Omega (X_1 + X_2 - (m_1 + m_2))^2 \, dP \\ &= \int_\Omega (X_1 - m_1)^2 \, dP + \int_\Omega (X_2 - m_2)^2 \, dP \\ &\qquad + 2 \int_\Omega (X_1 - m_1)(X_2 - m_2) \, dP \\ &= V(X_1) + V(X_2) + 2 \underbrace{E(X_1 - m_1)}_{=0} \underbrace{E(X_2 - m_2)}_{=0}, \end{aligned}$$

where we used independence in the next last step. □

E. BOREL–CANTELLI LEMMA

We introduce next a simple and very useful way to check if some sequence $A_1, \ldots, A_n, \ldots$ of events "occurs infinitely often".

DEFINITION. Let $A_1, \ldots, A_n, \ldots$ be events in a probability space. Then the event

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega \,|\, \omega \text{ belongs to infinitely many of the } A_n\},$$

is called "A_n *infinitely often*", abbreviated "A_n *i.o.*". □

BOREL–CANTELLI LEMMA. *If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$.*

PROOF. By definition A_n i.o. $= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$, and so for each n

$$P(A_n \text{ i.o.}) \le P\left(\bigcup_{m=n}^{\infty} A_m \right) \le \sum_{m=n}^{\infty} P(A_m).$$

The limit of the left-hand side is zero as $n \to \infty$ because $\sum P(A_m) < \infty$. □

APPLICATION. We illustrate a typical use of the Borel–Cantelli Lemma.

A sequence of random variables $\{X_k\}_{k=1}^{\infty}$ defined on some probability space *converges in probability* to a random variable X, provided

$$\lim_{k \to \infty} P(|X_k - X| > \epsilon) = 0$$

for each $\epsilon > 0$.

THEOREM. *If $X_k \to X$ in probability, then there exists a subsequence $\{X_{k_j}\}_{j=1}^{\infty} \subset \{X_k\}_{k=1}^{\infty}$ such that*

$$X_{k_j}(\omega) \to X(\omega) \quad \textit{for almost every } \omega.$$

PROOF. For each positive integer j we select k_j so large that

$$P(|X_{k_j} - X| > \frac{1}{j}) \leq \frac{1}{j^2},$$

and also $\dots k_{j-1} < k_j < \dots, k_j \to \infty$. Let $A_j := \{|X_{k_j} - X| > \frac{1}{j}\}$. Since $\sum \frac{1}{j^2} < \infty$, the Borel–Cantelli Lemma implies $P(A_j \ i.o.) = 0$. Therefore for almost all sample points ω, $|X_{k_j}(\omega) - X(\omega)| \leq \frac{1}{j}$ provided $j \geq J$, for some index J depending on ω. □

F. CHARACTERISTIC FUNCTIONS

It is convenient to introduce next a clever integral transform, which will later provide us with a useful means to identify normal random variables.

DEFINITION. Let $\mathbf{X}$ be an $\mathbb{R}^n$-valued random variable. Then

$$\phi_{\mathbf{X}}(\lambda) := E(e^{i\lambda \cdot \mathbf{X}}) \qquad (\lambda \in \mathbb{R}^n)$$

is the *characteristic function* of $\mathbf{X}$. □

EXAMPLE. If the real-valued random variable X is $N(m, \sigma^2)$, then

$$\phi_X(\lambda) = e^{im\lambda - \frac{\lambda^2\sigma^2}{2}} \qquad (\lambda \in \mathbb{R}).$$

To see this, let us suppose that $m = 0, \sigma = 1$ and calculate

$$\phi_X(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{e^{\frac{-\lambda^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-i\lambda)^2}{2}} dx.$$

We move the path of integration in the complex plane from the line $\{Im(z) = -\lambda\}$ to the real axis, and recall that $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$. (Here $Im(z)$ means the imaginary part of the complex number z.) Hence $\phi_X(\lambda) = e^{-\frac{\lambda^2}{2}}$. □

LEMMA. (i) *If $\mathbf{X}_1, \dots, \mathbf{X}_m$ are independent random variables, then for each $\lambda \in \mathbb{R}^n$*

$$\phi_{\mathbf{X}_1 + \dots + \mathbf{X}_m}(\lambda) = \phi_{\mathbf{X}_1}(\lambda) \dots \phi_{\mathbf{X}_m}(\lambda).$$

(ii) *If X is a real-valued random variable,*

$$\phi^{(k)}(0) = i^k E(X^k) \quad (k = 0, 1, \dots).$$

(iii) *If $\mathbf{X}$ and $\mathbf{Y}$ are random variables and*

$$\phi_{\mathbf{X}}(\lambda) = \phi_{\mathbf{Y}}(\lambda) \quad \textit{for all } \lambda,$$

then

$$F_X(x) = F_Y(x) \quad \textit{for all } x.$$

Assertion (iii) says the characteristic function of $\mathbf{X}$ determines the distribution of $\mathbf{X}$.

PROOF. 1. Let us calculate

$$\begin{aligned}\phi_{\mathbf{X}_1+\cdots+\mathbf{X}_m}(\lambda) &= E(e^{i\lambda\cdot(\mathbf{X}_1+\cdots+\mathbf{X}_m)}) \\ &= E(e^{i\lambda\cdot\mathbf{X}_1}e^{i\lambda\cdot\mathbf{X}_2}\cdots e^{i\lambda\cdot\mathbf{X}_m}) \\ &= E(e^{i\lambda\cdot\mathbf{X}_1})\cdots E(e^{i\lambda\cdot\mathbf{X}_m}) \quad \text{by independence} \\ &= \phi_{\mathbf{X}_1}(\lambda)\ldots\phi_{\mathbf{X}_m}(\lambda).\end{aligned}$$

2. We have $\phi'(\lambda) = iE(Xe^{i\lambda X})$, and so $\phi'(0) = iE(X)$. The formulas in (ii) for $k = 2, \ldots$ follow similarly.

3. See Breiman [**B**] for the proof of (iii). □

EXAMPLE. If X and Y are independent, real-valued random variables, and if X is $N(m_1, \sigma_1^2)$, Y is $N(m_2, \sigma_2^2)$, then

$$X + Y \text{ is } N(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

To see this, just calculate

$$\begin{aligned}\phi_{X+Y}(\lambda) = \phi_X(\lambda)\phi_Y(\lambda) &= e^{im_1\lambda - \frac{\lambda^2\sigma_1^2}{2}}e^{im_2\lambda - \frac{\lambda^2\sigma_2^2}{2}} \\ &= e^{i(m_1+m_2)\lambda - \frac{\lambda^2}{2}(\sigma_1^2+\sigma_2^2)}.\end{aligned}$$

□

G. STRONG LAW OF LARGE NUMBERS, CENTRAL LIMIT THEOREM

This section discusses a mathematical model for "repeated, independent experiments".

The idea is this. Suppose we are given a probability space and on it a real–valued random variable X, which records the outcome of some sort of random experiment. We can model repetitions of this experiment by introducing a sequence of random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$, each of which "has the same probabilistic information as X":

DEFINITION. A sequence $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ of random variables is called *identically distributed* if

$$F_{\mathbf{X}_1}(x) = F_{\mathbf{X}_2}(x) = \cdots = F_{\mathbf{X}_n}(x) = \ldots \quad \text{for all } x.$$

If we additionally assume that the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ are independent, we can regard this sequence as a model for repeated and independent runs of the experiment, the outcomes of which we can measure. More precisely, imagine that a "random" sample point $\omega \in \Omega$ is given and we can observe the sequence of values $\mathbf{X}_1(\omega), \mathbf{X}_2(\omega), \ldots, \mathbf{X}_n(\omega), \ldots$. What can we infer from these observations?

STRONG LAW OF LARGE NUMBERS. First we show that with probability one, we can deduce the common expected values of the random variables.

THEOREM (Strong Law of Large Numbers). *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ be a sequence of independent, identically distributed, integrable random variables defined on the same probability space.*

Write $m := E(\mathbf{X}_i)$ for $i = 1, \ldots$. Then

$$P\left(\lim_{n\to\infty} \frac{\mathbf{X}_1 + \cdots + \mathbf{X}_n}{n} = m\right) = 1.$$

Proof. 1. Supposing that the random variables are real–valued entails no loss of generality. We will as well suppose for simplicity that

$$E(X_i^4) < \infty \quad (i = 1, \ldots).$$

We may also assume $m = 0$, as we could otherwise consider $X_i - m$ in place of X_i.

2. Then

$$E\left(\left(\sum_{i=1}^n X_i\right)^4\right) = \sum_{i,j,k,l=1}^n E(X_i X_j X_k X_l).$$

If $i \neq j$, k, or l, independence implies

$$E(X_i X_j X_k X_l) = \underbrace{E(X_i)}_{=0} E(X_j X_k X_l).$$

Consequently, since the X_i are identically distributed, we have

$$\begin{aligned} E\left(\left(\sum_{i=1}^n X_i\right)^4\right) &= \sum_{i=1}^n E(X_i^4) + 3\sum_{\substack{i,j=1 \\ i\neq j}}^n E(X_i^2 X_j^2) \\ &= nE(X_1^4) + 3(n^2 - n)(E(X_1^2))^2 \\ &\leq n^2 C \end{aligned}$$

for some constant C.

Now fix $\varepsilon > 0$. Then

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \geq \varepsilon\right) &= P\left(\left|\sum_{i=1}^n X_i\right| \geq \varepsilon n\right) \\ &\leq \frac{1}{(\varepsilon n)^4} E\left(\left(\sum_{i=1}^n X_i\right)^4\right) \\ &\leq \frac{C}{\varepsilon^4}\frac{1}{n^2}. \end{aligned}$$

We used here the Chebyshev inequality. By the Borel–Cantelli Lemma, therefore,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \geq \varepsilon \text{ i.o.}\right) = 0.$$

3. Take $\varepsilon = \frac{1}{k}$. The foregoing says that

$$\limsup_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) \right| \leq \frac{1}{k},$$

except possibly for ω lying in an event B_k, with $P(B_k) = 0$. Write $B := \cup_{k=1}^{\infty} B_k$. Then $P(B) = 0$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = 0$$

for each sample point $\omega \notin B$. □

FLUCTUATIONS, LAPLACE–DE MOIVRE THEOREM. The Strong Law of Large Numbers says that for almost every sample point $\omega \in \Omega$,

$$\frac{X_1(\omega) + \cdots + X_n(\omega)}{n} \to m \quad \text{as } n \to \infty.$$

We turn next to the Laplace–De Moivre Theorem, and its generalization the Central Limit Theorem, which estimate the "fluctuations" we can expect in this limit.

Let us start with a simple calculation.

LEMMA. *Suppose the real–valued random variables $X_1, \ldots, X_n, \ldots$ are independent and identically distributed, with*

$$\begin{cases} P(X_i = 1) = p \\ P(X_i = 0) = q \end{cases}$$

for $p, q \geq 0$, $p + q = 1$. Then

$$E(X_1 + \cdots + X_n) = np$$
$$V(X_1 + \cdots + X_n) = npq.$$

PROOF. $E(X_1) = \int_\Omega X_1 \, dP = p$ and therefore $E(X_1 + \cdots + X_n) = np$. Also,

$$\begin{aligned} V(X_1) &= \int_\Omega (X_1 - p)^2 \, dP = (1-p)^2 P(X_1 = 1) + p^2 P(X_1 = 0) \\ &= q^2 p + p^2 q = qp. \end{aligned}$$

By independence, $V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n) = npq$. □

We can imagine these random variables as modeling for example repeated tosses of a biased coin, which has probability p of coming up heads, and probability $q = 1 - p$ of coming up tails.

THEOREM (Laplace–De Moivre). *Let $X_1, \ldots, X_n$ be the independent, identically distributed, real–valued random variables in the preceding Lemma. Define the sums*

$$S_n := X_1 + \cdots + X_n.$$

Then for all $-\infty < a < b < +\infty$,

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx.$$

A proof is in Appendix A.

Interpretation of the Laplace–De Moivre Theorem. In view of the Lemma,

$$\frac{S_n - np}{\sqrt{npq}} = \frac{S_n - E(S_n)}{V(S_n)^{1/2}}.$$

Hence the Laplace–De Moivre Theorem says that the sums S_n, properly renormalized, have a distribution which tends to the Gaussian $N(0,1)$ as $n \to \infty$.

Consider in particular the situation $p = q = \frac{1}{2}$. Suppose $a > 0$; then

$$\lim_{n\to\infty} P\left(-\frac{a\sqrt{n}}{2} \le S_n - \frac{n}{2} \le \frac{a\sqrt{n}}{2}\right) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{-\frac{x^2}{2}}\, dx.$$

If we fix $b > 0$ and write $a = \frac{2b}{\sqrt{n}}$, then for large n

$$P\left(-b \le S_n - \frac{n}{2} \le b\right) \approx \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\frac{2b}{\sqrt{n}}}^{\frac{2b}{\sqrt{n}}} e^{-\frac{x^2}{2}}\, dx}_{\to 0 \text{ as } n\to\infty.}$$

Thus for almost every ω, $\frac{1}{n} S_n(\omega) \to \frac{1}{2}$, in accord with the Strong Law of Large Numbers; but $\left|S_n(\omega) - \frac{n}{2}\right|$ "fluctuates" with probability 1 to exceed any finite bound b. □

CENTRAL LIMIT THEOREM. We now generalize the Laplace–De Moivre Theorem:

THEOREM (Central Limit Theorem). *Let $X_1, \ldots, X_n, \ldots$ be independent, identically distributed, real-valued random variables with*

$$E(X_i) = m,\ V(X_i) = \sigma^2 > 0.$$

for $i = 1, \ldots$. Set

$$S_n := X_1 + \cdots + X_n.$$

Then for all $-\infty < a < b < +\infty$

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - nm}{\sqrt{n}\sigma} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx. \tag{1}$$

Thus the conclusion of the Laplace–De Moivre Theorem holds not only for the 0– or 1–valued random variable considered before, but for any sequence of independent, identically distributed random variables with finite variance. We will later invoke this assertion to motivate our requirement that Brownian motion be normally distributed for each time $t \geq 0$.

OUTLINE OF PROOF. For simplicity assume $m = 0$, $\sigma = 1$, since we can always rescale to this case. Then

$$\phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = \phi_{\frac{X_1}{\sqrt{n}}}(\lambda) \dots \phi_{\frac{X_n}{\sqrt{n}}}(\lambda) = \left(\phi_{X_1}\left(\frac{\lambda}{\sqrt{n}}\right)\right)^n$$

for $\lambda \in \mathbb{R}$, because the random variables are independent and identically distributed.

Now $\phi = \phi_{X_1}$ satisfies

$$\phi(\mu) = \phi(0) + \phi'(0)\mu + \frac{1}{2}\phi''(0)\mu^2 + o(\mu^2) \quad \text{as } \mu \to 0,$$

with $\phi(0) = 1$, $\phi'(0) = iE(X_1) = 0$, $\phi''(0) = -E(X_1^2) = -1$. Consequently our setting $\mu = \frac{\lambda}{\sqrt{n}}$ gives

$$\phi_{X_1}\left(\frac{\lambda}{\sqrt{n}}\right) = 1 - \frac{\lambda^2}{2n} + o\left(\frac{\lambda^2}{n}\right),$$

and so

$$\phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = \left(1 - \frac{\lambda^2}{2n} + o\left(\frac{\lambda^2}{n}\right)\right)^n \to e^{-\frac{\lambda^2}{2}}$$

for all λ, as $n \to \infty$. But $e^{-\frac{\lambda^2}{2}}$ is the characteristic function of an $N(0,1)$ random variable. It turns out that this convergence of the characteristic functions implies the limit (1): see Breiman [**B**] for more. □

H. CONDITIONAL EXPECTATION

MOTIVATION. We earlier decided to define $P(A \mid B)$, the probability of A, given B, to be $\frac{P(A \cap B)}{P(B)}$, provided $P(B) > 0$. How then should we define

$$E(X \mid B),$$

the expected value of the random variable X, given the event B? Remember that we can think of B as the new probability space, with $\tilde{P} = \frac{P}{P(B)}$. Thus if $P(B) > 0$, we should set

$$\begin{aligned} E(X \mid B) &= \text{ mean value of } X \text{ over } B \\ &= \frac{1}{P(B)} \int_B X \, dP. \end{aligned}$$

Next we pose a more interesting question. What is a reasonable definition of

$$E(X \mid Y),$$

the expected value of the random variable X, given another random variable Y? In other words if "chance" selects a sample point $\omega \in \Omega$ and all we know about ω is the value $Y(\omega)$, what is our best guess as to the value $X(\omega)$?

This turns out to be a subtle, but extremely important issue, for which we provide two introductory discussions.

FIRST APPROACH TO CONDITIONAL EXPECTATION. We start with an example.

EXAMPLE. Assume we are given a probability space $(\Omega, \mathcal{U}, P)$, on which is defined a simple random variable Y. That is, $Y = \sum_{i=1}^{m} a_i \chi_{A_i}$, and so

$$Y = \begin{cases} a_1 & \text{on } A_1 \\ a_2 & \text{on } A_2 \\ \quad \vdots & \\ a_m & \text{on } A_m, \end{cases}$$

for distinct real numbers $a_1, a_2, \ldots, a_m$ and disjoint events $A_1, A_2, \ldots, A_m$, each of positive probability, whose union is Ω.

Next, let X be any other real–valued random variable on Ω. What is our best guess of X, given Y? Think about the problem this way: if we know the value of $Y(\omega)$, we can tell which event $A_1, A_2, \ldots, A_m$ contains ω. This, and only this, known, our best estimate for X should then be the average value of X over each appropriate event. That is, we should take

$$E(X \,|\, Y) := \begin{cases} \frac{1}{P(A_1)} \int_{A_1} X \, dP & \text{on } A_1 \\ \frac{1}{P(A_2)} \int_{A_2} X \, dP & \text{on } A_2 \\ \quad \vdots & \\ \frac{1}{P(A_m)} \int_{A_m} X \, dP & \text{on } A_m. \end{cases}$$

□

We note for this example that
- $E(X \,|\, Y)$ is a random variable, and *not* a constant.
- $E(X \,|\, Y)$ is $\mathcal{U}(Y)$-measurable.
- $\int_A X dP = \int_A E(X \,|\, Y) \, dP$ for all $A \in \mathcal{U}(Y)$.

Let us take these properties as the definition in the general case:

DEFINITION. Let Y be a random variable. Then $E(X \,|\, Y)$ is any $\mathcal{U}(Y)$-measurable random variable such that

$$\int_A X \, dP = \int_A E(X \,|\, Y) \, dP \quad \text{for all } A \in \mathcal{U}(Y).$$

Finally, notice that it is not really the values of Y that are important, but rather just the σ-algebra it generates. This motivates the next

DEFINITION. Let $(\Omega, \mathcal{U}, P)$ be a probability space and suppose $\mathcal{V}$ is a σ-algebra, $\mathcal{V} \subseteq \mathcal{U}$. If $X : \Omega \to \mathbb{R}^n$ is an integrable random variable, we define

$$E(X \,|\, \mathcal{V})$$

to be any random variable on Ω such that

(i) $E(X \,|\, \mathcal{V})$ is $\mathcal{V}$-measurable, and

(ii) $\int_A X \, dP = \int_A E(X \,|\, \mathcal{V}) \, dP$ for all $A \in \mathcal{V}$.

Interpretation. We can understand $E(X \,|\, \mathcal{V})$ as follows. We are given the "information" available in a σ-algebra $\mathcal{V}$, from which we intend to build an estimate of the random variable X. Condition (i) in the definition requires that $E(X \,|\, \mathcal{V})$ be constructed from the information in $\mathcal{V}$, and (ii) requires that our estimate be consistent with X, at least as regards integration over events in $\mathcal{V}$. We will later see that the conditional expectation $E(X \,|\, \mathcal{V})$, so defined, has various additional nice properties.

Remark. We can check without difficulty that
(i) $E(X \,|\, Y) = E(X \,|\, \mathcal{U}(Y))$.
(ii) $E(E(X \,|\, \mathcal{V})) = E(X)$.
(iii) $E(X) = E(X \,|\, \mathcal{W})$, where $\mathcal{W} = \{\emptyset, \Omega\}$ is the trivial σ-algebra. □

THEOREM. *Let X be an integrable random variable. Then for each σ-algebra $\mathcal{V} \subset \mathcal{U}$, the conditional expectation $E(X \,|\, \mathcal{V})$ exists and is unique up to $\mathcal{V}$-measurable sets of probability zero.*

We omit the proof, which uses a few advanced concepts from measure theory.

SECOND APPROACH TO CONDITIONAL EXPECTATION. An elegant alternative approach to conditional expectations is based upon projections onto closed subspaces, and is motivated by this example:

Least squares method. Consider for the moment $\mathbb{R}^n$ and suppose that V is a proper subspace.

Suppose we are given a vector $x \in \mathbb{R}^n$. The *least squares problem* asks us to find a vector $z \in V$ so that

$$|z - x| = \min_{y \in V} |y - x|.$$

It is not particularly difficult to show that, given x, there exists a unique vector $z \in V$ solving this minimization problem. We call v the *projection* of x onto V,

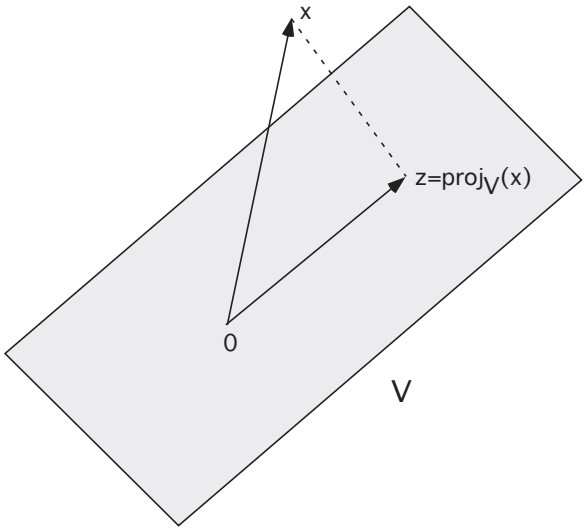
$$z = \mathrm{proj}_V(x). \tag{7}$$

Now we want to find formula characterizing z. For this take any other vector $w \in V$. Define then

$$i(\tau) := |z + \tau w - x|^2.$$

Since $z + \tau w \in V$ for all τ, we see that the function $i(\cdot)$ has a minimum at $\tau = 0$. Hence $0 = i'(0) = 2(z - x) \cdot w$; that is,

$$x \cdot w = z \cdot w \quad \text{for all } w \in V. \tag{8}$$



The geometric interpretation is that the "error" $x - z$ is perpendicular to the subspace V. □

Projection of random variables. Motivated by the example above, we return now to conditional expectation. Let us take the linear space $L^2(\Omega) = L^2(\Omega, \mathcal{U})$, which consists of all real-valued, $\mathcal{U}$–measurable random variables Y, such that

$$||Y|| := \left(\int_\Omega Y^2 \, dP \right)^{\frac{1}{2}} < \infty.$$

We call $||Y||$ the *norm* of Y; and if $X, Y \in L^2(\Omega)$, we define their *inner product* to be

$$(X, Y) := \int_\Omega XY \, dP = E(XY).$$

Next, take as before $\mathcal{V}$ to be a σ-algebra contained in $\mathcal{U}$. Consider then

$$V := L^2(\Omega, \mathcal{V}),$$

the space of square–integrable random variables *that are* $\mathcal{V}$*–measurable*. This is a closed subspace of $L^2(\Omega)$. Consequently if $X \in L^2(\Omega)$, we can define its *projection*

$$Z = \text{proj}_V(X), \tag{9}$$

by analogy with (7) in the finite dimensional case. Almost exactly as we established (8) above, we can likewise show

$$(X, W) = (Z, W) \quad \text{for all } W \in V.$$

Take in particular $W = \chi_A$ for any set $A \in \mathcal{V}$. In view of the definition of the inner product, it follows that

$$\int_A X \, dP = \int_A Z \, dP \quad \text{for all } A \in \mathcal{V}.$$

Since $Z \in V$ is $\mathcal{V}$-measurable, we see that Z is in fact $E(X \mid \mathcal{V})$, as defined in the earlier discussion. That is,

$$E(X \mid \mathcal{V}) = \text{proj}_V(X).$$

We could therefore alternatively take the last identity as a definition of conditional expectation. This point of view also makes it clear that $Z = E(X \mid \mathcal{V})$ solves the least squares problem:

$$||Z - X|| = \min_{Y \in V} ||Y - X||;$$

and so $E(X \mid \mathcal{V})$ *can be interpreted as that* $\mathcal{V}$*-measurable random variable which is the best least squares approximation of the random variable* X*.* □

The two introductory discussions now completed, we turn next to examining conditional expectation more closely.

THEOREM (Properties of conditional expectation).

(i) *If* X *is* $\mathcal{V}$*-measurable, then* $E(X \mid \mathcal{V}) = X$ *a.s.*

(ii) *If* a, b *are constants,* $E(aX + bY \mid \mathcal{V}) = aE(X \mid \mathcal{V}) + bE(Y \mid \mathcal{V})$ *a.s.*

(iii) *If* X *is* $\mathcal{V}$*-measurable and* XY *is integrable, then* $E(XY \mid \mathcal{V}) = XE(Y \mid \mathcal{V})$ *a.s.*

(iv) *If* X *is independent of* $\mathcal{V}$*, then* $E(X \mid \mathcal{V}) = E(X)$ *a.s.*

(v) *If* $\mathcal{W} \subseteq \mathcal{V}$*, we have*

$$E(X \mid \mathcal{W}) = E(E(X \mid \mathcal{V}) \mid \mathcal{W}) = E(E(X \mid \mathcal{W}) \mid \mathcal{V}) \text{ a.s.}$$

(vi) *The inequality* $X \leq Y$ *a.s. implies* $E(X \mid \mathcal{V}) \leq E(Y \mid \mathcal{V})$ *a.s.*

Proof.

1. Statement (i) is obvious, and (ii) is easy to check
2. By uniqueness a.s. of $E(XY \mid \mathcal{V})$, it is enough in proving (iii) to show

$$\int_A XE(Y \mid \mathcal{V})\, dP = \int_A XY\, dP \quad \text{for all } A \in \mathcal{V}. \tag{10}$$

First suppose $X = \sum_{i=1}^m b_i \chi_{B_i}$, where $B_i \in \mathcal{V}$ for $i = 1, \ldots, m$. Then

$$\begin{aligned}\int_A XE(Y \mid \mathcal{V})\, dP &= \sum_{i=1}^m b_i \underbrace{\int_{A \cap B_i} E(Y \mid \mathcal{V})\, dP}_{\in \mathcal{V}} \\ &= \sum_{i=1}^m b_i \int_{A \cap B_i} Y\, dP = \int_A XY\, dP.\end{aligned}$$

This proves (10) if X is a simple function. The general case follows by approximation.

3. To show (iv), it suffices to prove $\int_A E(X)\,dP = \int_A X\,dP$ for all $A \in \mathcal{V}$. Let us compute:

$$\int_A X\,dP = \int_\Omega \chi_A X\,dP = E(\chi_A X) = E(X)P(A) = \int_A E(X)\,dP,$$

the third equality owing to independence.

4. Assume $\mathcal{W} \subseteq \mathcal{V}$ and let $A \in \mathcal{W}$. Then

$$\int_A E(E(X \mid \mathcal{V}) \mid \mathcal{W})\,dP = \int_A E(X \mid \mathcal{V})\,dP = \int_A X\,dP,$$

since $A \in \mathcal{W} \subseteq \mathcal{V}$. Thus $E(X \mid \mathcal{W}) = E(E(X \mid \mathcal{V}) \mid \mathcal{W})$ a.s.

Furthermore, assertion (i) implies that $E(E(X \mid \mathcal{W}) \mid \mathcal{V}) = E(X \mid \mathcal{W})$, since $E(X \mid \mathcal{W})$ is $\mathcal{W}$-measurable and so also $\mathcal{V}$-measurable. This establishes assertion (v).

5. Finally, suppose $X \leq Y$, and note that

$$\begin{aligned}\int_A E(Y \mid \mathcal{V}) - E(X \mid \mathcal{V})\,dP &= \int_A E(Y - X \mid \mathcal{V})\,dP \\ &= \int_A Y - X\,dP \geq 0\end{aligned}$$

for all $A \in \mathcal{V}$. Take $A := \{E(Y \mid \mathcal{V}) - E(X \mid \mathcal{V}) \leq 0\}$. This event lies in $\mathcal{V}$, and we deduce from the previous inequality that $P(A) = 0$. □

LEMMA (Conditional Jensen's Inequality). *Suppose $\Phi : \mathbb{R} \to \mathbb{R}$ is convex, with $E(|\Phi(X)|) < \infty$. Then*

$$\Phi(E(X \mid \mathcal{V})) \leq E(\Phi(X) \mid \mathcal{V}).$$

We leave the proof as an exercise.

I. MARTINGALES

MOTIVATION. Suppose $Y_1, Y_2, \ldots$ are independent real-valued random variables, with

$$E(Y_i) = 0 \quad (i = 1, 2, \ldots).$$

Define the sum $S_n := Y_1 + \cdots + Y_n$.

What is our best guess of S_{n+k}, given the values of $S_1, \ldots, S_n$? The answer is

$$\begin{aligned}E(S_{n+k} \mid S_1, \ldots, S_n) &= E(Y_1 + \cdots + Y_n \mid S_1, \ldots, S_n) \\ &\quad + E(Y_{n+1} + \cdots + Y_{n+k} \mid S_1, \ldots, S_n) \\ &= Y_1 + \cdots + Y_n + \underbrace{E(Y_{n+1} + \cdots + Y_{n+k})}_{=0} = S_n.\end{aligned} \tag{11}$$

Thus the best estimate of the "future value" of S_{n+k}, *given* the history up to time n, is just S_n.

If we interpret Y_i as the payoff of a "fair" gambling game at time i, and therefore S_n as the total winnings at time n, the calculation above says that at any time one's

future *expected* winnings, given the winnings to date, is just the current amount of money. So the formula (11) characterizes a "fair" game.

We incorporate these ideas into a formal definition:

DEFINITION. Let $X_1, \ldots, X_n, \ldots$ be a sequence of real-valued random variables, with $E(|X_i|) < \infty \quad (i = 1, 2, \ldots)$. If

$$X_k = E(X_j \,|\, X_1, \ldots, X_k) \quad \text{a.s. for all } j \geq k,$$

we call $\{X_i\}_{i=1}^{\infty}$ a (*discrete*) *martingale*.

DEFINITION. Let $X(\cdot)$ be a real–valued stochastic process. Then

$$\mathcal{U}(t) := \mathcal{U}(X(s) \,|\, 0 \leq s \leq t),$$

the σ-algebra generated by the random variables $X(s)$ for $0 \leq s \leq t$, is called the *history* of the process until (and including) time $t \geq 0$.

DEFINITIONS. Let $X(\cdot)$ be a stochastic process, such that $E(|X(t)|) < \infty$ for all $t \geq 0$.

(i) If

$$X(s) = E(X(t) \,|\, \mathcal{U}(s)) \text{ a.s.} \quad \text{for all } t \geq s \geq 0,$$

then $X(\cdot)$ is called a *martingale*.

(ii) If

$$X(s) \leq E(X(t) \,|\, \mathcal{U}(s)) \text{ a.s.} \quad \text{for all } t \geq s \geq 0,$$

$X(\cdot)$ is a *submartingale*. □

EXAMPLE. Let $W(\cdot)$ be a 1-dimensional Wiener process, as defined later in Chapter 3. Then

$$W(\cdot) \text{ is a martingale.}$$

To see this, write $\mathcal{W}(t) := \mathcal{U}(W(s)|\ 0 \leq s \leq t)$, and let $t \geq s$. Then

$$\begin{aligned} E(W(t) \,|\, \mathcal{W}(s)) &= E(W(t) - W(s) \,|\, \mathcal{W}(s)) + E(W(s) \,|\, \mathcal{W}(s)) \\ &= E(W(t) - W(s)) + W(s) = W(s) \quad \text{a.s.} \end{aligned}$$

(The reader should refer back to this calculation after reading Chapter 3.) □

LEMMA. *Suppose $X(\cdot)$ is a real-valued martingale and $\Phi : \mathbb{R} \to \mathbb{R}$ is convex. Then if $E(|\Phi(X(t))|) < \infty$ for all $t \geq 0$,*

$$\Phi(X(\cdot)) \textit{ is a submartingale.}$$

We omit the proof, which uses Jensen's inequality.

Martingales are important in probability theory mainly because they admit the following powerful estimates:

THEOREM (Discrete martingale inequalities).

(i) *If* $\{X_n\}_{n=1}^{\infty}$ *is a submartingale, then*

$$P\left(\max_{1\le k\le n} X_k \ge \lambda\right) \le \frac{1}{\lambda} E(X_n^+)$$

for all $n = 1, \ldots$ *and* $\lambda > 0$.

(ii) *If* $\{X_n\}_{n=1}^{\infty}$ *is a martingale and* $1 < p < \infty$, *then*

$$E\left(\max_{1\le k\le n} |X_k|^p\right) \le \left(\frac{p}{p-1}\right)^p E(|X_n|^p)$$

for all $n = 1, \ldots$.

A proof is provided in Appendix B. Notice that (i) is a generalization of the Chebyshev inequality. We can also extend these estimates to continuous–time martingales.

THEOREM (Martingale inequalities). *Let* $X(\cdot)$ *be a stochastic process with continuous sample paths a.s.*

(i) *If* $X(\cdot)$ *is a submartingale, then*

$$P\left(\max_{0\le s\le t} X(s) \ge \lambda\right) \le \frac{1}{\lambda} E(X(t)^+) \quad \text{for all } \lambda > 0,\ t \ge 0.$$

(ii) *If* $X(\cdot)$ *is a martingale and* $1 < p < \infty$, *then*

$$E\left(\max_{0\le s\le t} |X(s)|^p\right) \le \left(\frac{p}{p-1}\right)^p E(|X(t)|^p).$$

Outline of Proof. Choose $\lambda > 0$, $t > 0$ and select $0 = t_0 < t_1 < \cdots < t_n = t$. We check that $\{X(t_i)\}_{i=1}^n$ is a martingale and apply the discrete martingale inequality. Next choose a finer and finer partition of $[0, t]$ and pass to limits.

The proof of assertion (ii) is similar.

□

CHAPTER 3: BROWNIAN MOTION AND "WHITE NOISE"

A. Motivation and definitions
B. Construction of Brownian motion
C. Sample paths
D. Markov property

A. MOTIVATION AND DEFINITIONS

SOME HISTORY. *R. Brown* in 1826–27 observed the irregular motion of pollen particles suspended in water. He and others noted that

• the path of a given particle is very irregular, having a tangent at no point, and

• the motions of two distinct particles appear to be independent.

In 1900 *L. Bachelier* attempted to describe fluctuations in stock prices mathematically and essentially discovered first certain results later rederived and extended by *A. Einstein* in 1905. Einstein studied the Brownian phenomena this way. Let us consider a long, thin tube filled with clear water, into which we inject at time $t = 0$ a unit amount of ink, at the location $x = 0$. Now let $f(x,t)$ denote the density of ink particles at position $x \in \mathbb{R}$ and time $t \geq 0$. Initially we have

$$f(x,0) = \delta_0, \text{ the unit mass at } 0.$$

Next, suppose that the probability density of the event that an ink particle moves from x to $x+y$ in (small) time τ is $\rho(\tau, y)$. Then

$$\begin{aligned} f(x,t+\tau) &= \int_{-\infty}^{\infty} f(x-y,t)\rho(\tau,y)\,dy \\ &= \int_{-\infty}^{\infty} \left(f - f_x y + \frac{1}{2} f_{xx} y^2 + \dots \right) \rho(\tau,y)\,dy. \end{aligned} \tag{1}$$

But since ρ is a probability density, $\int_{-\infty}^{\infty} \rho\,dy = 1$; whereas $\rho(\tau,-y) = \rho(\tau,y)$ by symmetry. Consequently $\int_{-\infty}^{\infty} y\rho\,dy = 0$. We further assume that $\int_{-\infty}^{\infty} y^2\rho\,dy$, the variance of ρ, is *linear* in τ:

$$\int_{-\infty}^{\infty} y^2\rho\,dy = D\tau, \ D > 0.$$

We insert these identities into (1), thereby to obtain

$$\frac{f(x,t+\tau) - f(x,t)}{\tau} = \frac{D f_{xx}(x,t)}{2} \quad \{+ \text{ higher order terms}\}.$$

Sending now $\tau \to 0$, we discover

$$f_t = \frac{D}{2} f_{xx}$$

This is the *diffusion equation*, also known as the *heat equation*. This partial differential equation, with the initial condition $f(x,0) = \delta_0$, has the solution

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

This says the *probability density at time* t is $N(0, Dt)$, for some constant D.

In fact, Einstein computed:

$$D = \frac{RT}{N_A f}, \quad \text{where} \quad \begin{cases} R = \text{gas constant} \\ T = \text{absolute temperature} \\ f = \text{friction coefficient} \\ N_A = \text{Avogadro's number.} \end{cases}$$

This equation and the observed properties of Brownian motion allowed *J. Perrin* to compute N_A ($\approx 6 \times 10^{23}$ = the number of molecules in a mole) and help to confirm the atomic theory of matter.

N. Wiener in the 1920's (and later) put the theory on a firm mathematical basis. His ideas are at the heart of the mathematics in §B–D below.

RANDOM WALKS. A variant of Einstein's argument follows. We introduce a 2-dimensional rectangular lattice, comprising the sites $\{(m\Delta x, n\Delta t) \,|\, m = 0, \pm 1, \pm 2, \dots; n = 0, 1, 2, \dots\}$. Consider a particle starting at $x = 0$ and time $t = 0$, and at each time $n\Delta t$ moves to the left an amount Δx with probability $1/2$, to the right an amount Δx with probability $1/2$. Let $p(m, n)$ denote the probability that the particle is at position $m\Delta x$ at time $n\Delta t$. Then

$$p(m, 0) = \begin{cases} 0 & m \neq 0 \\ 1 & m = 0. \end{cases}$$

Also

$$p(m, n+1) = \frac{1}{2}p(m-1, n) + \frac{1}{2}p(m+1, n),$$

and hence

$$p(m, n+1) - p(m, n) = \frac{1}{2}(p(m+1, n) - 2p(m, n) + p(m-1, n)).$$

Now assume

$$\frac{(\Delta x)^2}{\Delta t} = D \quad \text{for some positive constant } D.$$

This implies

$$\frac{p(m, n+1) - p(m, n)}{\Delta t} = \frac{D}{2}\left(\frac{p(m+1, n) - 2p(m, n) + p(m-1, n)}{(\Delta x)^2}\right).$$

Let $\Delta t \to 0$, $\Delta x \to 0$, $m\Delta x \to x$, $n\Delta t \to t$, with $\frac{(\Delta x)^2}{\Delta t} \equiv D$. Then presumably $p(m, n) \to f(x, t)$, which we now interpret as the probability density that particle is at x at time t. The above difference equation becomes formally in the limit

$$f_t = \frac{D}{2} f_{xx},$$

and so we arrive at the diffusion equation again.

MATHEMATICAL JUSTIFICATION. A more careful study of this technique of passing to limits with random walks on a lattice depends upon the Laplace–De Moivre Theorem.

As above we assume the particle moves to the left or right a distance Δx with probability $1/2$. Let $X(t)$ denote the position of particle at time $t = n\Delta t \quad (n = 0, \dots)$. Define

$$S_n := \sum_{i=1}^{n} X_i,$$

where the X_i are independent random variables such that

$$\begin{cases} P(X_i = 0) = 1/2 \\ P(X_i = 1) = 1/2 \end{cases}$$

for $i = 1, \dots$. Then $V(X_i) = \frac{1}{4}$.

Now S_n is the number of moves to the right by time $t = n\Delta t$. Consequently

$$X(t) = S_n \Delta x + (n - S_n)(-\Delta x) = (2S_n - n)\Delta x.$$

Note also

$$\begin{aligned} V(X(t)) &= (\Delta x)^2 V(2S_n - n) \\ &= (\Delta x)^2 4V(S_n) = (\Delta x)^2 4nV(X_1) \\ &= (\Delta x)^2 n = \frac{(\Delta x)^2}{\Delta t} t. \end{aligned}$$

Again assume $\frac{(\Delta x)^2}{\Delta t} = D$. Then

$$X(t) = (2S_n - n)\Delta x = \left(\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \right) \sqrt{n} \Delta x = \left(\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \right) \sqrt{tD}.$$

The Laplace–De Moivre Theorem thus implies

$$\begin{aligned} \lim_{\substack{n \to \infty \\ t = n\Delta t, \frac{(\Delta x)^2}{\Delta t} = D}} P(a \le X(t) \le b) &= \lim_{n \to \infty} \left(\frac{a}{\sqrt{tD}} \le \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \le \frac{b}{\sqrt{tD}} \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{tD}}}^{\frac{b}{\sqrt{tD}}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi Dt}} \int_a^b e^{-\frac{x^2}{2Dt}} dx. \end{aligned}$$

Once again, and rigorously this time, we obtain the $N(0, Dt)$ distribution. □

Inspired by all these considerations, we now introduce Brownian motion, for which we take $D = 1$:

DEFINITION. A real-valued stochastic process $W(\cdot)$ is called a *Brownian motion* or *Wiener process* if

(i) $W(0) = 0$ a.s.,

(ii) $W(t) - W(s)$ is $N(0, t-s)$ for all $t \geq s \geq 0$,

(iii) for all times $0 < t_1 < t_2 < \cdots < t_n$, the random variables $W(t_1), W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$ are independent ("independent increments").

Notice in particular that

$$E(W(t)) = 0,\ E(W^2(t)) = t \quad \text{for each time } t \geq 0.$$

The Central Limit Theorem provides some further motivation for our definition of Brownian motion, since we can expect that any suitably scaled sum of independent, random disturbances affecting the position of a moving particle will result in a Gaussian distribution.

B. CONSTRUCTION OF BROWNIAN MOTION

COMPUTATION OF JOINT PROBABILITIES. From the definition we know that if $W(\cdot)$ is a Brownian motion, then for all $t > 0$ and $a \leq b$,

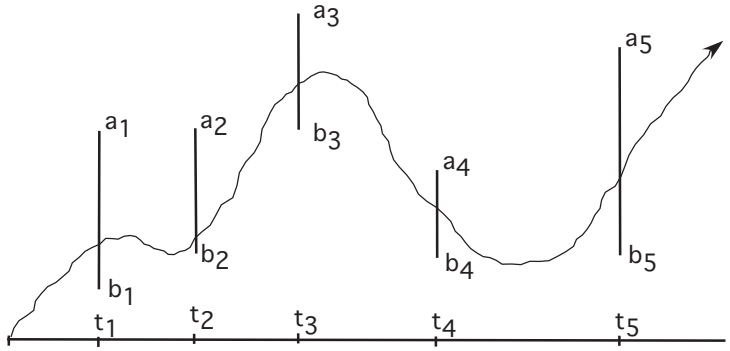
$$P(a \leq W(t) \leq b) = \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-\frac{x^2}{2t}}\, dx,$$

since $W(t)$ is $N(0, t)$.

Suppose we now choose times $0 < t_1 < \cdots < t_n$ and real numbers $a_i \leq b_i$, for $i = 1, \ldots, n$. What is the joint probability

$$P(a_1 \leq W(t_1) \leq b_1, \cdots, a_n \leq W(t_n) \leq b_n)?$$

In other words, what is the probability that a sample path of Brownian motion takes values between a_i and b_i at time t_i for each $i = 1, \ldots n$?



We can *guess* the answer as follows. We know

$$P(a_1 \leq W(t_1) \leq b_1) = \int_{a_1}^{b_1} \frac{e^{-\frac{x_1^2}{2t_1}}}{\sqrt{2\pi t_1}}\, dx_1;$$

and *given* that $W(t_1) = x_1$, $a_1 \le x_1 \le b_1$, then presumably the process is $N(x_1, t_2 - t_1)$ on the interval $[t_1, t_2]$. Thus the probability that $a_2 \le W(t_2) \le b_1$, given that $W(t_1) = x_1$, should equal

$$\int_{a_2}^{b_2} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{|x_2 - x_1|^2}{2(t_2 - t_1)}} dx_2.$$

Hence it should be that

$$P(a_1 \le W(t_1) \le b_1, a_2 \le W(t_2) \le b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} g(x_1, t_1 \,|\, 0) g(x_2, t_2 - t_1 \,|\, x_1)\, dx_2 dx_1$$

for

$$g(x, t \,|\, y) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}.$$

In general, we would therefore guess that

(2)
$$P(a_1 \le W(t_1) \le b_1, \dots, a_n \le W(t_n) \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} g(x_1, t_1 \,|\, 0) g(x_2, t_2 - t_1 \,|\, x_1) \dots g(x_n, t_n - t_{n-1} \,|\, x_{n-1})\, dx_n \dots dx_1.$$

The next assertion confirms and extends this formula.

THEOREM. *Let $W(\cdot)$ be a one-dimensional Wiener process. Then for all positive integers n, all choices of times $0 = t_0 < t_1 < \cdots < t_n$ and each function $f : \mathbb{R}^n \to \mathbb{R}$, we have*

$$Ef(W(t_1), \dots, W(t_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) g(x_1, t_1 \,|\, 0) g(x_2, t_2 - t_1 \,|\, x_1) \dots g(x_n, t_n - t_{n-1} \,|\, x_{n-1})\, dx_n \dots dx_1.$$

Our taking

$$f(x_1, \dots, x_n) = \chi_{[a_1, b_1]}(x_1) \cdots \chi_{[a_n, b_n]}(x_n)$$

gives (2).

PROOF. Let us write $X_i := W(t_i), Y_i := X_i - X_{i-1}$ for $i = 1, \dots, n$. We also define

$$h(y_1, y_2, \dots, y_n) := f(y_1, y_1 + y_2, \dots, y_1 + \cdots + y_n).$$

Then

$$\begin{aligned}
Ef(W(t_1),\dots,W(t_n)) &= Eh(Y_1,\dots,Y_n)\\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(y_1,\dots,y_n)g(y_1,t_1\,|\,0)g(y_2,t_2-t_1\,|\,0)\\
&\qquad\qquad \dots g(y_n,t_n-t_{n-1}\,|\,0)dy_n\dots dy_1\\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\dots,x_n)g(x_1,t_1\,|\,0)g(x_2,t_2-t_1\,|\,x_1)\\
&\qquad\qquad \dots g(x_n,t_n-t_{n-1}\,|\,x_{n-1})\,dx_n\dots dx_1.
\end{aligned}$$

For the second equality we recalled that the random variables $Y_i = W(t_i) - W(t_{i-1})$ are independent for $i = 1,\dots,n$, and that each Y_i is $N(0, t_i - t_{i-1})$. We also changed variables using the identities $y_i = x_i - x_{i-1}$ for $i = 1,\dots,n$ and $x_0 = 0$. The Jacobian for this change of variables equals 1. □

BUILDING A ONE-DIMENSIONAL WIENER PROCESS. The main issue now is to demonstrate that a Brownian motion actually exists.

Our method will be to develop a formal expansion of white noise $\xi(\cdot)$ in terms of a cleverly selected orthonormal basis of $L^2(0,1)$, the space of all real-valued, square–integrable funtions defined on $(0,1)$. We will then integrate the resulting expression in time, show that this series converges, and prove then that we have built a Wiener process. This procedure is a form of "wavelet analysis": see Pinsky [**P**].

We start with an easy lemma.

LEMMA. *Suppose $W(\cdot)$ is a one-dimensional Brownian motion. Then*

$$E(W(t)W(s)) = t \wedge s = \min\{s,t\} \quad \textit{for } t \geq 0,\ s \geq 0.$$

PROOF. Assume $t \geq s \geq 0$. Then

$$\begin{aligned}
E(W(t)W(s)) &= E((W(s) + W(t) - W(s))W(s))\\
&= E(W^2(s)) + E((W(t) - W(s))W(s))\\
&= s + \underbrace{E(W(t) - W(s))}_{=0}\underbrace{E(W(s))}_{=0}\\
&= s = t \wedge s,
\end{aligned}$$

since $W(s)$ is $N(0,s)$ and $W(t) - W(s)$ is independent of $W(s)$. □

HEURISTICS. Remember from Chapter 1 that the formal time-derivative

$$\dot{W}(t) = \frac{dW(t)}{dt} = \xi(t)$$

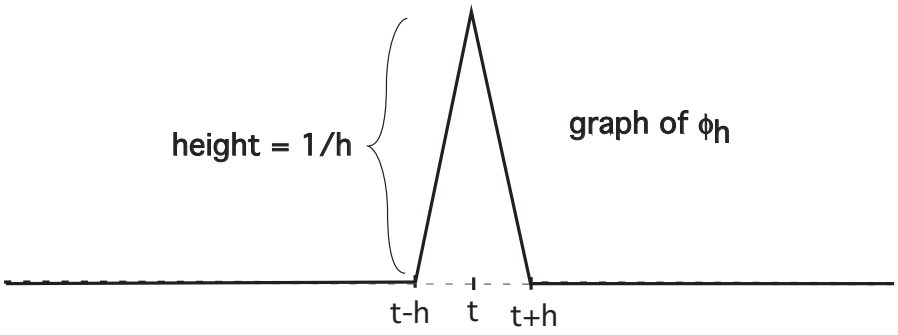
is "1-dimensional white noise". As we will see later however, for a.e. ω the sample path $t \mapsto W(t,\omega)$ is in fact differentiable for no time $t \geq 0$. Thus $\dot{W}(t) = \xi(t)$ does not really exist.

However, we do have the heuristic formula

$$\text{“}E(\xi(t)\xi(s)) = \delta_0(s-t)\text{”}, \tag{3}$$

where δ_0 is the unit mass at 0. A formal “proof” is this. Suppose $h > 0$, fix $t > 0$, and set

$$\begin{aligned}\phi_h(s) &:= E\left(\left(\frac{W(t+h)-W(t)}{h}\right)\left(\frac{W(s+h)-W(s)}{h}\right)\right)\\ &= \frac{1}{h^2}[E(W(t+h)W(s+h)) - E(W(t+h)W(s)) - E(W(t)W(s+h)) + E(W(t)W(s))]\\ &= \frac{1}{h^2}[((t+h)\wedge(s+h)) - ((t+h)\wedge s) - (t\wedge(s+h)) + (t\wedge s)].\end{aligned}$$



Then $\phi_h(s) \to 0$ as $h \to 0$, $t \neq s$. But $\int \phi_h(s)\,ds = 1$, and so presumably $\phi_h(s) \to \delta_0(s-t)$ in some sense, as $h \to 0$. In addition, we expect that $\phi_h(s) \to E(\xi(t)\xi(s))$. This gives the formula (3) above. □

Remark: Why $\dot{\mathbf{W}}(\,\cdot\,) = \boldsymbol{\xi}(\,\cdot\,)$ is called white noise. If $X(\cdot)$ is any real-valued stochastic process with $E(X^2(t)) < \infty$ for all $t \geq 0$, we define

$$r(t,s) := E(X(t)X(s)) \quad (t,s \geq 0),$$

the *autocorrelation function* of $X(\cdot)$. If $r(t,s) = c(t-s)$ for some function $c : \mathbb{R} \to \mathbb{R}$ and if $E(X(t)) = E(X(s))$ for all $t, s \geq 0$, $X(\cdot)$ is called *stationary in the wide sense*. A white noise process $\xi(\cdot)$ is by definition Gaussian, wide sense stationary, with $c(\cdot) = \delta_0$.

In general we define

$$f(\lambda) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda t}c(t)\,dt \quad (\lambda \in \mathbb{R})$$

to be the *spectral density* of a process $X(\cdot)$. For white noise, we have

$$f(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda t}\delta_0\,dt = \frac{1}{2\pi} \quad \text{for all } \lambda.$$

Thus the spectral density of $\xi(\cdot)$ is flat; that is, all "frequencies" contribute equally in the correlation function, just as—by analogy—all colors contribute equally to make white light. □

RANDOM FOURIER SERIES. Suppose now $\{\psi_n\}_{n=0}^{\infty}$ is a complete, orthonormal basis of $L^2(0,1)$, where $\psi_n = \psi_n(t)$ are functions of $0 \le t \le 1$ only and so are not random variables. The orthonormality means that

$$\int_0^1 \psi_n(s)\psi_m(s)\,ds = \delta_{mn} \quad \text{for all } m, n.$$

We write formally

$$\xi(t) = \sum_{n=0}^{\infty} A_n \psi_n(t) \quad (0 \le t \le 1). \tag{4}$$

It is easy to see that then

$$A_n = \int_0^1 \xi(t)\psi_n(t)\,dt.$$

We expect that the A_n are independent and Gaussian, with $E(A_n) = 0$. Therefore to be consistent we must have for $m \neq n$

$$\begin{aligned} 0 = E(A_n)E(A_m) = E(A_nA_m) &= \int_0^1 \int_0^1 E(\xi(t)\xi(s))\psi_n(t)\psi_m(s)\,dtds \\ &= \int_0^1 \int_0^1 \delta_0(s-t)\psi_n(t)\psi_m(s)\,dtds \quad \text{by (3)} \\ &= \int_0^1 \psi_n(s)\psi_m(s)\,ds. \end{aligned}$$

But this is already automatically true as the ψ_n are orthogonal. Similarly,

$$E(A_n^2) = \int_0^1 \psi_n^2(s)\,ds = 1.$$

Consequently if the A_n are independent and $N(0,1)$, it is reasonable to believe that formula (4) makes sense. But then the Brownian motion $W(\cdot)$ should be given by

$$W(t) := \int_0^t \xi(s)\,ds = \sum_{n=0}^{\infty} A_n \int_0^t \psi_n(s)\,ds. \tag{5}$$

This seems to be true for any orthonormal basis, and we will next make this rigorous by choosing a particularly nice basis.

LÉVY–CIESIELSKI CONSTRUCTION OF BROWNIAN MOTION

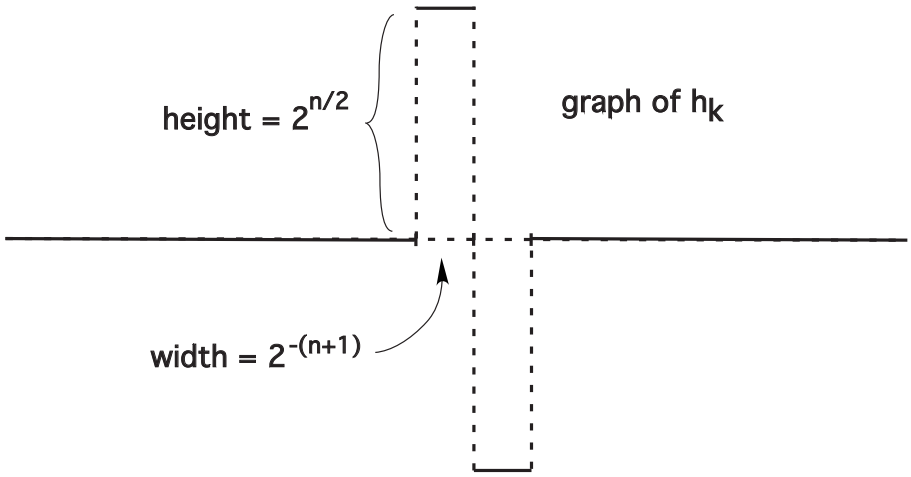
DEFINITION. The family $\{h_k(\cdot)\}_{k=0}^{\infty}$ of *Haar functions* are defined for $0 \le t \le 1$ as follows:

$$h_0(t) := 1 \quad \text{for } 0 \le t \le 1.$$

$$h_1(t) := \begin{cases} 1 & \text{for } 0 \le t \le \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} < t \le 1. \end{cases}$$

If $2^n \le k < 2^{n+1}$, $n = 1, 2, \ldots$, we set

$$h_k(t) := \begin{cases} 2^{n/2} & \text{for } \frac{k-2^n}{2^n} \le t \le \frac{k-2^n+1/2}{2^n} \\ -2^{n/2} & \text{for } \frac{k-2^n+1/2}{2^n} < t \le \frac{k-2^n+1}{2^n} \\ 0 & \text{otherwise.} \end{cases}$$



Graph of a Haar function

LEMMA 1. *The functions $\{h_k(\cdot)\}_{k=0}^{\infty}$ form a complete, orthonormal basis of $L^2(0,1)$.*

Proof. 1. We have $\int_0^1 h_k^2\, dt = 2^n \left(\frac{1}{2^{n+1}} + \frac{1}{2^{n+1}}\right) = 1$.

Note also that for all $l > k$, either $h_k h_l = 0$ for all t or else h_k is constant on the support of h_l. In this second case

$$\int_0^1 h_l h_k\, dt = \pm 2^{n/2} \int_0^1 h_l\, dt = 0.$$

2. Suppose $f \in L^2(0,1)$, $\int_0^1 f h_k\, dt = 0$ for all $k = 0, 1, \ldots$. We will prove $f = 0$ almost everywhere.

If $n = 0$, we have $\int_0^1 f\, dt = 0$. Let $n = 1$. Then $\int_0^{1/2} f\, dt = \int_{1/2}^1 f\, dt$; and both are equal to zero, since $0 = \int_0^{1/2} f\, dt + \int_{1/2}^1 f\, dt = \int_0^1 f\, dt$. Continuing in this way, we deduce $\int_{\frac{k}{2^{n+1}}}^{\frac{k+1}{2^{n+1}}} f\, dt = 0$ for all $0 \le k < 2^{n+1}$. Thus $\int_s^r f\, dt = 0$ for all dyadic

rationals $0 \le s \le r \le 1$, and so for all $0 \le s \le r \le 1$. But

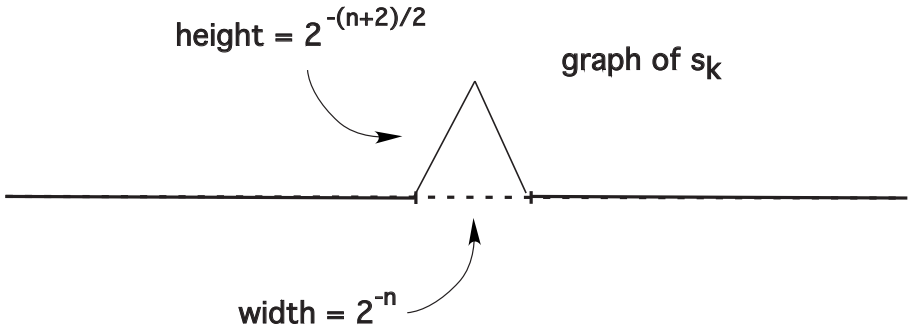
$$f(r) = \frac{d}{dr} \int_0^r f(t)\, dt = 0 \quad \text{a.e. } r.$$

□

DEFINITION. For $k = 0, 1, 2, \ldots,$

$$s_k(t) := \int_0^t h_k(s)\, ds \quad (0 \le t \le 1)$$

is the k^{th}–*Schauder function.*



GRAPH OF A SCHAUDER FUNCTION

The graph of s_k is a "tent" of height $2^{-n/2-1}$, lying above the interval $[\frac{k-2^n}{2^n}, \frac{k-2^n+1}{2^n}]$. Consequently if $2^n \le k < 2^{n+1}$, then

$$\max_{0 \le t \le 1} |s_k(t)| = 2^{-n/2-1}.$$

Our goal is to define

$$W(t) := \sum_{k=0}^{\infty} A_k s_k(t)$$

for times $0 \le t \le 1$, where the coefficients $\{A_k\}_{k=0}^{\infty}$ are independent, $N(0,1)$ random variables defined on some probability space.

We must first of all check whether this series converges.

LEMMA 2. *Let $\{a_k\}_{k=0}^{\infty}$ be a sequence of real numbers such that*

$$|a_k| = O(k^{\delta}) \quad \text{as } k \to \infty$$

for some $0 \le \delta < 1/2$. Then the series

$$\sum_{k=0}^{\infty} a_k s_k(t)$$

converges uniformly for $0 \le t \le 1$.

PROOF. Fix $\varepsilon > 0$. Notice that for $2^n \le k < 2^{n+1}$, the functions $s_k(\cdot)$ have disjoint supports. Set

$$b_n := \max_{2^n \le k < 2^{n+1}} |a_k| \le C(2^{n+1})^\delta.$$

Then for $0 \le t \le 1$,

$$\begin{aligned}\sum_{k=2^m}^{\infty} |a_k||s_k(t)| &\le \sum_{n=m}^{\infty} b_n \max_{\substack{2^n \le k < 2^{n+1} \\ 0 \le t \le 1}} |s_k(t)| \\ &\le C \sum_{n=m}^{\infty} (2^{n+1})^\delta 2^{-n/2-1} < \varepsilon\end{aligned}$$

for m large enough, since $0 \le \delta < 1/2$. □

LEMMA 3. *Suppose $\{A_k\}_{k=1}^{\infty}$ are independent, $N(0,1)$ random variables. Then for almost every ω,*

$$|A_k(\omega)| = O(\sqrt{\log k}) \quad \text{as } k \to \infty.$$

In particular, the numbers $\{A_k(\omega)\}_{k=1}^{\infty}$ almost surely satisfy the hypothesis of Lemma 2 above.

PROOF. For all $x > 0$, $k = 2, \ldots$, we have

$$\begin{aligned}P(|A_k| > x) &= \frac{2}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{s^2}{2}}\, ds \\ &\le \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{4}} \int_x^{\infty} e^{-\frac{s^2}{4}}\, ds \\ &\le C e^{-\frac{x^2}{4}},\end{aligned}$$

for some constant C. Set $x := 4\sqrt{\log k}$; then

$$P(|A_k| \ge 4\sqrt{\log k}) \le C e^{-4 \log k} = C \frac{1}{k^4}.$$

Since $\sum \frac{1}{k^4} < \infty$, the Borel–Cantelli Lemma implies

$$P(|A_k| \ge 4\sqrt{\log k} \text{ i.o.}) = 0.$$

Therefore for almost every sample point ω, we have

$$|A_k(\omega)| \le 4\sqrt{\log k} \quad \text{provided } k \ge K,$$

where K depends on ω. □

LEMMA 4. $\sum_{k=0}^{\infty} s_k(s) s_k(t) = t \wedge s$ *for each* $0 \le s,\, t \le 1$.

PROOF. Define for $0 \le s \le 1$,

$$\phi_s(\tau) := \begin{cases} 1 & 0 \le \tau \le s \\ 0 & s < \tau \le 1. \end{cases}$$

Then if $s \le t$, Lemma 1 implies

$$s = \int_0^1 \phi_t \phi_s \, d\tau = \sum_{k=0}^{\infty} a_k b_k,$$

where

$$a_k = \int_0^1 \phi_t h_k \, d\tau = \int_0^t h_k \, d\tau = s_k(t), \quad b_k = \int_0^1 \phi_s h_k \, d\tau = s_k(s).$$

□

THEOREM. *Let $\{A_k\}_{k=0}^{\infty}$ be a sequence of independent, $N(0,1)$ random variables defined on the same probability space. Then the sum*

$$W(t,\omega) := \sum_{k=0}^{\infty} A_k(\omega) s_k(t) \quad (\ 0 \le t \le 1)$$

converges uniformly in t, for a.e. ω. Furthermore

(i) *$W(\cdot)$ is a Brownian motion for $0 \le t \le 1$, and*

(ii) *for a.e. ω, the sample path $t \mapsto W(t,\omega)$ is continuous.*

PROOF. 1. The uniform convergence is a consequence of Lemmas 2 and 3; this implies (ii).

2. To prove $W(\cdot)$ is a Brownian motion, we first note that clearly $W(0) = 0$ a.s.

We assert as well that $W(t) - W(s)$ is $N(0, t-s)$ for all $0 \le s \le t \le 1$. To prove this, let us compute

$$\begin{aligned} E(e^{i\lambda(W(t)-W(s))}) &= E(e^{i\lambda \sum_{k=0}^{\infty} A_k(s_k(t)-s_k(s))}) \\ &= \prod_{k=0}^{\infty} E(e^{i\lambda A_k(s_k(t)-s_k(s))}) \quad \text{by independence} \\ &= \prod_{k=0}^{\infty} e^{-\frac{\lambda^2}{2}(s_k(t)-s_k(s))^2} \quad \text{since } A_k \text{ is } N(0,1) \\ &= e^{-\frac{\lambda^2}{2}\sum_{k=0}^{\infty}(s_k(t)-s_k(s))^2} \\ &= e^{-\frac{\lambda^2}{2}\sum_{k=0}^{\infty} s_k^2(t) - 2s_k(t)s_k(s) + s_k^2(s)} \\ &= e^{-\frac{\lambda^2}{2}(t-2s+s)} \quad \text{by Lemma 4} \\ &= e^{-\frac{\lambda^2}{2}(t-s)}. \end{aligned}$$

By uniqueness of characteristic functions, the increment $W(t)-W(s)$ is $N(0,t-s)$, as asserted.

3. Next we claim for all $m = 1, 2, \ldots$ and for all $0 = t_0 < t_1 < \cdots < t_m \leq 1$, that

$$E(e^{i\sum_{j=1}^m \lambda_j (W(t_j)-W(t_{j-1}))}) = \prod_{j=1}^m e^{-\frac{\lambda_j^2}{2}(t_j-t_{j-1})}. \tag{6}$$

Once this is proved, we will know from uniqueness of characteristic functions that

$$F_{W(t_1),\ldots,W(t_m)-W(t_{m-1})}(x_1,\ldots,x_m) = F_{W(t_1)}(x_1)\cdots F_{W(t_m)-W(t_{m-1})}(x_m)$$

for all $x_1, \ldots x_m \in \mathbb{R}$. This proves that

$$W(t_1), \ldots, W(t_m) - W(t_{m-1}) \quad \text{are independent.}$$

Thus (6) will establish the Theorem.

Now in the case $m = 2$, we have

$$\begin{aligned}
E(e^{i[\lambda_1 W(t_1)+\lambda_2(W(t_2)-W(t_1))]}) &= E(e^{i[(\lambda_1-\lambda_2)W(t_1)+\lambda_2 W(t_2)]}) \\
&= E(e^{i(\lambda_1-\lambda_2)\sum_{k=0}^\infty A_k s_k(t_1)+i\lambda_2\sum_{k=0}^\infty A_k s_k(t_2)}) \\
&= \prod_{k=0}^\infty E(e^{iA_k[(\lambda_1-\lambda_2)s_k(t_1)+\lambda_2 s_k(t_2)]}) \\
&= \prod_{k=0}^\infty e^{-\frac{1}{2}((\lambda_1-\lambda_2)s_k(t_1)+\lambda_2 s_k(t_2))^2} \\
&= e^{-\frac{1}{2}\sum_{k=0}^\infty (\lambda_1-\lambda_2)^2 s_k^2(t_1)+2(\lambda_1-\lambda_2)\lambda_2 s_k(t_1)s_k(t_2)+\lambda_2^2 s_k^2(t_2)} \\
&= e^{-\frac{1}{2}[(\lambda_1-\lambda_2)^2 t_1+2(\lambda_1-\lambda_2)\lambda_2 t_1+\lambda_2^2 t_2]} \quad \text{by Lemma 4} \\
&= e^{-\frac{1}{2}[\lambda_1^2 t_1+\lambda_2^2(t_2-t_1)]}.
\end{aligned}$$

This is (6) for $m = 2$, and the general case follows similarly. □

THEOREM (Existence of one-dimensional Brownian motion). *Let $(\Omega, \mathcal{U}, P)$ be a probability space on which countably many $N(0,1)$, independent random variables $\{A_n\}_{n=1}^\infty$ are defined. Then there exists a 1-dimensional Brownian motion $W(\cdot)$ defined for $\omega \in \Omega$, $t \geq 0$.*

Outline of proof. The theorem above demonstrated how to build a Brownian motion on $0 \leq t \leq 1$. As we can reindex the $N(0,1)$ random variables to obtain countably many families of countably many random variables, we can therefore build countably many independent Brownian motions $W^n(t)$ for $0 \leq t \leq 1$.

We assemble these inductively by setting

$$W(t) := W(n-1) + W^n(t-(n-1)) \quad \text{for } n-1 \leq t \leq n.$$

Then $W(\cdot)$ is a one-dimensional Brownian motion, defined for all times $t \geq 0$. □

This theorem shows we can construct a Brownian motion defined on any probability space on which there exist countably many independent $N(0,1)$ random variables.

We mostly followed Lamperti [**L1**] for the foregoing theory.

3. BROWNIAN MOTION IN $\mathbb{R}^n$

It is straightforward to extend our definitions to Brownian motions taking values in $\mathbb{R}^n$.

DEFINITION. An $\mathbb{R}^n$-valued stochastic process $\mathbf{W}(\cdot) = (W^1(\cdot), \dots, W^n(\cdot))$ is an n-dimensional *Wiener process* (or *Brownian motion*) provided

(i) for each $k = 1, \dots, n$, $W^k(\cdot)$ is a 1-dimensional Wiener process,
and

(ii) the σ-algebras $\mathcal{W}^k := \mathcal{U}(W^k(t) \,|\, t \geq 0)$ are independent, $k = 1, \dots, n$.

By the arguments above we can build a probability space and on it n independent 1-dimensional Wiener processes $W^k(\cdot)$ $(k = 1, \dots, n)$. Then $\mathbf{W}(\cdot) := (W^1(\cdot), \dots, W^n(\cdot))$ is an n-dimensional Brownian motion.

LEMMA. *If $\mathbf{W}(\cdot)$ is an n-dimensional Wiener process, then*

$$(i) \qquad E(W^k(t)W^l(s)) = (t \wedge s)\delta_{kl} \quad (k, l = 1, \dots, n),$$

$$(ii) \quad E((W^k(t) - W^k(s))(W^l(t) - W^l(s))) = (t-s)\delta_{kl} \quad (k, l = 1, \dots, n;\ t \geq s \geq 0.)$$

PROOF. If $k \neq l$, $E(W^k(t)W^l(s)) = E(W^k(t))E(W^l(s)) = 0$, by independence. The proof of (ii) is similar. □

THEOREM. (i) *If $\mathbf{W}(\cdot)$ is an n-dimensional Brownian motion, then $\mathbf{W}(t)$ is $N(0, tI)$ for each time $t > 0$. Therefore*

$$P(\mathbf{W}(t) \in A) = \frac{1}{(2\pi t)^{n/2}} \int_A e^{-\frac{|x|^2}{2t}}\, dx$$

for each Borel subset $A \subseteq \mathbb{R}^n$.

(ii) *More generally, for each $m = 1, 2, \dots$ and each function $f : \mathbb{R}^n \times \mathbb{R}^n \times \cdots \mathbb{R}^n \to \mathbb{R}$, we have*

(7)

$$Ef(\mathbf{W}(t_1), \dots, \mathbf{W}(t_m)) = \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} f(x_1, \dots, x_m) g(x_1, t_1 \,|\, 0) g(x_2, t_2 - t_1 \,|\, x_1) \dots g(x_m, t_m - t_{m-1} \,|\, x_{m-1})\, dx_m \dots dx_1.$$

where

$$g(x, t \,|\, y) := \frac{1}{(2\pi t)^{n/2}} e^{-\frac{|x-y|^2}{2t}}.$$

PROOF. For each time $t > 0$, the random variables $W^1(t), \ldots, W^n(t)$ are independent. Consequently for each point $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we have

$$\begin{aligned} f_{\mathbf{W}(t)}(x_1, \ldots, x_n) &= f_{W^1(t)}(x_1) \cdots f_{W^n(t)}(x_n) \\ &= \frac{1}{(2\pi t)^{1/2}} e^{-\frac{x_1^2}{2t}} \cdots \frac{1}{(2\pi t)^{1/2}} e^{-\frac{x_n^2}{2t}} \\ &= \frac{1}{(2\pi t)^{n/2}} e^{-\frac{|x|^2}{2t}} = g(x, t \,|\, 0). \end{aligned}$$

We prove formula (7) as in the one-dimensional case. □

C. SAMPLE PATH PROPERTIES

In this section we will demonstrate that for almost every ω, the sample path $t \mapsto \mathbf{W}(t, \omega)$ is uniformly Hölder continuous for each exponent $\gamma < \frac{1}{2}$, but is nowhere Hölder continuous with any exponent $\gamma > \frac{1}{2}$. In particular $t \mapsto \mathbf{W}(t, \omega)$ almost surely is nowhere differentiable and is of infinite variation for each time interval.

DEFINITIONS. (i) Let $0 < \gamma \leq 1$. A function $f : [0, T] \to \mathbb{R}$ is called *uniformly Hölder continuous* with exponent $\gamma > 0$ if there exists a constant K such that

$$|f(t) - f(s)| \leq K|t - s|^\gamma \quad \text{for all } s, t \in [0, T].$$

(ii) We say f is Hölder continuous with exponent $\gamma > 0$ *at the point* s if there exists a constant K such that

$$|f(t) - f(s)| \leq K|t - s|^\gamma \quad \text{for all } t \in [0, T].$$

1. CONTINUITY OF SAMPLE PATHS.

A good general theorem to prove Hölder continuity is this important theorem of Kolmogorov:

THEOREM. *Let $\mathbf{X}(\cdot)$ be a stochastic process with continuous sample paths a.s., such that*

$$E(|\mathbf{X}(t) - \mathbf{X}(s)|^\beta) \leq C|t - s|^{1+\alpha}$$

for constants $\beta, \alpha > 0$, $C \geq 0$ and for all $0 \leq t, s$.

Then for each $0 < \gamma < \frac{\alpha}{\beta}$, $T > 0$, and almost every ω, there exists a constant $K = K(\omega, \gamma, T)$ such that

$$|\mathbf{X}(t, \omega) - \mathbf{X}(s, \omega)| \leq K|t - s|^\gamma \quad \textit{for all } 0 \leq s, t \leq T.$$

Hence the sample path $t \mapsto \mathbf{X}(t, \omega)$ is uniformly Hölder continuous with exponent γ on $[0, T]$.

APPLICATION TO BROWNIAN MOTION. Consider $\mathbf{W}(\cdot)$, an n-dimensional Brownian motion. We have for all integers $m = 1, 2, \ldots$

$$\begin{aligned} E(|\mathbf{W}(t) - \mathbf{W}(s)|^{2m}) &= \frac{1}{(2\pi r)^{n/2}} \int_{\mathbb{R}^n} |x|^{2m} e^{-\frac{|x|^2}{2r}} dx \quad \text{for } r = t - s > 0 \\ &= \frac{1}{(2\pi)^{n/2}} r^m \int_{\mathbb{R}^n} |y|^{2m} e^{-\frac{|y|^2}{2}} dy \quad \left(y = \frac{x}{\sqrt{r}} \right) \\ &= Cr^m = C|t - s|^m. \end{aligned}$$

Thus the hypotheses of Kolmogorov's theorem hold for $\beta = 2m$, $\alpha = m - 1$. The process $\mathbf{W}(\cdot)$ is thus Hölder continuous a.s. for exponents

$$0 < \gamma < \frac{\alpha}{\beta} = \frac{1}{2} - \frac{1}{2m} \quad \text{for all } m.$$

Thus for almost all ω and any $T > 0$, the sample path $t \mapsto \mathbf{W}(t, \omega)$ is uniformly Hölder continuous on $[0, T]$ for each exponent $0 < \gamma < 1/2$. □

PROOF OF THEOREM. 1. For simplicity, take $T = 1$. Pick any

$$0 < \gamma < \frac{\alpha}{\beta}. \tag{8}$$

Now define for $n = 1, \ldots,$

$$A_n := \left\{ \left| \mathbf{X}(\frac{i+1}{2^n}) - \mathbf{X}(\frac{i}{2^n}) \right| > \frac{1}{2^{n\gamma}} \text{ for } \textit{some integer } 0 \le i < 2^n \right\}.$$

Then

$$\begin{aligned} P(A_n) &\le \sum_{i=0}^{2^n - 1} P\left(\left| \mathbf{X}(\frac{i+1}{2^n}) - \mathbf{X}(\frac{i}{2^n}) \right| > \frac{1}{2^{n\gamma}} \right) \\ &\le \sum_{i=0}^{2^n - 1} E\left(\left| \mathbf{X}(\frac{i+1}{2^n}) - \mathbf{X}(\frac{i}{2^n}) \right|^{\beta} \right) \left(\frac{1}{2^{n\gamma}} \right)^{-\beta} \quad \text{by Chebyshev's inequality} \\ &\le C \sum_{i=0}^{2^n - 1} \left(\frac{1}{2^n} \right)^{1+\alpha} \left(\frac{1}{2^{n\gamma}} \right)^{-\beta} \\ &= C 2^{n(-\alpha + \gamma\beta)}. \end{aligned}$$

Since (8) forces $-\alpha + \gamma\beta < 0$, we deduce $\sum_{n=1}^{\infty} P(A_n) < \infty$; whence the Borel–Cantelli Lemma implies

$$P(A_n \text{ i.o.}) = 0.$$

So for a.e. ω there exists $m = m(\omega)$ such that

$$\left| \mathbf{X}(\frac{i+1}{2^n}, \omega) - \mathbf{X}(\frac{i}{2^n}, \omega) \right| \le \frac{1}{2^{n\gamma}} \quad \text{for } 0 \le i \le 2^n - 1$$

provided $n \geq m$. But then we have

$$\text{(9)} \qquad \begin{cases} \left|\mathbf{X}(\frac{i+1}{2^n}, \omega) - \mathbf{X}(\frac{i}{2^n}, \omega)\right| \leq K \frac{1}{2^{n\gamma}} & \text{for } 0 \leq i \leq 2^n - 1 \\ \text{for } \textit{all } n \geq 0, \end{cases}$$

if we select $K = K(\omega)$ large enough.

2.* We now claim (9) implies the stated Hölder continuity. To see this, fix $\omega \in \Omega$ for which (9) holds. Let $t_1, t_2 \in [0,1]$ be dyadic rationals, $0 < t_2 - t_1 < 1$. Select $n \geq 1$ so that

$$\text{(10)} \qquad 2^{-n} \leq t < 2^{-(n-1)} \quad \text{for } t := t_2 - t_1.$$

We can write

$$\begin{cases} t_1 = \frac{i}{2^n} - \frac{1}{2^{p_1}} - \cdots - \frac{1}{2^{p_k}} & (n < p_1 < \cdots < p_k) \\ t_2 = \frac{j}{2^n} + \frac{1}{2^{q_1}} + \cdots + \frac{1}{2^{q_l}} & (n < q_1 < \cdots < q_l) \end{cases}$$

for

$$t_1 \leq \frac{i}{2^n} \leq \frac{j}{2^n} \leq t_2.$$

Then

$$\frac{j-i}{2^n} \leq t < \frac{1}{2^{n-1}}$$

and so $j = i$ or $i+1$. In view of (9),

$$|\mathbf{X}(i/2^n, \omega) - \mathbf{X}(j/2^n, \omega)| \leq K \left|\frac{i-j}{2^n}\right|^\gamma \leq K t^\gamma.$$

Furthermore

$$|\mathbf{X}(i/2^n - 1/2^{p_1} - \cdots - 1/2^{p_r}, \omega) - \mathbf{X}(i/2^n - 1/2^{p_1} - \cdots - 1/2^{p_{r-1}}, \omega)| \leq K \left|\frac{1}{2^{p_r}}\right|^\gamma$$

for $r = 1, \ldots, k$; and consequently

$$\begin{aligned} |\mathbf{X}(t_1, \omega) - \mathbf{X}(i/2^n, \omega)| &\leq K \sum_{r=1}^{k} \left|\frac{1}{2^{p_r}}\right|^\gamma \\ &\leq \frac{K}{2^{n\gamma}} \sum_{r=1}^{\infty} \frac{1}{2^{r\gamma}} \quad \text{since } p_r > n \\ &= \frac{C}{2^{n\gamma}} \leq C t^\gamma \quad \text{by (10).} \end{aligned}$$

In the same way we deduce

$$|\mathbf{X}(t_2, \omega) - \mathbf{X}(j/2^n, \omega)| \leq C t^\gamma.$$

Add up the estimates above, to discover

$$|\mathbf{X}(t_1, \omega) - \mathbf{X}(t_2, \omega)| \leq C |t_1 - t_2|^\gamma$$

*Omit the second step in this proof on first reading.

for all dyadic rationals $t_1, t_2 \in [0,1]$ and some constant $C = C(\omega)$. Since $t \mapsto \mathbf{X}(t,\omega)$ is continuous for a.e. ω, the estimate above holds for all $t_1, t_2 \in [0,1]$. □

Remark. The proof above can in fact be modified to show that if $\mathbf{X}(\cdot)$ is a stochastic process such that

$$E(|\mathbf{X}(t) - \mathbf{X}(s)|^\beta) \le C|t-s|^{1+\alpha} \quad (\alpha, \beta > 0, C \ge 0),$$

then $\mathbf{X}(\cdot)$ has a version $\tilde{\mathbf{X}}(\cdot)$ such that a.e. sample path is Hölder continuous for each exponent $0 < \gamma < \alpha/\beta$. (We call $\tilde{\mathbf{X}}(\cdot)$ a *version* of $\mathbf{X}(\cdot)$ if $P(\mathbf{X}(t) = \tilde{\mathbf{X}}(t)) = 1$ for all $t \ge 0$.)

So any Wiener process has a version with continuous sample paths a.s. □

2. NOWHERE DIFFERENTIABILITY

Next we prove that sample paths of Brownian motion are with probability one nowhere Hölder continuous with exponent greater than $\frac{1}{2}$, and thus are nowhere differentiable.

THEOREM. (i) *For each $\frac{1}{2} < \gamma \le 1$ and almost every ω, $t \mapsto \mathbf{W}(t,\omega)$ is nowhere Hölder continuous with exponent γ.*

(ii) *In particular, for almost every ω, the sample path $t \mapsto \mathbf{W}(t,\omega)$ is nowhere differentiable and is of infinite variation on each subinterval.*

PROOF. (Dvoretzky, Erdös, Kakutani) 1. It suffices to consider a one-dimensional Brownian motion, and we may for simplicity consider only times $0 \le t \le 1$.

Fix an integer N so large that

$$N\left(\gamma - \frac{1}{2}\right) > 1.$$

Now if the function $t \mapsto W(t,\omega)$ is Hölder continuous with exponent γ at some point $0 \le s < 1$, then

$$|W(t,\omega) - W(s,\omega)| \le K|t-s|^\gamma \quad \text{for all } t \in [0,1] \text{ and some constant } K.$$

For $n \gg 1$, set $i = [ns] + 1$ and note that for $j = i, i+1, \dots, i+N-1$

$$\begin{aligned}\left|W(\frac{j}{n},\omega) - W(\frac{j+1}{n},\omega)\right| &\le \left|W(s,\omega) - W(\frac{j}{n},\omega)\right| \\ &\quad + \left|W(s,\omega) - W(\frac{j+1}{n},\omega)\right| \\ &\le K\left(\left|s - \frac{j}{n}\right|^\gamma + \left|s - \frac{j+1}{n}\right|^\gamma\right) \\ &\le \frac{M}{n^\gamma}\end{aligned}$$

for some constant M. Thus

$$\omega \in A^i_{M,n} := \left\{ \left| W(\frac{j}{n}) - W(\frac{j+1}{n}) \right| \leq \frac{M}{n^\gamma} \text{ for } j = i, \ldots, i+N-1 \right\}$$

for some $1 \leq i \leq n$, some $M \geq 1$, and all large n.

Therefore the set of $\omega \in \Omega$ such that $W(\omega, \cdot)$ is Hölder continuous with exponent γ at some time $0 \leq s < 1$ is contained in

$$\bigcup_{M=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \bigcup_{i=1}^{n} A^i_{M,n}.$$

We will show this event has probability 0.

2. For all k and M,

$$\begin{aligned} P\left(\bigcap_{n=k}^{\infty} \bigcup_{i=1}^{n} A^i_{M,n} \right) &\leq \liminf_{n\to\infty} P\left(\bigcup_{i=1}^{n} A^i_{M,n} \right) \\ &\leq \liminf_{n\to\infty} \sum_{i=1}^{n} P(A^i_{M,n}) \\ &\leq \liminf_{n\to\infty} n \left(P\left(|W(\frac{1}{n})| \leq \frac{M}{n^\gamma} \right) \right)^N, \end{aligned}$$

since the random variables $W(\frac{j+1}{n}) - W(\frac{j}{n})$ are $N\left(0, \frac{1}{n}\right)$ and independent. Now

$$\begin{aligned} P\left(|W(\frac{1}{n})| \leq \frac{M}{n^\gamma} \right) &= \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-Mn^{-\gamma}}^{Mn^{-\gamma}} e^{-\frac{nx^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-Mn^{1/2-\gamma}}^{Mn^{1/2-\gamma}} e^{-\frac{y^2}{2}} dy \\ &\leq Cn^{1/2-\gamma}. \end{aligned}$$

We use this calculation to deduce:

$$P\left(\bigcap_{n=k}^{\infty} \bigcup_{i=1}^{n} A^i_{M,n} \right) \leq \liminf_{n\to\infty} nC[n^{1/2-\gamma}]^N = 0,$$

since $N(\gamma - 1/2) > 1$. This holds for all k, M. Thus

$$P\left(\bigcup_{M=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \bigcup_{i=1}^{n} A^i_{M,n} \right) = 0,$$

and assertion (i) of the Theorem follows.

3. If $W(t, \omega)$ is differentiable at s, then $W(t, \omega)$ would be Hölder continuous (with exponent 1) at s. But this is almost surely not so. If $W(t, \omega)$ were of finite variation on some subinterval, it would then be differentiable almost everywhere there. $\square$

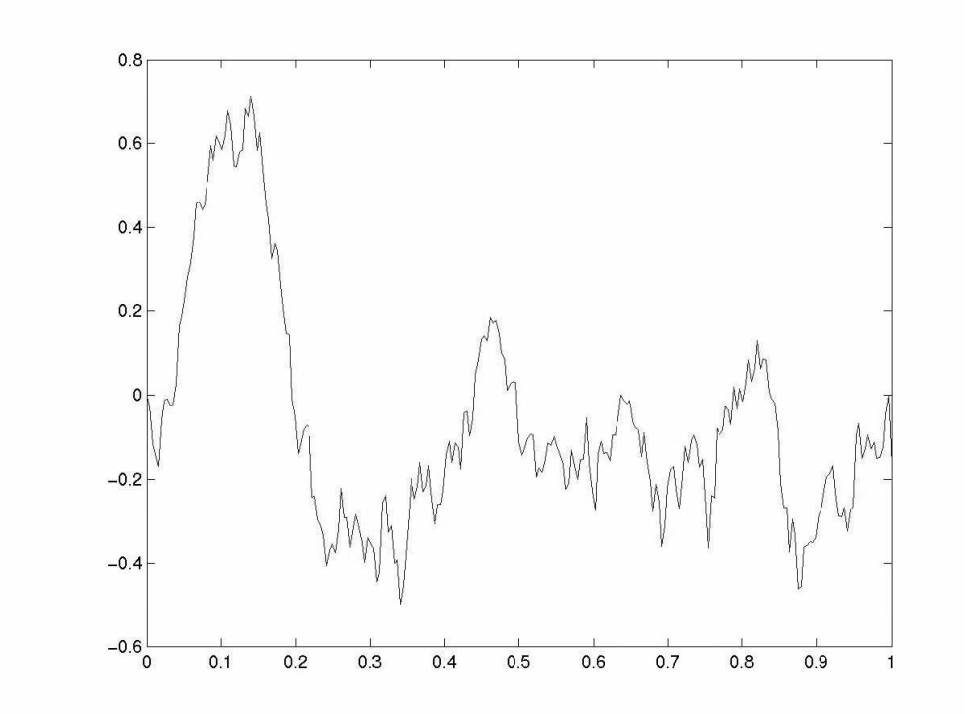
Interpretation. The idea underlying the proof is that if

$$|W(t,\omega) - W(s,\omega)| \le K|t-s|^\gamma \quad \text{for all } t,$$

then

$$|W(\frac{j}{n},\omega) - W(\frac{j+1}{n},\omega)| \le \frac{M}{n^\gamma}$$

for all $n \gg 1$ and at least N values of j. But these are independent events of small probability. The probability that the above inequality holds for all these j's is a small number to the large power N, and is therefore extremely small. □



A sample path of Brownian motion

D. MARKOV PROPERTY

DEFINITION. If $\mathcal{V}$ is a σ-algebra, $\mathcal{V} \subseteq \mathcal{U}$, then

$$P(A \,|\, \mathcal{V}) := E(\chi_A \,|\, \mathcal{V}) \quad \text{for } A \in \mathcal{U}.$$

Therefore $P(A \,|\, \mathcal{V})$ is a random variable, the *conditional probability of A, given $\mathcal{V}$.*

DEFINITION. If $\mathbf{X}(\cdot)$ is a stochastic process, the σ-algebra

$$\mathcal{U}(s) := \mathcal{U}(\mathbf{X}(r) \,|\, 0 \le r \le s)$$

is called the *history* of the process up to and including time s.

We can informally interpret $\mathcal{U}(s)$ as recording the information available from our observing $\mathbf{X}(r)$ for all times $0 \le r \le s$.

DEFINITION. An $\mathbb{R}^n$-valued stochastic process $\mathbf{X}(\cdot)$ is called a *Markov process* if

$$P(\mathbf{X}(t) \in B \,|\, \mathcal{U}(s)) = P(\mathbf{X}(t) \in B \,|\, \mathbf{X}(s)) \text{ a.s.}$$

for all $0 \le s \le t$ and all Borel subset B of $\mathbb{R}^n$.

The idea of this definition is that, *given* the current value $\mathbf{X}(s)$, you can predict the probabilities of future values of $\mathbf{X}(t)$ just as well as if you knew the entire history of the process before time s. Loosely speaking, the process only "knows" its value at time s and does not "remember" how it got there.

THEOREM. *Let $\mathbf{W}(\cdot)$ be an n-dimensional Wiener process. Then $\mathbf{W}(\cdot)$ is a Markov process, and*

$$P(\mathbf{W}(t) \in B \,|\, \mathbf{W}(s)) = \frac{1}{(2\pi(t-s))^{n/2}} \int_B e^{-\frac{|x-\mathbf{W}(s)|^2}{2(t-s)}} dx \quad a.s. \tag{13}$$

for all $0 \le s < t$, and Borel sets B .

Note carefully that each side of this identity is a random variable.

PROOF. We will only prove (13). Let A be a Borel set and write

$$\Phi(y) := \frac{1}{(2\pi(t-s))^{n/2}} \int_A e^{-\frac{|x-y|^2}{2(t-s)}} dx.$$

As $\Phi(\mathbf{W}(s))$ is $\mathcal{U}(\mathbf{W}(s))$ measurable, we must show

$$\int_C \chi_{\{\mathbf{W}(t) \in A\}} dP = \int_C \Phi(\mathbf{W}(s))\, dP \quad \text{for all } C \in \mathcal{U}(\mathbf{W}(s)). \tag{14}$$

Now if $C \in \mathcal{U}(\mathbf{W}(s))$, then $C = \{\mathbf{W}(s) \in B\}$ for some Borel set $B \subseteq \mathbb{R}^n$. Hence

$$\begin{aligned} \int_C \chi_{\{\mathbf{W}(t) \in A\}} dP &= P(\mathbf{W}(s) \in B, \mathbf{W}(t) \in A) \\ &= \int_B \int_A g(y, s \,|\, 0) g(x, t-s \,|\, y)\, dx dy \\ &= \int_B g(y, s \,|\, 0) \Phi(y)\, dy. \end{aligned}$$

On the other hand,

$$\begin{aligned} \int_C \Phi(\mathbf{W}(s)) dP &= \int_\Omega \chi_B(\mathbf{W}(s)) \Phi(\mathbf{W}(s))\, dP \\ &= \int_{\mathbb{R}^n} \chi_B(y) \Phi(y) \frac{e^{-\frac{|y|^2}{2s}}}{(2\pi s)^{n/2}}\, dy \\ &= \int_B g(y, s \,|\, 0) \Phi(y)\, dy, \end{aligned}$$

and this last expression agrees with that above. This verifies (14), and so establishes (13). □

Interpretation. The Markov property partially explains the nondifferentiability of sample paths for Brownian motion, as discussed before in §C.

If $\mathbf{W}(s, \omega) = b$, say, then the future behavior of $\mathbf{W}(t, \omega)$ depends only upon this fact and not on how $\mathbf{W}(t, \omega)$ approached the point b as $t \to s^-$. Thus the path "cannot remember" how to leave b in such a way that $\mathbf{W}(\cdot, \omega)$ will have a tangent there. □

CHAPTER 4: STOCHASTIC INTEGRALS, ITÔ'S FORMULA

A. **Motivation**
B. **Definition and properties of Itô integral**
C. **Indefinite Itô integrals**
D. **Itô's formula**
E. **Itô integral in higher dimensions**

A. MOTIVATION

Remember from Chapter 1 that we want to develop a theory of stochastic differential equations of the form

$$(SDE) \qquad \begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0, \end{cases}$$

which we will in Chapter 5 interpret to mean

$$(1) \qquad \mathbf{X}(t) = \mathbf{X}_0 + \int_0^t \mathbf{b}(\mathbf{X}, s)\, ds + \int_0^t \mathbf{B}(\mathbf{X}, s)\, d\mathbf{W}$$

for all times $t \geq 0$. But before we can study and solve such an integral equation, we must first *define*

$$\int_0^T \mathbf{G}\, d\mathbf{W}$$

for some wide class of stochastic processes G, so that the right-hand side of (1) at least makes sense. Observe also that this is not at all obvious. For instance, since $t \mapsto \mathbf{W}(t, \omega)$ is of infinite variation for almost every ω, then $\int_0^T \mathbf{G}\, d\mathbf{W}$ simply cannot be understood as an ordinary integral.

A FIRST DEFINITION. Suppose now $n = m = 1$. One possible definition is due to Paley, Wiener and Zygmund [P-W-Z]. Suppose $g : [0, 1] \to \mathbb{R}$ is continuously differentiable, with $g(0) = g(1) = 0$. Note carefully: g is an ordinary, deterministic function and *not* **a stochastic process. Then let us** *define*

$$\int_0^1 g\, dW := -\int_0^1 g' W\, dt.$$

Note that $\int_0^1 g\, dW$ is therefore a random variable. Let us check out the properties following from this definition:

LEMMA (Properties of the Paley–Wiener–Zygmund integral).
(i) $E\left(\int_0^1 g\, dW\right) = 0.$
(ii) $E\left(\left(\int_0^1 g\, dW\right)^2\right) = \int_0^1 g^2\, dt.$

PROOF. 1. $E\left(\int_0^1 g\,dW\right) = -\int_0^1 g' \underbrace{E(W(t))}_{=0}\,dt.$

2. To confirm (ii), we calculate

$$\begin{aligned}
E\left(\left(\int_0^1 g\,dW\right)^2\right) &= E\left(\int_0^1 g'(t)W(t)\,dt \int_0^1 g'(s)W(s)\,ds\right) \\
&= \int_0^1 \int_0^1 g'(t)g'(s)\underbrace{E(W(t)W(s))}_{=t\wedge s} ds dt \\
&= \int_0^1 g'(t)\left(\int_0^t sg'(s)\,ds + \int_t^1 tg'(s)\,ds\right) dt \\
&= \int_0^1 g'(t)\left(tg(t) - \int_0^t g\,ds - tg(t)\right) dt \\
&= \int_0^1 g'(t)\left(-\int_0^t g\,ds\right) dt = \int_0^1 g^2\,dt.
\end{aligned}$$

□

Discussion. Suppose now $g \in L^2(0,1)$. We can take a sequence of C^1 functions g_n, as above, such that $\int_0^1 (g_n - g)^2\,dt \to 0$. In view of property (ii),

$$E\left(\left(\int_0^1 g_m\,dW - \int_0^1 g_n\,dW\right)^2\right) = \int_0^1 (g_m - g_n)^2\,dt,$$

and therefore $\{\int_0^1 g_n\,dW\}_{n=1}^\infty$ is a Cauchy sequence in $L^2(\Omega)$. Consequently we can define

$$\int_0^1 g\,dW := \lim_{n\to\infty} \int_0^1 g_n\,dW.$$

The extended definition still satisfies properties (i) and (ii).

This is a reasonable definition of $\int_0^1 g\,dW$, except that this only makes sense for functions $g \in L^2(0,1)$, and *not* for stochastic processes. If we wish to define the integral in (1),

$$\int_0^t \mathbf{B}(\mathbf{X}, s)\,d\mathbf{W},$$

then the integrand $\mathbf{B}(\mathbf{X}, t)$ is a stochastic process and the definition above will not suffice.

We must devise a definition for a wider class of integrands (although the definition we finally decide on will agree with that of Paley, Wiener, Zygmund if g happens to be a deterministic C^1 function, with $g(0) = g(1) = 0$).

RIEMANN SUMS. To continue our study of stochastic integrals with random integrands, let us think about what might be an appropriate definition for

$$\int_0^T W\,dW = ?,$$

where $W(\cdot)$ is a 1-dimensional Brownian motion. A reasonable procedure is to construct a *Riemann sum approximation*, and then–if possible–to pass to limits.

DEFINITIONS. (i) If $[0,T]$ is an interval, a *partition* P of $[0,T]$ is a finite collection of points in $[0,T]$:

$$P := \{0 = t_0 < t_1 < \cdots < t_m = T\}.$$

(ii) Let the *mesh size* of P be $|P| := \max_{0\le k\le m-1} |t_{k+1} - t_k|$.
(iii) For fixed $0 \le \lambda \le 1$ and P a given partition of $[0,T]$, set

$$\tau_k := (1-\lambda)t_k + \lambda t_{k+1} \quad (k = 0, \ldots, m-1).$$

For such a partition P and for $0 \le \lambda \le 1$, we define

$$R = R(P,\lambda) := \sum_{k=0}^{m-1} W(\tau_k)(W(t_{k+1}) - W(t_k)).$$

This is the corresponding Riemann sum approximation of $\int_0^T W\,dW$. The key question is this: what happens if $|P| \to 0$, with λ fixed?

LEMMA (Quadratic variation). *Let $[a,b]$ be an interval in $[0,\infty)$, and suppose*

$$P^n := \{a = t_0^n < t_1^n < \cdots < t_{m_n}^n = b\}$$

are partitions of $[a,b]$, with $|P^n| \to 0$ as $n \to \infty$. Then

$$\sum_{k=0}^{m_n-1} (W(t_{k+1}^n) - W(t_k^n))^2 \to b - a$$

in $L^2(\Omega)$ as $n \to \infty$.

This assertion partly justifies the heuristic idea, introduced in Chapter 1, that

$$dW \approx (dt)^{1/2}.$$

Proof. Set $Q_n := \sum_{k=0}^{m_n-1}(W(t_{k+1}^n) - W(t_k^n))^2$. Then

$$Q_n - (b-a) = \sum_{k=0}^{m_n-1} ((W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n)).$$

Hence

$$E((Q_n - (b-a))^2) = \sum_{k=0}^{m_n-1}\sum_{j=0}^{m_n-1} E([(W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n)] \\ [(W(t_{j+1}^n) - W(t_j^n))^2 - (t_{j+1}^n - t_j^n)]).$$

For $k \ne j$, the term in the double sum is

$$E((W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n))E(\cdots),$$

according to the independent increments, and thus equals 0, as $W(t) - W(s)$ is $N(0, t-s)$ for all $t \geq s \geq 0$. Hence

$$E((Q_n - (b-a))^2) = \sum_{k=0}^{m_n - 1} E((Y_k^2 - 1)^2 (t_{k+1}^n - t_k^n)^2),$$

where

$$Y_k = Y_k^n := \frac{W(t_{k+1}^n) - W(t_k^n)}{\sqrt{t_{k+1}^n - t_k^n}} \quad \text{is } N(0,1).$$

Therefore for some constant C we have

$$\begin{aligned} E((Q_n - (a-b))^2) &\leq C \sum_{k=0}^{m_n - 1} (t_{k+1}^n - t_k^n)^2 \\ &\leq C \,|\, P^n \,|\, (b-a) \to 0, \qquad \text{as } n \to \infty. \end{aligned}$$

□

Remark. Passing if necessary to a subsequence,

$$\sum_{k=0}^{m_n - 1} (W(t_{k+1}^n) - W(t_k^n))^2 \to b - a \quad \text{a.s.}$$

Pick an ω for which this holds and also for which the sample path is uniformly Hölder continuous with some exponent $0 < \gamma < \frac{1}{2}$. Then

$$b - a \leq K \limsup_{n \to \infty} |P_n|^\gamma \sum_{k=0}^{m_n - 1} |W(t_{k+1}^n) - W(t_k^n)|$$

for a constant K. Since $|P_n| \to 0$, we see again that sample paths have infinite variation with probability one:

$$\sup_P \left\{ \sum_{k=0}^{m-1} |W(t_{k+1}) - W(t_k)| \right\} = \infty.$$

□

Let us now return to the question posed above, as to the limit of the Riemann sum approximations.

LEMMA. *If P^n denotes a partition of $[0,T]$ and $0 \leq \lambda \leq 1$ is fixed, define*

$$R_n := \sum_{k=0}^{m_n - 1} W(\tau_k^n)(W(t_{k+1}^n) - W(t_k^n)).$$

Then

$$\lim_{n \to \infty} R_n = \frac{W(T)^2}{2} + \left(\lambda - \frac{1}{2}\right) T,$$

the limit taken in $L^2(\Omega)$. That is,

$$E\left(\left(R_n - \frac{W(T)^2}{2} - \left(\lambda - \frac{1}{2}\right)T\right)^2\right) \to 0.$$

In particular *the limit of the Riemann sum approximations depends upon the choice of intermediate points $t_k^n \le \tau_k^n \le t_{k+1}^n$*, where $\tau_k^n = (1-\lambda)t_k^n + \lambda t_{k+1}^n$.

PROOF. We have

$$\begin{aligned} R_n &:= \sum_{k=0}^{m_n-1} W(\tau_k^n)(W(t_{k+1}^n) - W(t_k^n)) \\ &= \frac{W^2(T)}{2} - \underbrace{\frac{1}{2}\sum_{k=0}^{m_n-1}(W(t_{k+1}^n) - W(t_k^n))^2}_{=:A} \\ &\quad + \underbrace{\sum_{k=0}^{m_n-1}(W(\tau_k^n) - W(t_k^n))^2}_{=:B} + \underbrace{\sum_{k=0}^{m_n-1}(W(t_{k+1}^n) - W(\tau_k^n))(W(\tau_k^n) - W(t_k^n))}_{=:C}. \end{aligned}$$

According to the foregoing Lemma, $A \to \frac{T}{2}$ in $L^2(\Omega)$ as $n \to \infty$. A similar argument shows that $B \to \lambda T$ as $n \to \infty$. Next we study the term C:

$$\begin{aligned} &E([\sum_{k=0}^{m_n-1}(W(t_{k+1}^n) - W(\tau_k^n))(W(\tau_k^n) - W(t_k^n))]^2) \\ &\qquad = \sum_{k=0}^{m_n-1} E([W(t_{k+1}^n) - W(\tau_k^n)]^2)E([W(\tau_k^n) - W(t_k^n)]^2) \\ &\qquad\qquad \text{(independent increments)} \\ &\qquad = \sum_{k=0}^{m_n-1}(1-\lambda)(t_{k+1}^n - t_k^n)\lambda(t_{k+1}^n - t_k^n) \\ &\qquad \le \lambda(1-\lambda)T|P^n| \to 0. \end{aligned}$$

Hence $C \to 0$ in $L^2(\Omega)$ as $n \to \infty$.

We combine the limiting expressions for the terms A, B, C, and thereby establish the Lemma. □

It turns out that Itô's definition (later, in §B) of $\int_0^T W\,dW$ corresponds to the choice $\lambda = 0$. That is,

$$\int_0^T W\,dW = \frac{W^2(T)}{2} - \frac{T}{2}$$

and, more generally,

$$\int_s^r W\,dW = \frac{W^2(r) - W^2(s)}{2} - \frac{(r-s)}{2} \qquad \text{for all } r \geq s \geq 0.$$

This is not what one would guess offhand. An alternative definition, due to *Stratonovich*, takes $\lambda = \frac{1}{2}$; so that

$$\int_0^T W \circ dW = \frac{W^2(T)}{2} \qquad \text{(Stratonovich integral)}.$$

See Chapter 6 for more.

More discussion. What are the advantages of taking $\lambda = 0$ and getting

$$\int_0^T W\,dW = \frac{W^2(T)}{2} - \frac{T}{2}?$$

First and most importantly, building the Riemann sum approximation by evaluating the integrand at the left-hand endpoint $\tau_k^n = t_k^n$ on each subinterval $[t_k^n, t_{k=1}^n]$ will ultimately permit the definition of

$$\int_0^T G\,dW$$

for a wide class of so-called "nonanticipating" stochastic processes $G(\cdot)$. Exact definitions are later, but the idea is that t represents time, and since we do not know what $W(\cdot)$ will do on $[t_k^n, t_{k+1}^n]$, it is best to use the known value of $G(t_k^n)$ in the approximation. Indeed, $G(\cdot)$ will in general depend on Brownian motion $W(\cdot)$, and we do not know at time t_k^n its future value at the future time $\tau_k^n = (1-\lambda)t_k^n + \lambda t_{k+1}^n$, if $\lambda > 0$. □

B. DEFINITION AND PROPERTIES OF ITÔ'S INTEGRAL

Let $W(\cdot)$ be a 1-dimensional Brownian motion defined on some probability space $(\Omega, \mathcal{U}, P)$.

DEFINITIONS. (i) The σ-algebra $\mathcal{W}(t) := \mathcal{U}(W(s) \,|\, 0 \leq s \leq t)$ is called the *history* of the Brownian motion up to (and including) time t.

(ii) The σ-algebra $\mathcal{W}^+(t) := \mathcal{U}(W(s) - W(t) \,|\, s \geq t)$ is the *future* of the Brownian motion beyond time t. □

DEFINITION. A family $\mathcal{F}(\cdot)$ of σ-algebras $\subseteq \mathcal{U}$ is called *nonanticipating* (with respect to $W(\cdot)$) if

(a) $\mathcal{F}(t) \supseteq \mathcal{F}(s)$ for all $t \geq s \geq 0$

(b) $\mathcal{F}(t) \supseteq \mathcal{W}(t)$ for all $t \geq 0$

(c) $\mathcal{F}(t)$ is independent of $\mathcal{W}^+(t)$ for all $t \geq 0$.

We also refer to $\mathcal{F}(\cdot)$ as a *filtration*.

IMPORTANT REMARK. We should informally think of $\mathcal{F}(t)$ as "containing all information available to us at time t". Our primary example will be $\mathcal{F}(t) := \mathcal{U}(W(s)\ (0 \leq s \leq t), X_0)$, where X_0 is a random variable *independent* of $\mathcal{W}^+(0)$.

This will be employed in Chapter 5, where X_0 will be the (possibly random) initial condition for a stochastic differential equation. □

DEFINITION. A real-valued stochastic process $G(\cdot)$ is called *nonanticipating* (with respect to $\mathcal{F}(\cdot)$) if for each time $t \geq 0$, $G(t)$ is $\mathcal{F}(t)$–measurable.

The idea is that for each time $t \geq 0$, the random variable $G(t)$ "depends upon only the information available in the σ-algebra $\mathcal{F}(t)$".

Discussion. We will actually need a slightly stronger notion, namely that $G(\cdot)$ be *progressively measurable*. This is however a bit subtle to define, and we will not do so here. The idea is that $G(\cdot)$ is nonanticipating and, in addition, is appropriately jointly measurable in the variables t and ω together.

These measure theoretic issues can be confusing to students, and so we pause here to emphasize the basic point, to be developed below. *For progressively measurable integrands $G(\cdot)$, we will be able to define, and understand, the stochastic integral $\int_0^T G\,dW$ in terms of some simple, useful and elegant formulas.* In other words, we will see that since at each moment of time "G depends only upon the past history of the Brownian motion", some nice identities hold, which would be false if G "depends upon the future behavior of the Brownian motion".

DEFINITIONS. (i) We denote by $\mathbb{L}^2(0,T)$ the space of all real–valued, progressively measurable stochastic processes $G(\cdot)$ such that

$$E\left(\int_0^T G^2\,dt\right) < \infty.$$

(ii) Likewise, $\mathbb{L}^1(0,T)$ is the space of all real–valued, progressively measurable processes $F(\cdot)$ such that

$$E\left(\int_0^T |F|\,dt\right) < \infty.$$

DEFINITION. A process $G \in \mathbb{L}^2(0,T)$ is called a *step process* if there exists a partition $P = \{0 = t_0 < t_1 < \cdots < t_m = T\}$ such that

$$G(t) \equiv G_k \qquad for \;\; t_k \leq t < t_{k+1} \quad (k = 0, \ldots, m-1).$$

Then each G_k is an $\mathcal{F}(t_k)$-measurable random variable, since G is nonanticipating.

DEFINITION. Let $G \in \mathbb{L}^2(0,T)$ be a step process, as above. Then

$$\int_0^T G\,dW := \sum_{k=0}^{m-1} G_k(W(t_{k+1}) - W(t_k))$$

is the *Itô stochastic integral* of G on the interval $(0,T)$.

Note carefully that this is a random variable.

LEMMA (Properties of stochastic integral for step processes). *We have for all constants $a, b \in \mathbb{R}$ and for all step processes $G, H \in \mathbb{L}^2(0,T)$:*

$$(i) \qquad \int_0^T aG + bH \, dW = a \int_0^T G \, dW + b \int_0^T H \, dW,$$

$$(ii) \qquad E\left(\int_0^T G \, dW \right) = 0,$$

$$(iii) \qquad E\left(\left(\int_0^T G \, dW \right)^2 \right) = E\left(\int_0^T G^2 \, dt \right).$$

Proof. 1. The first assertion is easy to check.
Suppose next $G(t) \equiv G_k$ for $t_k \leq t < t_{k+1}$. Then

$$E\left(\int_0^T G \, dW \right) = \sum_{k=0}^{m-1} E(G_k(W(t_{k+1}) - W(t_k))).$$

Now G_k is $\mathcal{F}(t_k)$-measurable and $\mathcal{F}(t_k)$ is independent of $\mathcal{W}^+(t_k)$. On the other hand, $W(t_{k+1}) - W(t_k)$ is $\mathcal{W}^+(t_k)$-measurable, and so G_k is independent of $W(t_{k+1}) - W(t_k)$. Hence

$$E(G_k(W(t_{k+1}) - W(t_k))) = E(G_k) \underbrace{E(W(t_{k+1}) - W(t_k))}_{=0}.$$

2. Furthermore,

$$E\left(\left(\int_0^T G \, dW \right)^2 \right) = \sum_{k,j=1}^{m-1} E\left(G_k G_j (W(t_{k+1}) - W(t_k))(W(t_{j+1}) - W(t_j)) \right).$$

Now if $j < k$, then $W(t_{k+1}) - W(t_k)$ is independent of $G_k G_j (W(t_{j+1}) - W(t_j))$. Thus

$$\begin{aligned} &E(G_k G_j (W(t_{k+1}) - W(t_k))(W(t_{j+1}) - W(t_j))) \\ &= \underbrace{E(G_k G_j (W(t_{j+1}) - W(t_j)))}_{<\infty} \underbrace{E(W(t_{k+1}) - W(t_k))}_{=0}. \end{aligned}$$

Consequently

$$E\left(\left(\int_0^T G\,dW\right)^2\right) = \sum_{k=0}^{m-1} E(G_k^2(W(t_{k+1}) - W(t_k))^2)$$
$$= \sum_{k=0}^{m-1} E(G_k^2)\underbrace{E((W(t_{k+1}) - W(t_k))^2)}_{=t_{k+1}-t_k}$$
$$= E\left(\int_0^T G^2\,dt\right).$$

□

APPROXIMATION BY STEP FUNCTIONS. The plan now is to approximate an arbitrary process $G \in \mathbb{L}^2(0,T)$ by step processes in $\mathbb{L}^2(0,T)$, and then pass to limits to define the Itô integral of G.

LEMMA (Approximation by step processes). *If $G \in \mathbb{L}^2(0,T)$, there exists a sequence of bounded step processes $G^n \in \mathbb{L}^2(0,T)$ such that*

$$E\left(\int_0^T |G - G^n|^2\,dt\right) \to 0.$$

Outline of proof. We omit the proof, but the idea is this: if $t \mapsto G(t,\omega)$ is continuous for almost every ω, we can set

$$G^n(t) := G(\frac{k}{n}) \text{ for } \frac{k}{n} \le t < \frac{k+1}{n},\ k = 0, \dots, [nT].$$

For a general $G \in \mathbb{L}^2(0,T)$, define

$$G^m(t) := \int_0^t me^{m(s-t)}G(s)\,ds.$$

Then $G^m \in \mathbb{L}^2(0,T)$, $t \mapsto G^m(t,\omega)$ is continuous for a.e. ω, and

$$\int_0^T |G^m - G|^2 dt \to 0 \text{ a.s.}$$

Now approximate G^m by step processes, as above. □

DEFINITION. If $G \in \mathbb{L}^2(0,T)$, take step processes G^n as above. Then

$$E\left(\left(\int_0^T G^n - G^m\,dW\right)^2\right) = E\left(\int_0^T (G^n - G^m)^2\,dt\right) \to 0 \quad \text{as } n, m \to \infty$$

and so the limit

$$\int_0^T G\,dW := \lim_{n\to\infty} \int_0^T G^n\,dW$$

exists in $L^2(\Omega)$.

It is not hard to check that this definition does not depend upon the particular sequence of step process approximations in $\mathbb{L}^2(0,T)$.

THEOREM (PROPERTIES OF ITÔ INTEGRAL). *For all constants $a, b \in \mathbb{R}$ and for all $G, H \in \mathbb{L}^2(0,T)$, we have*

$$(i) \qquad \int_0^T aG + bH \, dW = a \int_0^T G \, dW + b \int_0^T H \, dW,$$

$$(ii) \qquad E\left(\int_0^T G \, dW \right) = 0,$$

$$(iii) \qquad E\left(\left(\int_0^T G \, dW \right)^2 \right) = E\left(\int_0^T G^2 \, dt \right),$$

$$(iv) \qquad E\left(\int_0^T G \, dW \int_0^T H \, dW \right) = E\left(\int_0^T GH \, dt \right).$$

PROOF. 1. Assertion (i) follows at once from the corresponding linearity property for step processes.

Statements (ii) and (iii) are also easy consequences of the similar rules for step processes.

2. Finally, assertion (iv) results from (iii) and the identity $2ab = (a+b)^2 - a^2 - b^2$, and is left as an exercise. □

EXTENDING THE DEFINITION. For many applications, it is important to consider a wider class of integrands, instead of just $\mathbb{L}^2(0,T)$. To this end we define $\mathbb{M}^2(0,T)$ to be the space of all real–valued, progressively measurable processes $G(\cdot)$ such that

$$\int_0^T G^2 \, dt < \infty \quad \text{a.s.}$$

It is possible to extend the definition of the Itô integral to cover $G \in \mathbb{M}^2(0,T)$, although we will not do so in these notes. The idea is to find a sequence of step processes $G^n \in \mathbb{M}^2(0,T)$ such that

$$\int_0^T (G - G^n)^2 \, dt \to 0 \quad \text{a.s. as } n \to \infty.$$

It turns out that we can then define

$$\int_0^T G \, dW := \lim_{n \to \infty} \int_0^T G^n \, dW,$$

the expressions on the right converging in probability. See for instance Friedman [**F**] or Gihman–Skorohod [**G-S**] for details. □

More on Riemann sums. In particular, if $G \in \mathbb{M}^2(0,T)$ and $t \mapsto G(t,\omega)$ is continuous for a.e. ω, then

$$\sum_{k=0}^{m_n-1} G(t_k^n)(W(t_{k+1}^n) - W(t_k^n)) \to \int_0^T G\,dW$$

in probability, where $P^n = \{0 = t^n < \cdots < t_{m_n}^n = T\}$ is any sequence of partitions, with $|P^n| \to 0$. This confirms the consistency of Itô's integral with the earlier calculations involving Riemann sums, evaluated at $\tau_k^n = t_k^n$. □

C. INDEFINITE ITÔ INTEGRALS

DEFINITION. For $G \in \mathbb{L}^2(0,T)$, set

$$I(t) := \int_0^t G\,dW \quad (0 \le t \le T),$$

the *indefinite integral of* $G(\cdot)$. Note $I(0) = 0$.

In this section we note some properties of the process $I(\cdot)$, namely that it is a martingale and has continuous sample paths a.s. These facts will be quite useful for proving Itô's formula later in §D and in solving the stochastic differential equations in Chapter 5.

THEOREM. (i) *If $G \in \mathbb{L}^2(0,T)$, then the indefinite integral $I(\cdot)$ is a martingale.*

(ii) *Furthermore, $I(\cdot)$ has a version with continuous sample paths a.s.*

Henceforth when we refer to $I(\cdot)$, we will always mean this version. We will not prove assertion (i); a proof of (ii) is in Appendix C.

D. ITÔ'S FORMULA

DEFINITION. Suppose that $X(\cdot)$ is a real–valued stochastic process satisfying

$$X(r) = X(s) + \int_s^r F\,dt + \int_s^r G\,dW$$

for some $F \in \mathbb{L}^1(0,T)$, $G \in \mathbb{L}^2(0,T)$ and all times $0 \le s \le r \le T$. We say that $X(\cdot)$ has the *stochastic differential*

$$dX = F dt + G dW$$

for $0 \le t \le T$. □

Note carefully that the differential symbols are simply an abbreviation for the integral expressions above: strictly speaking "dX", "dt", and "dW" have no meaning alone.

THEOREM (Itô's Formula). *Suppose that $X(\cdot)$ has a stochastic differential*

$$dX = Fdt + GdW,$$

for $F \in \mathbb{L}^1(0,T), G \in \mathbb{L}^2(0,T)$. Assume $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ is continuous and that $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$, $\frac{\partial^2 u}{\partial x^2}$ exist and are continuous.
Set

$$Y(t) := u(X(t), t).$$

Then Y has the stochastic differential

$$\begin{aligned} dY &= \tfrac{\partial u}{\partial t}dt + \tfrac{\partial u}{\partial x}dX + \tfrac{1}{2}\tfrac{\partial^2 u}{\partial x^2}G^2 dt \\ &= \left(\tfrac{\partial u}{\partial t} + \tfrac{\partial u}{\partial x}F + \tfrac{1}{2}\tfrac{\partial^2 u}{\partial x^2}G^2\right)dt + \tfrac{\partial u}{\partial x}GdW. \end{aligned} \tag{2}$$

We call (2) Itô's formula or Itô's chain rule.

Remarks. (i) The argument of u, $\frac{\partial u}{\partial t}$, etc. above is $(X(t), t)$.
(ii) In view of our definitions, the expression (2) means for all $0 \le s \le r \le T$,

$$\begin{aligned} Y(r) - Y(s) &= u(X(r), r) - u(X(s), s) \\ &= \int_s^r \frac{\partial u}{\partial t}(X,t) + \frac{\partial u}{\partial x}(X,t)F + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(X,t)G^2\, dt \\ &\quad + \int_s^r \frac{\partial u}{\partial x}(X,t)G\, dW \quad \text{almost surely.} \end{aligned} \tag{3}$$

(iii) Since $X(t) = X(0) + \int_0^t F\, ds + \int_0^t G\, dW$, $X(\cdot)$ has continuous sample paths almost surely. Thus for almost every ω, the functions $t \mapsto \frac{\partial u}{\partial t}(X(t), t)$, $\frac{\partial u}{\partial x}(X(t), t)$, $\frac{\partial^2 u}{\partial x^2}(X(t), t)$ are continuous and so the integrals in (3) are defined. □

ILLUSTRATIONS OF ITÔ'S FORMULA. We will prove Itô's formula below, but first here are some applications:

Example 1. Let $X(\cdot) = W(\cdot)$, $u(x) = x^m$. Then $dX = dW$ and thus $F \equiv 0$, $G \equiv 1$. Hence Itô's formula gives

$$d(W^m) = mW^{m-1}dW + \frac{1}{2}m(m-1)W^{m-2}dt.$$

In particular the case $m = 2$ reads

$$d(W^2) = 2W\,dW + dt.$$

This integrated is the identity

$$\int_s^r W\, dW = \frac{W^2(r) - W^2(s)}{2} - \frac{(r-s)}{2},$$

a formula we have established from first principles before. □

EXAMPLE 2. Again take $X(\cdot) = W(\cdot)$, $u(x,t) = e^{\lambda x - \frac{\lambda^2 t}{2}}$, $F \equiv 0$, $G \equiv 1$. Then

$$d\left(e^{\lambda W(t) - \frac{\lambda^2 t}{2}}\right) = \left(-\frac{\lambda^2}{2} e^{\lambda W(t) - \frac{\lambda^2 t}{2}} + \frac{\lambda^2}{2} e^{\lambda W(t) - \frac{\lambda^2 t}{2}}\right) dt + \lambda e^{\lambda W(t) - \frac{\lambda^2 t}{2}} dW$$

by Itô's formula. Thus

$$\begin{cases} dY = \lambda Y dW \\ Y(0) = 1. \end{cases}$$

This is a *stochastic differential equation*, about which more in Chapters 5 and 6. □

In the Itô stochastic calculus the expression $e^{\lambda W(t) - \frac{\lambda^2 t}{2}}$ plays the role that $e^{\lambda t}$ plays in ordinary calculus. We build upon this observation:

EXAMPLE 3. For $n = 0, 1, \ldots$, define

$$h_n(x,t) := \frac{(-t)^n}{n!} e^{x^2/2t} \frac{d^n}{dx^n}\left(e^{-x^2/2t}\right),$$

the n-th *Hermite polynomial*. Then

$$\begin{aligned} h_0(x,t) &= 1,\ h_1(x,t) = x \\ h_2(x,t) &= \frac{x^2}{2} - \frac{t}{2},\ h_3(x,t) = \frac{x^3}{6} - \frac{tx}{2} \\ h_4(x,t) &= \frac{x^4}{24} - \frac{tx^2}{4} + \frac{t^2}{8},\ \text{etc.} \end{aligned}$$

THEOREM (STOCHASTIC CALCULUS WITH HERMITE POLYNOMIALS). *We have*

$$\int_0^t h_n(W,s)\, dW = h_{n+1}(W(t),t) \quad \textit{for } t \geq 0 \textit{and } n = 0, 1, \ldots;$$

that is,

$$dh_{n+1}(W,t) = h_n(W,t)\, dW.$$

Consequently in the Itô stochastic calculus the expression $h_n(W(t),t)$ plays the role that $\frac{t^n}{n!}$ plays in ordinary calculus.

PROOF. (from McKean [McK]) Since

$$\frac{d^n}{d\lambda^n}(e^{-\frac{(x-\lambda t)^2}{2t}})|_{\lambda=0} = (-t)^n \frac{d^n}{dx^n}(e^{-x^2/2t}),$$

we have

$$\begin{aligned} \frac{d^n}{d\lambda^n}(e^{\lambda x - \frac{\lambda^2 t}{2}})|_{\lambda=0} &= (-t)^n e^{x^2/2t} \frac{d^n}{dx^n}(e^{-x^2/2t}) \\ &= n! h_n(x,t). \end{aligned}$$

Hence

$$e^{\lambda x - \frac{\lambda^2 t}{2}} = \sum_{n=0}^{\infty} \lambda^n h_n(x,t),$$

and so

$$Y(t) = e^{\lambda W(t) - \frac{\lambda^2 t}{2}} = \sum_{n=0}^{\infty} \lambda^n h_n(W(t), t).$$

But $Y(\cdot)$ solves

$$\begin{cases} dY = \lambda Y dW \\ Y(0) = 1; \end{cases}$$

that is,

$$Y(t) = 1 + \lambda \int_0^t Y \, dW \quad \text{for all } t \geq 0.$$

Plug in the expansion above for $Y(t)$:

$$\begin{aligned} \sum_{n=0}^{\infty} \lambda^n h_n(W(t), t) &= 1 + \lambda \int_0^t \sum_{n=0}^{\infty} \lambda^n h_n(W(s), s) \, dW \\ &= 1 + \sum_{n=1}^{\infty} \lambda^n \int_0^t h_{n-1}(W(s), s) \, dW. \end{aligned}$$

This identity holds for all λ and so the coefficients of λ^n on both sides are equal. □

PROOF OF ITÔ'S FORMULA. We now begin the proof of Itô's formula, by verifying directly two important special cases:

LEMMA (Two simple stochastic differentials). *We have*
(i) $d(W^2) = 2W dW + dt$,
and
(ii) $d(tW) = W dt + t dW$.

Proof. We have already established formula (i). To verify (ii), note that

$$\int_0^r t \, dW = \lim_{n \to \infty} \sum_{k=0}^{m_n - 1} t_k^n (W(t_{k+1}^n) - W(t_k^n)),$$

where $P^n = \{0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = r\}$ is a sequence of partitions of $[0, r]$, with $|P^n| \to 0$. The limit above is taken in $L^2(\Omega)$.

Similarly, since $t \mapsto W(t)$ is continuous a.s.,

$$\int_0^r W \, dt = \lim_{n \to \infty} \sum_{k=0}^{m_n - 1} W(t_{k+1}^n)(t_{k+1}^n - t_k^n),$$

since for amost every ω the sum is an ordinary Riemann sum approximation and for this we can take the *right*-hand endpoint t_{k+1}^n at which to evaluate the continuous integrand.

We add these formulas to obtain

$$\int_0^r t \, dW + \int_0^r W \, dt = rW(r).$$

These integral identities for all $r \geq 0$ are abbreviated $d(tW) = tdW + Wdt$. $\square$

These special cases in hand, we now prove:

THEOREM (Itô product rule). *Suppose*

$$\begin{cases} dX_1 = F_1 dt + G_1 dW \\ dX_2 = F_2 dt + G_2 dW \end{cases} \qquad (0 \leq t \leq T),$$

for $F_i \in \mathbb{L}^1(0,T)$, $G_i \in \mathbb{L}^2(0,T)$ $(i = 1,2)$. *Then*

$$d(X_1 X_2) = X_2 dX_1 + X_1 dX_2 + G_1 G_2 dt. \tag{4}$$

Remarks. (i) The expression $G_1 G_2 dt$ here is the *Itô correction term*. The integrated version of the product rule is the *Itô integration-by-parts formula*:

$$\int_s^r X_2 \, dX_1 = X_1(r)X_2(r) - X_1(s)X_2(s) - \int_s^r X_1 \, dX_2 - \int_s^r G_1 G_2 \, dt. \tag{5}$$

(ii) If either G_1 or G_2 is identically equal to 0, we get the ordinary calculus integration-by-parts formula. This confirms that the Paley–Wiener–Zygmund definition

$$\int_0^1 g \, dW = -\int_0^1 g' W \, dt,$$

for deterministic C^1 functions g, with $g(0) = g(1) = 0$, agrees with the Itô definition. $\square$

Proof. 1. Choose $0 \leq r \leq T$.

First of all, assume for simplicity that $X_1(0) = X_2(0) = 0$, $F_i(t) \equiv F_i$, $G_i(t) \equiv G_i$, where F_i, G_i are time–independent, $\mathcal{F}(0)$-measurable random variables $(i = 1,2)$. Then

$$X_i(t) = F_i t + G_i W(t) \quad (t \geq 0, \ i = 1,2).$$

Thus

$$\begin{aligned}
&\int_0^r X_2\,dX_1 + X_1\,dX_2 + G_1G_2\,dt \\
&= \int_0^r X_1F_2 + X_2F_1\,dt + \int_0^r X_1G_2 + X_2G_1\,dW \\
&\qquad + \int_0^r G_1G_2\,dt \\
&= \int_0^r (F_1t + G_1W)F_2 + (F_2t + G_2W)F_1\,dt \\
&\qquad + \int_0^r (F_1t + G_1W)G_2 + (F_2t + G_2W)G_1\,dW + G_1G_2r \\
&= F_1F_2r^2 + (G_1F_2 + G_2F_1)\left[\int_0^r W\,dt + \int_0^r t\,dW\right] \\
&\qquad + 2G_1G_2\int_0^r W\,dW + G_1G_2r.
\end{aligned}$$

We now use the Lemma above to compute $2\int_0^r W\,dW = W^2(r) - r$ and $\int_0^r W dt + \int_0^r t\,dW = rW(r)$. Employing these identities, we deduce:

$$\begin{aligned}
&\int_0^r X_2\,dX_1 + X_1\,dX_2 + G_1G_2\,dt \\
&\qquad = F_1F_2r^2 + (G_1F_2 + G_2F_1)rW(r) + G_1G_2W^2(r) \\
&\qquad = X_1(r)X_2(r).
\end{aligned}$$

This is formula (5) for the special circumstance that $s = 0$, $X_i(0) = 0$, and F_i, G_i time–independent random variables.

The case that $s \geq 0$, $X_1(s), X_2(s)$ are arbitrary, and F_i, G_i are constant $\mathcal{F}(s)$-measurable random variables has a similar proof.

2. If F_i, G_i are step processes, we apply Step 1 on each subinterval $[t_k, t_{k+1})$ on which F_i and G_i are constant random variables, and add the resulting integral expressions.

3. In the general situation, we select step processes $F_i^n \in \mathbb{L}^1(0,T)$, $G_i^n \in \mathbb{L}^2(0,T)$, with

$$\begin{aligned}
&E(\textstyle\int_0^T |F_i^n - F_i|\,dt) \to 0 \\
&E(\textstyle\int_0^T (G_i^n - G_i)^2\,dt) \to 0
\end{aligned} \qquad \text{as } n \to \infty,\ i = 1, 2.$$

Define

$$X_i^n(t) := X_i(0) + \int_0^t F_i^n\,ds + \int_0^t G_i^n\,dW \quad (i = 1, 2).$$

We apply Step 2 to $X_i^n(\cdot)$ on (s, r) and pass to limits, to obtain the formula

$$X_1(r)X_2(r) = X_1(s)X_2(s) + \int_s^r X_1dX_2 + X_2dX_1 + G_1G_2\,dt.$$

□

Now we are ready for

CONCLUSION OF THE PROOF OF ITÔ'S FORMULA. Suppose $dX = Fdt + GdW$, with $F \in \mathbb{L}^1(0,T)$, $G \in \mathbb{L}^2(0,T)$.

1. We start with the case $u(x) = x^m$, $m = 0, 1, \ldots$, and first of all claim that

$$d(X^m) = mX^{m-1}dX + \frac{1}{2}m(m-1)X^{m-2}G^2dt. \tag{6}$$

This is clear for $m = 0, 1$, and the case $m = 2$ follows from the Itô product formula. Now assume the stated formula for $m-1$:

$$\begin{aligned} d(X^{m-1}) &= (m-1)X^{m-2}dX + \frac{1}{2}(m-1)(m-2)X^{m-3}G^2dt \\ &= (m-1)X^{m-2}(Fdt + GdW) + \frac{1}{2}(m-1)(m-2)X^{m-3}G^2dt, \end{aligned}$$

and we prove it for m:

$$\begin{aligned} d(X^m) &= d(XX^{m-1}) \\ &= Xd(X^{m-1}) + X^{m-1}dX + (m-1)X^{m-2}G^2dt \\ &\qquad\qquad \text{(by the product rule)} \\ &= X\left((m-1)X^{m-2}dX + \frac{1}{2}(m-1)(m-2)X^{m-3}G^2dt\right) \\ &\quad + (m-1)X^{m-2}G^2dt + X^{m-1}dX \\ &= mX^{m-1}dX + \frac{1}{2}m(m-1)X^{m-2}G^2dt, \end{aligned}$$

because $m - 1 + \frac{1}{2}(m-1)(m-2) = \frac{1}{2}m(m-1)$. This proves (6).

Since Itô's formula thus holds for the functions $u(x) = x^m$, $m = 0, 1, \ldots$ and since the operator "d" is linear, Itô's formula is valid for all polynomials u in the variable x.

2. Suppose now $u(x,t) = f(x)g(t)$, where f and g are polynomials. Then

$$\begin{aligned} d(u(X,t)) &= d(f(X)g) \\ &= f(X)dg + gdf(X) \\ &= f(X)g'dt + g[f'(X)dX + \frac{1}{2}f''(X)G^2dt] \\ &= \frac{\partial u}{\partial t}dt + \frac{\partial u}{\partial x}dX + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}G^2dt. \end{aligned}$$

This calculation confirms Itô's formula for $u(x,t) = f(x)g(t)$, where f and g are polynomials. Thus it is true as well for any function u having the form

$$u(x,t) = \sum_{i=1}^{m} f^i(x)g^i(t),$$

where f^i and g^i polynomials. That is, Itô's formula is valid for all polynomial functions u of the variables x, t.

3. Given u as in Itô's formula, there exists a sequence of polynomials u^n such that

$$u^n \to u, \ \frac{\partial u^n}{\partial t} \to \frac{\partial u}{\partial t}$$
$$\frac{\partial u^n}{\partial x} \to \frac{\partial u}{\partial x}, \ \frac{\partial^2 u^n}{\partial x^2} \to \frac{\partial^2 u}{\partial x^2},$$

uniformly on compact subsets of $\mathbb{R} \times [0,T]$. Invoking Step 2, we know that for all $0 \le r \le T$,

$$u^n(X(r),r) - u^n(X(0),0) = \int_0^r \frac{\partial u^n}{\partial t} + \frac{\partial u^n}{\partial x}F + \frac{1}{2}\frac{\partial^2 u^n}{\partial x^2}G^2\,dt + \int_0^r \frac{\partial u^n}{\partial x}G\,dW \text{ almost surely;}$$

the argument of the partial derivatives of u^n is $(X(t),t)$.

We may pass to limits as $n \to \infty$ in this expression, thereby proving Itô's formula in general. □

A similar proof gives this:

GENERALIZED ITÔ FORMULA. Suppose $dX^i = F^i dt + G^i dW$, with for $F^i \in \mathbb{L}^1(0,T)$, $G^i \in \mathbb{L}^2(0,T)$, for $i = 1, \dots, n$.

If $u : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ is continuous, with continuous partial derivatives $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x_i}$, $\frac{\partial^2 u}{\partial x_i \partial x_j}$, $(i,j = 1, \dots, n)$, then

$$d(u(X^1, \dots, X^n, t)) = \frac{\partial u}{\partial t}dt + \sum_{i=1}^n \frac{\partial u}{\partial x_i}dX^i + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 u}{\partial x_i \partial x_j}G^i G^j dt.$$

□

E. ITÔ'S INTEGRAL IN HIGHER DIMENSIONS

Notation. (i) Let $\mathbf{W}(\cdot) = (W^1(\cdot), \dots, W^m(\cdot))$ be an m-dimensional Brownian motion.

(ii) We assume $\mathcal{F}(\cdot)$ is a family of nonanticipating σ-algebras, meaning that

(a) $\mathcal{F}(t) \supseteq \mathcal{F}(s)$ for all $t \ge s \ge 0$

(b) $\mathcal{F}(t) \supseteq \mathcal{W}(t) = \mathcal{U}(\mathbf{W}(s) \,|\, 0 \le s \le t)$

(c) $\mathcal{F}(t)$ is independent of $\mathcal{W}^+(t) := \mathcal{U}(\mathbf{W}(s) - \mathbf{W}(t) \,|\, t \le s < \infty)$.

DEFINITIONS. (i) An $\mathbb{M}^{n\times m}$-valued stochastic process $\mathbf{G} = ((G^{ij}))$ belongs to $\mathbb{L}^2_{n\times m}(0,T)$ if

$$G^{ij} \in \mathbb{L}^2(0,T) \qquad (i = 1, \dots n; \ j = 1, \dots m).$$

(ii) An $\mathbb{R}^n$-valued stochastic process $\mathbf{F} = (F^1, F^2, \dots, F^n)$ belongs to $\mathbb{L}^1_n(0,T)$ if

$$F^i \in \mathbb{L}^1(0,T) \qquad (i = 1, \dots n).$$

DEFINITION. If $\mathbf{G} \in \mathbb{L}^2_{n\times m}(0,T)$, then

$$\int_0^T \mathbf{G}\, d\mathbf{W}$$

is an $\mathbb{R}^n$–valued random variable, whose i-th component is

$$\sum_{j=1}^m \int_0^T G^{ij}\, dW^j \quad (i = 1, \ldots, n).$$

□

Approximating by step processes as before, we can establish this

LEMMA. *If* $\mathbf{G} \in \mathbb{L}^2_{n\times m}(0,T)$, *then*

$$E\left(\int_0^T \mathbf{G}\, d\mathbf{W}\right) = 0,$$

and

$$E\left(\left|\int_0^T \mathbf{G}\, d\mathbf{W}\right|^2\right) = E\left(\int_0^T |\mathbf{G}|^2\, dt\right),$$

where $|\mathbf{G}|^2 := \sum_{\substack{1\le i\le n\\ 1\le j\le m}} |G^{ij}|^2$.

DEFINITION. If $\mathbf{X}(\cdot) = (X^1(\cdot), \ldots, X^n(\cdot))$ is an $\mathbb{R}^n$-valued stochastic process such that

$$\mathbf{X}(r) = \mathbf{X}(s) + \int_s^r \mathbf{F}\, dt + \int_s^r \mathbf{G}\, d\mathbf{W}$$

for some $\mathbf{F} \in \mathbb{L}^1_n(0,T)$, $\mathbf{G} \in \mathbb{L}^2_{n\times m}(0,T)$ and all $0 \le s \le r \le T$, we say $\mathbf{X}(\cdot)$ has the *stochastic differential*

$$d\mathbf{X} = \mathbf{F}dt + \mathbf{G}d\mathbf{W}.$$

This means that

$$dX^i = F^i dt + \sum_{j=1}^m G^{ij} dW^j \quad \text{for } i = 1, \ldots, n.$$

THEOREM (Itô's formula in **n**-dimensions). *Suppose that* $d\mathbf{X} = \mathbf{F}dt + \mathbf{G}d\mathbf{W}$, *as above. Let* $u : \mathbb{R}^n \times [0,T]$ *be continuous, with continuous partial derivatives* $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x_i}$, $\frac{\partial^2 u}{\partial x_i \partial x_j}$, $(i,j = 1, \ldots, n)$. *Then*

$$\begin{aligned} d(u(\mathbf{X}(t),t)) = {} & \tfrac{\partial u}{\partial t} dt + \textstyle\sum_{i=1}^n \tfrac{\partial u}{\partial x_i} dX^i \\ & + \tfrac{1}{2} \textstyle\sum_{i,j=1}^n \tfrac{\partial^2 u}{\partial x_i \partial x_j} \sum_{l=1}^m G^{il} G^{jl} dt, \end{aligned} \tag{5}$$

where the argument of the partial derivatives of u *is* $(\mathbf{X}(t), t)$.

An outline of the proof follows some preliminary results:

LEMMA (ANOTHER SIMPLE STOCHASTIC DIFFERENTIAL). *Let $W(\cdot)$ and $\bar{W}(\cdot)$ be* independent 1-*dimensional Brownian motions. Then*

$$d(W\bar{W}) = WdW + \bar{W}dW.$$

Compare this to the case $W = \bar{W}$. There is no correction term "dt" here, since $W, \bar{W}$ are independent.

PROOF. 1. To begin, set $X(t) := \frac{W(t)+\bar{W}(t)}{\sqrt{2}}$.

We claim that $X(\cdot)$ is a 1-dimensional Brownian motion. To see this, note firstly that $X(0) = 0$ a.s. and $X(\cdot)$ has independent increments. Next observe that since X is the sum of two independent, $N(0, \frac{t}{2})$ random variables, $X(t)$ is $N(0, t)$. A similar observation shows that $X(t) - X(s)$ is $N(0, t-s)$ for $t \geq s$. This establishes the claim.

2. From the 1-dimensional Itô calculus, we know

$$\begin{cases} d(X^2) &= 2XdX + dt, \\ d(W^2) &= 2WdW + dt, \\ d(\bar{W}^2) &= 2\bar{W}d\bar{W} + dt. \end{cases}$$

Thus

$$\begin{aligned} d(W\bar{W}) &= d\left(X^2 - \frac{1}{2}W^2 - \frac{1}{2}\bar{W}^2\right) \\ &= 2XdX + dt - \frac{1}{2}(2WdW + dt) \\ &\qquad - \frac{1}{2}(2\bar{W}d\bar{W} + dt) \\ &= (W + \bar{W})(dW + d\bar{W}) - WdW - \bar{W}d\bar{W} \\ &= WdW + \bar{W}dW. \end{aligned}$$

□

We will also need the following modification of the product rule:

LEMMA (ITÔ PRODUCT RULE WITH SEVERAL BROWNIAN MOTIONS). *Suppose*

$$dX_1 = F_1 dt + \sum_{k=1}^{m} G_1^k dW^k$$

and

$$dX_2 = F_2 dt + \sum_{l=1}^{m} G_2^l dW^l,$$

where $F_i \in \mathbb{L}^1(0,T)$ and $G_i^k \in \mathbb{L}^2(0,T)$ for $i = 1, 2;\ k = 1, \ldots, m$. Then

$$d(X_1X_2) = X_1dX_2 + X_2dX_1 + \sum_{k=1}^{m} G_1^k G_2^k dt.$$

The proof is a modification of that for the one–dimensional Itô product rule, as before, with the new feature that

$$d(W^iW^j) = W^i dW^j + W^j dW^i + \delta_{ij}dt,$$

according to the Lemma above.

The Itô formula in n-dimensions can now be proved by a suitable modification of the one-dimensional proof. We first establish the formula for a multinomials $u = u(x) = x_1^{k_1} \dots x_m^{k_m}$, proving this by an induction on $k_1, \dots, k_m$, using the Lemma above. This done, the formula follows easily for polynomials $u = u(x,t)$ in the variables $x = (x_1, \dots, x_n)$ and t, and then, after an approximation, for all functions u as stated.

CLOSING REMARKS.

1. ALTERNATIVE NOTATION. When

$$d\mathbf{X} = \mathbf{F}dt + \mathbf{G}d\mathbf{W},$$

we sometimes write

$$H^{ij} := \sum_{k=1}^{m} G^{ik}G^{jk}.$$

Then Itô's formula reads

$$du(\mathbf{X},t) = \left(\frac{\partial u}{\partial t} + F \cdot Du + \frac{1}{2}H : D^2u\right)dt + Du \cdot \mathbf{G}d\mathbf{W},$$

where $Du = \left(\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_n}\right)$ is the gradient of u in the x-variables, $D^2u = \left(\left(\frac{\partial^2 u}{\partial x_i \partial x_j}\right)\right)$ is the Hessian matrix, and

$$F \cdot Du = \sum_{i=1}^{n} F^i \frac{\partial u}{\partial x_i},$$

$$H : D^2u = \sum_{i,j=1}^{n} H^{ij} \frac{\partial^2 u}{\partial x_i \partial x_j},$$

$$Du \cdot \mathbf{G}d\mathbf{W} = \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\partial u}{\partial x_i} G^{ik} dW^k.$$

2. HOW TO REMEMBER ITÔ'S FORMULA.
We may *symbolically* compute

$$d(u(\mathbf{X},t)) = \frac{\partial u}{\partial t}dt + \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} dX^i + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 u}{\partial x_i \partial x_j} dX^i dX^j,$$

and then simplify the term "$dX^i dX^j$" by expanding it out and using the *formal* multiplication rules

$$(dt)^2 = 0,\ dtdW^k = 0,\ dW^k dW^l = \delta_{kl}dt \quad (k,l = 1, \dots, m).$$

The foregoing theory provides a rigorous meaning for all this.

CHAPTER 5: STOCHASTIC DIFFERENTIAL EQUATIONS

A. Definitions and examples
B. Existence and uniqueness of solutions
C. Properties of solutions
D. Linear stochastic differential equations

A. DEFINITIONS AND EXAMPLES

We are finally ready to study stochastic differential equations:

Notation. (i) Let $\mathbf{W}(\cdot)$ be an m-dimensional Brownian motion and $\mathbf{X}_0$ an n-dimensional random variable which is independent of $\mathbf{W}(\cdot)$. We will henceforth take

$$\mathcal{F}(t) := \mathcal{U}(\mathbf{X}_0, \mathbf{W}(s)\ (0 \le s \le t)) \quad (t \ge 0),$$

the σ-algebra generated by $\mathbf{X}_0$ and the history of the Wiener process up to (and including) time t.

(ii) Assume $T > 0$ is given, and

$$\mathbf{b} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n,$$

$$\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n \times m}$$

are given functions. (Note carefully: these are *not* random variables.) We display the components of these functions by writing

$$\mathbf{b} = (b^1, b^2, \dots, b^n), \quad \mathbf{B} = \begin{pmatrix} b^{11} & \dots & b^{1m} \\ \vdots & \ddots & \vdots \\ b^{n1} & \dots & b^{nm} \end{pmatrix}.$$

DEFINITION. We say that an $\mathbb{R}^n$-valued stochastic process $\mathbf{X}(\cdot)$ is a *solution* of the *Itô stochastic differential equation*

$$(SDE) \qquad \begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases}$$

for $0 \le t \le T$, provided

(i) $\mathbf{X}(\cdot)$ is progressively measurable with respect to $\mathcal{F}(\cdot)$,

(ii) $\mathbf{F} := \mathbf{b}(\mathbf{X}, t) \in \mathbb{L}^1_n(0,T)$,

(iii) $\mathbf{G} := \mathbf{B}(\mathbf{X}, t) \in \mathbb{L}^2_{n \times m}(0,T)$,

and

(iv) $\mathbf{X}(t) = \mathbf{X}_0 + \int_0^t \mathbf{b}(\mathbf{X}(s), s)\, ds + \int_0^t \mathbf{B}(\mathbf{X}(s), s)\, d\mathbf{W}$ a.s. for all $0 \le t \le T$.

Remarks. (i) A higher order SDE of the form

$$Y^{(n)} = f(t, Y, \dots, Y^{(n-1)}) + g(t, Y, \dots, Y^{(n-1)})\xi,$$

where as usual ξ denotes "white noise", can be rewritten into the form above by the device of setting

$$\mathbf{X}(t) = \begin{pmatrix} Y(t) \\ \dot{Y}(t) \\ \vdots \\ Y^{(n-1)}(t) \end{pmatrix} = \begin{pmatrix} X^1(t) \\ X^2(t) \\ \vdots \\ X^n(t) \end{pmatrix}.$$

Then

$$d\mathbf{X} = \begin{pmatrix} X^2 \\ \vdots \\ f(\cdots) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \vdots \\ g(\cdots) \end{pmatrix} d\mathbf{W}.$$

(ii) In view of (iii), we can always assume $\mathbf{X}(\cdot)$ has continuous sample paths almost surely. □

EXAMPLES OF LINEAR STOCHASTIC DIFFERENTIAL EQUATIONS.

EXAMPLE 1. Let $m = n = 1$ and suppose g is a continuous function (not a random variable). Then the unique solution of

$$\begin{cases} dX = gX\,dW \\ X(0) = 1 \end{cases} \tag{1}$$

is

$$X(t) = e^{-\frac{1}{2}\int_0^t g^2\,ds + \int_0^t g\,dW}$$

for $0 \le t \le T$. To verify this, note that

$$Y(t) := -\frac{1}{2}\int_0^t g^2\,ds + \int_0^t g\,dW$$

satisfies

$$dY = -\frac{1}{2}g^2dt + gdW.$$

Thus Itô's lemma for $u(x) = e^x$ gives

$$\begin{aligned} dX &= \frac{\partial u}{\partial x}dY + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}g^2dt \\ &= e^Y\left(-\frac{1}{2}g^2dt + gdW + \frac{1}{2}g^2dt\right) \\ &= gX\,dW, \text{ as claimed.} \end{aligned}$$

We will prove uniqueness later, in §B. □

EXAMPLE 2. Similarly, the unique solution of

$$\begin{cases} dX = fX\,dt + gX\,dW \\ X(0) = 1 \end{cases} \tag{2}$$

is

$$X(t) = e^{\int_0^t f - \frac{1}{2}g^2\,ds + \int_0^t g\,dW}$$

for $0 \le t \le T$. □

EXAMPLE 3 (STOCK PRICES). Let $P(t)$ denote the price of a stock at time t. We can model the evolution of $P(t)$ in time by supposing that $\frac{dP}{P}$, the relative change of price, evolves according to the SDE

$$\frac{dP}{P} = \mu dt + \sigma dW$$

for certain constants $\mu > 0$ and σ, called the *drift* and the *volatility* of the stock. Hence

$$dP = \mu P dt + \sigma P dW; \tag{3}$$

and so

$$\begin{aligned} d(\log(P)) &= \frac{dP}{P} - \frac{1}{2}\frac{\sigma^2 P^2 dt}{P^2} \quad \text{by Itô's formula} \\ &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW. \end{aligned}$$

Consequently

$$P(t) = p_0 e^{\sigma W(t) + \left(\mu - \frac{\sigma^2}{2}\right)t},$$

similarly to Example 2. Observe that the price is always positive, assuming the initial price p_0 is positive.

Since (3) implies

$$P(t) = p_0 + \int_0^t \mu P\,ds + \int_0^t \sigma P\,dW$$

and $E\left(\int_0^t \sigma P\,dW\right) = 0$, we see that

$$E(P(t)) = p_0 + \int_0^t \mu E(P(s))\,ds.$$

Hence

$$E(P(t)) = p_0 e^{\mu t} \quad \text{for } t \ge 0.$$

The expected value of the stock price consequently agrees with the deterministic solution of (3) corresponding to $\sigma = 0$. □

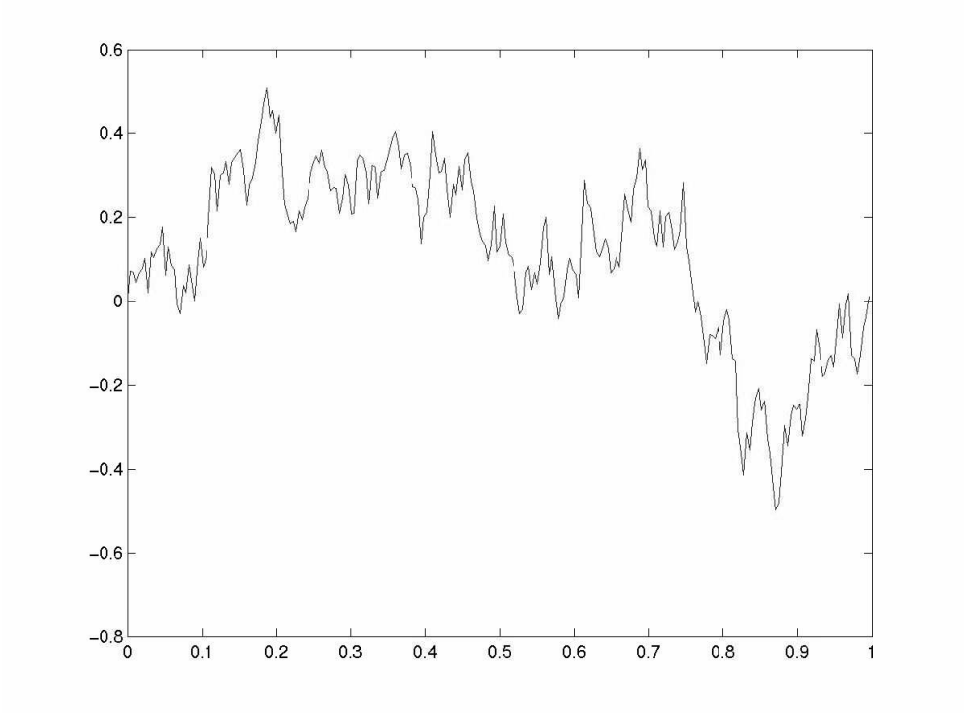
EXAMPLE 4 (BROWNIAN BRIDGE). The solution of the SDE

$$\begin{cases} dB = -\frac{B}{1-t}dt + dW & (0 \le t < 1) \\ B(0) = 0 \end{cases} \tag{4}$$

is

$$B(t) = (1-t)\int_0^t \frac{1}{1-s}\, dW \qquad (0 \le t < 1),$$

as we confirm by a direct calculation. It turns out also that $\lim_{t \to 1-} B(t) = 0$ almost surely. We call $B(\cdot)$ a *Brownian bridge*, between the origin at time 0 and at time 1.



A SAMPLE PATH OF THE BROWNIAN BRIDGE

EXAMPLE 5 (LANGEVIN'S EQUATION). A possible improvement of our mathematical model of the motion of a Brownian particle models *frictional forces* as follows for the one-dimensional case:

$$\dot{X} = -bX + \sigma\xi,$$

where $\xi(\cdot)$ is "white noise", $b > 0$ is a coefficient of friction, and σ is a diffusion coefficient. In this interpretation $X(\cdot)$ is the *velocity* of the Brownian particle: see Example 6 for the *position* process $Y(\cdot)$. We interpret this to mean

$$\begin{cases} dX = -bX\,dt + \sigma dW \\ X(0) = X_0, \end{cases} \tag{5}$$

for some initial distribution X_0, independent of the Brownian motion. This is the *Langevin equation.*

The solution is

$$X(t) = e^{-bt}X_0 + \sigma \int_0^t e^{-b(t-s)}dW \quad (t \geq 0),$$

as is straightforward to verify. Observe that

$$E(X(t)) = e^{-bt}E(X_0)$$

and

$$\begin{aligned} E(X^2(t)) &= E\left(e^{-2bt}X_0^2 + 2\sigma e^{-bt}X_0 \int_0^t e^{-b(t-s)}dW \right. \\ &\qquad \left. + \sigma^2 \left(\int_0^t e^{-b(t-s)}dW\right)^2\right) \\ &= e^{-2bt}E(X_0^2) + 2\sigma e^{-bt}E(X_0)E\left(\int_0^t e^{-b(t-s)}dW\right) \\ &\qquad + \sigma^2 \int_0^t e^{-2b(t-s)}ds \\ &= e^{-2bt}E(X_0^2) + \frac{\sigma^2}{2b}(1 - e^{-2bt}). \end{aligned}$$

Thus the variance

$$V(X(t)) = E(X^2(t)) - E(X(t))^2$$

is given by

$$V(X(t)) = e^{-2bt}V(X_0) + \frac{\sigma^2}{2b}(1 - e^{-2bt}),$$

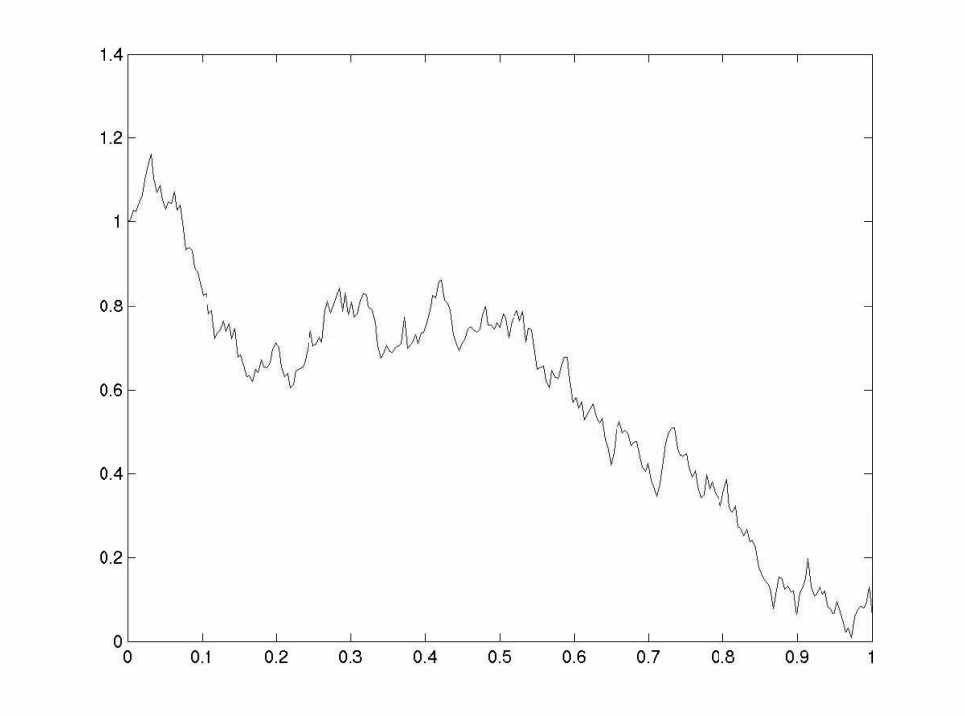
assuming, of course, $V(X_0) < \infty$. For any such initial condition X_0 we therefore have

$$\begin{cases} E(X(t)) \to 0 \\ V(X(t)) \to \frac{\sigma^2}{2b} \end{cases} \qquad \text{as } t \to \infty.$$

From the explicit form of the solution we see that the distribution of $X(t)$ approaches $N\left(0, \frac{\sigma^2}{2b}\right)$ as $t \to \infty$. We interpret this to mean that irrespective of the initial distribution, the solution of the SDE for large time "settles down" into a Gaussian distribution whose variance $\frac{\sigma^2}{2b}$ represents a balance between the random disturbing force $\sigma\xi(\cdot)$ and the frictional damping force $-bX(\cdot)$. □

Example 6 (Ornstein–Uhlenbeck process). A better model of Brownian movement is provided by the Ornstein–Uhlenbeck equation

$$\begin{cases} \ddot{Y} = -b\dot{Y} + \sigma\xi \\ Y(0) = Y_0,\ \dot{Y}(0) = Y_1, \end{cases}$$



A simulation of Langevin's equation

where $Y(t)$ is the *position* of Brownian particle at time t, Y_0 and Y_1 are given Gaussian random variables. As before $b > 0$ is the friction coefficient, σ is the diffusion coefficient, and $\xi(\cdot)$ as usual is "white noise".

Then $X := \dot{Y}$, the *velocity* process, satisfies the Langevin equation

$$\begin{cases} dX = -bX\,dt + \sigma\,dW \\ X(0) = Y_1, \end{cases} \tag{6}$$

studied in Example 5. We assume Y_1 to be normal, whence explicit formula for the solution,

$$X(t) = e^{-bt}Y_1 + \sigma \int_0^t e^{-b(t-s)}\,dW,$$

shows $X(t)$ to be Gaussian for all times $t \geq 0$. Now the *position process* is

$$Y(t) = Y_0 + \int_0^t X\,ds.$$

Therefore

$$\begin{aligned} E(Y(t)) &= E(Y_0) + \int_0^t E(X(s))\,ds \\ &= E(Y_0) + \int_0^t e^{-bs}E(Y_1)\,ds \\ &= E(Y_0) + \left(\frac{1-e^{-bt}}{b}\right)E(Y_1); \end{aligned}$$

and a somewhat lengthly calculation shows

$$V(Y(t)) = V(Y_0) + \frac{\sigma^2}{b^2}t + \frac{\sigma^2}{2b^3}(-3 + 4e^{-bt} - e^{-2bt}).$$

Nelson [**N, p. 57**] discusses this model as compared with Einstein's . □

EXAMPLE 7 (RANDOM HARMONIC OSCILLATOR). This is the SDE

$$\begin{cases} \ddot{X} = -\lambda^2 X - b\dot{X} + \sigma\xi \\ X(0) = X_0, \dot{X}(0) = X_1, \end{cases}$$

where $-\lambda^2 X$ represents a linear, restoring force and $-b\dot{X}$ is a frictional damping term.

An explicit solution can be worked out using the general formulas presented below in §D. For the special case $X_1 = 0$, $b = 0$, $\sigma = 1$, we have

$$X(t) = X_0 \cos(\lambda t) + \frac{1}{\lambda}\int_0^t \sin(\lambda(t-s))\, dW.$$

□

B. EXISTENCE AND UNIQUENESS OF SOLUTIONS

In this section we address the problem of building solutions to stochastic differential equations. We start with a simple case:

1. AN EXAMPLE IN ONE DIMENSION. Let us first suppose $b : \mathbb{R} \to \mathbb{R}$ is C^1, with $|b'| \leq L$ for some constant L, and try to solve the one–dimensional stochastic differential equation

$$\begin{cases} dX = b(X)dt + dW \\ X(0) = x \end{cases} \tag{7}$$

where $x \in \mathbb{R}$.

Now the SDE means

$$X(t) = x + \int_0^t b(X)\, ds + W(t),$$

for all times $t \geq 0$, and this formulation suggests that we try a *successive approximation method* to construct a solution. So define $X^0(t) \equiv x$, and then

$$X^{n+1}(t) := x + \int_0^t b(X^n)\ ds + W(t) \qquad (t \geq 0)$$

for $n = 0, 1, \ldots$. Next write

$$D^n(t) := \max_{0 \leq s \leq t} |\, X^{n+1}(s) - X^n(s)| \qquad (n = 0, \ldots),$$

and notice that for a given continuous sample path of the Brownian motion, we have

$$D^0(t) = \max_{0 \leq s \leq t} \left| \int_0^s b(x)\, dr + W(s) \right| \leq C$$

for all times $0 \le t \le T$, where C depends on ω.

We now *claim* that

$$D^n(t) \le C\frac{L^n}{n!}t^n$$

for $n = 0, 1, \ldots,\ 0 \le t \le T$. To see this note that

$$\begin{aligned} D^n(t) &= \max_{0\le s\le t}\left|\int_0^s b(X^n(r)) - b(X^{n-1}(r))\,dr\right| \\ &\le L\int_0^t D^{n-1}(s)\,ds \\ &\le L\int_0^t C\frac{L^{n-1}s^{n-1}}{(n-1)!}\,ds \quad \text{by the induction assumption} \\ &= C\frac{L^n t^n}{n!}. \end{aligned}$$

In view of the claim, for $m \ge n$ we have

$$\max_{0\le t\le T}|X^m(t) - X^n(t)| \le C\sum_{k=n}^{\infty}\frac{L^k T^k}{k!} \to 0 \quad \text{as } n \to \infty.$$

Thus for almost every ω, $X^n(\cdot)$ converges uniformly for $0 \le t \le T$ to a limit process $X(\cdot)$ which, as is easy to check, solves (7). □

2. SOLVING SDE BY CHANGING VARIABLES. Next is a procedure for solving SDE by means of a clever change of variables (McKean [**McK, p. 60**]).

Given a general one–dimensional SDE of the form

$$\begin{cases} dX = b(X)dt + \sigma(X)dW \\ X(0) = x, \end{cases} \tag{8}$$

let us first solve

$$\begin{cases} dY = f(Y)dt + dW \\ Y(0) = y, \end{cases} \tag{9}$$

where f will be selected later, and try to find a function u such that

$$X := u(Y)$$

solves our SDE (8). Note that we can in principle at least solve (9), according to the previous example. Assuming for the moment u and f are known, we compute using Itô's formula that

$$\begin{aligned} dX &= u'(Y)dY + \frac{1}{2}u''(Y)dt \\ &= \left[u'f + \frac{1}{2}u''\right]dt + u'dW. \end{aligned}$$

Thus $X(\cdot)$ solves (8) provided

$$\begin{cases} u'(Y) = \sigma(X) = \sigma(u(Y)), \\ u'(Y)f(Y) + \frac{1}{2}u''(Y) = b(X) = b(u(Y)), \end{cases}$$

and

$$u(y) = x.$$

So let us first solve the ODE

$$\begin{cases} u'(z) = \sigma(u(z)) \\ u(y) = x \end{cases} \qquad (z \in \mathbb{R}),$$

where $' = \frac{d}{dz}$, and then, once u is known, solve for

$$f(z) = \frac{1}{\sigma(u(z))}\left[b(u(z)) - \frac{1}{2}u''(z)\right].$$

We will not discuss here conditions under which all of this is possible: see Lamperti [**L2**]. □

Notice that both of the methods described above avoid all use of martingale estimates.

3. A GENERAL EXISTENCE AND UNIQUENESS THEOREM

We start with a useful calculus lemma:

GRONWALL'S LEMMA. *Let ϕ and f be nonnegative, continuous functions defined for $0 \le t \le T$, and let $C_0 \ge 0$ denote a constant. If*

$$\phi(t) \le C_0 + \int_0^t f\phi\, ds \quad \text{for all } 0 \le t \le T,$$

then

$$\phi(t) \le C_0 e^{\int_0^t f\, ds} \quad \text{for all } 0 \le t \le T.$$

PROOF. Set $\Phi(t) := C_0 + \int_0^t f\phi\, ds$. Then $\Phi' = f\phi \le f\Phi$, and so

$$\left(e^{-\int_0^t f\, ds}\Phi\right)' = (\Phi' - f\Phi)e^{-\int_0^t f\, ds} \le (f\phi - f\phi)e^{-\int_0^t f\, ds} = 0.$$

Therefore

$$\Phi(t)e^{-\int_0^t f\, ds} \le \Phi(0)e^{-\int_0^0 f\, ds} = C_0,$$

and thus

$$\phi(t) \le \Phi(t) \le C_0 e^{\int_0^t f\, ds}.$$

□

EXISTENCE AND UNIQUENESS THEOREM. *Suppose that* $\mathbf{b} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ *and* $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{m\times n}$ *are continuous and satisfy the following conditions:*

$$(a) \qquad \begin{array}{l} |\mathbf{b}(x,t) - \mathbf{b}(\hat{x},t)| \le L|x-\hat{x}| \\ |\mathbf{B}(x,t) - \mathbf{B}(\hat{x},t)| \le L|x-\hat{x}| \end{array} \quad \text{for all } 0 \le t \le T,\ x, \hat{x} \in \mathbb{R}^n$$

$$(b) \qquad \begin{array}{l} |\mathbf{b}(x,t)| \le L(1+|x|) \\ |\mathbf{B}(x,t)| \le L(1+|x|) \end{array} \quad \text{for all } 0 \le t \le T,\ x \in \mathbb{R}^n,$$

for some constant L.

Let $\mathbf{X}_0$ *be any* $\mathbb{R}^n$*-valued random variable such that*

$$(c) \qquad E(|\mathbf{X}_0|^2) < \infty$$

and

$$(d) \qquad \mathbf{X}_0 \text{ is independent of } \mathcal{W}^+(0),$$

where $\mathbf{W}(\cdot)$ *is a given* m*-dimensional Brownian motion.*

Then there exists a unique solution $\mathbf{X} \in \mathbb{L}^2_n(0,T)$ *of the stochastic differential equation:*

$$\text{(SDE)} \qquad \begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W} \quad (0 \le t \le T) \\ \mathbf{X}(0) = \mathbf{X}_0. \end{cases}$$

Remarks. (i) "Unique" means that if $\mathbf{X}, \hat{\mathbf{X}} \in \mathbb{L}^2_n(0,T)$, with continuous sample paths almost surely, and both solve (SDE), then

$$P(\mathbf{X}(t) = \hat{\mathbf{X}}(t) \text{ for all } 0 \le t \le T) = 1.$$

(ii) Hypotheses (a) says that $\mathbf{b}$ and $\mathbf{B}$ are *uniformly Lipschitz continuous* in the variable x. Notice also that hypothesis (b) actually follows from (a).

□

PROOF. 1. *Uniqueness.* Suppose $\mathbf{X}$ and $\hat{\mathbf{X}}$ are solutions, as above. Then for all $0 \le t \le T$,

$$\mathbf{X}(t) - \hat{\mathbf{X}}(t) = \int_0^t \mathbf{b}(\mathbf{X},s) - \mathbf{b}(\hat{\mathbf{X}},s)\,ds + \int_0^t \mathbf{B}(\mathbf{X},s) - \mathbf{B}(\hat{\mathbf{X}},s)\,d\mathbf{W}.$$

Since $(a+b)^2 \le 2a^2 + 2b^2$, we can estimate

$$\begin{aligned} E(|\mathbf{X}(t) - \hat{\mathbf{X}}(t)|^2) \le\ & 2E\left(\left|\int_0^t \mathbf{b}(\mathbf{X},s) - \mathbf{b}(\hat{\mathbf{X}},s)\,ds\right|^2\right) \\ & + 2E\left(\left|\int_0^t \mathbf{B}(\mathbf{X},s) - \mathbf{B}(\hat{\mathbf{X}},s)\,d\mathbf{W}\right|^2\right). \end{aligned}$$

The Cauchy–Schwarz inequality implies that

$$\left|\int_0^t \mathbf{f}\,ds\right|^2 \le t\int_0^t |\mathbf{f}|^2\,ds$$

for any $t>0$ and $\mathbf{f}:[0,t]\to\mathbb{R}^n$. We use this to estimate

$$\begin{aligned} E\left(\left|\int_0^t \mathbf{b}(\mathbf{X},s)-\mathbf{b}(\hat{\mathbf{X}},s)\,ds\right|^2\right) &\le TE\left(\int_0^t \left|\mathbf{b}(\mathbf{X},s)-\mathbf{b}(\hat{\mathbf{X}},s)\right|^2 ds\right)\\ &\le L^2T\int_0^t E(|\mathbf{X}-\hat{\mathbf{X}}|^2)\,ds. \end{aligned}$$

Furthermore

$$\begin{aligned} E\left(\left|\int_0^t \mathbf{B}(\mathbf{X},s)-\mathbf{B}(\hat{\mathbf{X}},s)\,d\mathbf{W}\right|^2\right) &= E\left(\int_0^t \left|\mathbf{B}(\mathbf{X},s)-\mathbf{B}(\hat{\mathbf{X}},s)\right|^2 ds\right)\\ &\le L^2\int_0^t E(|\mathbf{X}-\hat{\mathbf{X}}|^2)\,ds. \end{aligned}$$

Therefore for some appropriate constant C we have

$$E(|\mathbf{X}(t)-\hat{\mathbf{X}}(t)|^2)\le C\int_0^t E(|\mathbf{X}-\hat{\mathbf{X}}|^2)\,ds,$$

provided $0\le t\le T$. If we now set $\phi(t):=E(|\mathbf{X}(t)-\hat{\mathbf{X}}(t)|^2)$, then the foregoing reads

$$\phi(t)\le C\int_0^t \phi(s)\,ds \quad \text{for all } 0\le t\le T.$$

Therefore Gronwall's Lemma, with $C_0=0$, implies $\phi\equiv 0$. Thus $\mathbf{X}(t)=\hat{\mathbf{X}}(t)$ a.s. for all $0\le t\le T$, and so $\mathbf{X}(r)=\hat{\mathbf{X}}(r)$ for all rational $0\le r\le T$, except for some set of probability zero. As $\mathbf{X}$ and $\hat{\mathbf{X}}$ have continuous sample paths almost surely,

$$P\left(\max_{0\le t\le T}|\mathbf{X}(t)-\hat{\mathbf{X}}(t)|>0\right)=0.$$

2. *Existence.* We will utilize the iterative scheme introduced earlier. Define

$$\begin{cases} \mathbf{X}^0(t):=\mathbf{X}_0\\ \mathbf{X}^{n+1}(t):=\mathbf{X}_0+\int_0^t \mathbf{b}(\mathbf{X}^n(s),s)\,ds+\int_0^t \mathbf{B}(\mathbf{X}^n(s),s)\,d\mathbf{W}, \end{cases}$$

for $n=0,1,\dots$ and $0\le t\le T$. Define also

$$d^n(t):=E(|\mathbf{X}^{n+1}(t)-\mathbf{X}^n(t)|^2).$$

We *claim* that

$$d^n(t)\le \frac{(Mt)^{n+1}}{(n+1)!} \quad \text{for all } n=0,\dots,\ 0\le t\le T$$

for some constant M, depending on L, T and $\mathbf{X}_0$. Indeed for $n = 0$, we have

$$\begin{aligned} d^0(t) &= E(|\mathbf{X}^1(t) - \mathbf{X}^0(t)|^2) \\ &= E\left(\left|\int_0^t \mathbf{b}(\mathbf{X}_0, s)\, ds + \int_0^t \mathbf{B}(\mathbf{X}_0, s)\, d\mathbf{W}\right|^2\right) \\ &\le 2E\left(\left|\int_0^t L(1 + |\mathbf{X}_0|)\, ds\right|^2\right) + 2E\left(\int_0^t L^2(1 + |\mathbf{X}_0|^2)\, ds\right) \\ &\le tM \end{aligned}$$

for some large enough constant M. This confirms the claim for $n = 0$.

Next assume the claim is valid for some $n - 1$. Then

$$\begin{aligned} d^n(t) &= E(|\mathbf{X}^{n+1}(t) - \mathbf{X}^n(t)|^2) \\ &= E\left(\left|\int_0^t \mathbf{b}(\mathbf{X}^n, s) - \mathbf{b}(\mathbf{X}^{n-1}, s)\, ds \right.\right. \\ &\qquad \left.\left. + \int_0^t \mathbf{B}(\mathbf{X}^n, s) - \mathbf{B}(\mathbf{X}^{n-1}, s)\, d\mathbf{W}\right|^2\right) \\ &\le 2TL^2 E\left(\int_0^t |\mathbf{X}^n - \mathbf{X}^{n-1}|^2\, ds\right) \\ &\qquad + 2L^2 E\left(\int_0^t |\mathbf{X}^n - \mathbf{X}^{n-1}|^2\, ds\right) \\ &\le 2L^2(1 + T)\int_0^t \frac{M^n s^n}{n!} ds \quad \text{by the induction hypothesis} \\ &\le \frac{M^{n+1}t^{n+1}}{(n+1)!}, \end{aligned}$$

provided we choose $M \ge 2L^2(1 + T)$. This proves the claim.

3. Now note

$$\begin{aligned} \max_{0\le t\le T} |\mathbf{X}^{n+1}(t) - \mathbf{X}^n(t)|^2 &\le 2TL^2 \int_0^T |\mathbf{X}^n - \mathbf{X}^{n-1}|^2\, ds \\ &\qquad + 2 \max_{0\le t\le T} \left|\int_0^t \mathbf{B}(\mathbf{X}^n, s) - \mathbf{B}(\mathbf{X}^{n-1}, s)\, d\mathbf{W}\right|^2. \end{aligned}$$

Consequently the martingale inequality from Chapter 2 implies

$$\begin{aligned} E\left(\max_{0\le t\le T} |\mathbf{X}^{n+1}(t) - \mathbf{X}^n(t)|^2\right) &\le 2TL^2 \int_0^T E(|\mathbf{X}^n - \mathbf{X}^{n-1}|^2)\, ds \\ &\qquad + 8L^2 \int_0^T E(|\mathbf{X}^n - \mathbf{X}^{n-1}|^2)\, ds \\ &\le C\frac{(MT)^n}{n!} \quad \text{by the claim above.} \end{aligned}$$

4. The Borel–Cantelli Lemma thus applies, since

$$P\left(\max_{0\le t\le T}|\mathbf{X}^{n+1}(t)-\mathbf{X}^n(t)|>\frac{1}{2^n}\right)\le 2^{2n}E\left(\max_{0\le t\le T}|\mathbf{X}^{n+1}(t)-\mathbf{X}^n(t)|^2\right)$$
$$\le 2^{2n}\frac{C(MT)^n}{n!}$$

and

$$\sum_{n=1}^{\infty}2^{2n}\frac{(MT)^n}{n!}<\infty.$$

Thus

$$P\left(\max_{0\le t\le T}|\mathbf{X}^{n+1}(t)-\mathbf{X}^n(t)|>\frac{1}{2^n}\text{ i.o.}\right)=0.$$

In light of this, for almost every ω

$$\mathbf{X}^n=\mathbf{X}^0+\sum_{j=0}^{n-1}(\mathbf{X}^{j+1}-\mathbf{X}^j)$$

converges uniformly on $[0,T]$ to a process $\mathbf{X}(\cdot)$. We pass to limits in the definition of $\mathbf{X}^{n+1}(\cdot)$, to prove

$$\mathbf{X}(t)=X_0+\int_0^t\mathbf{b}(\mathbf{X},s)\,ds+\int_0^t\mathbf{B}(\mathbf{X},s)\,d\mathbf{W}\quad\text{ for }0\le t\le T.$$

That is,

$$\text{(SDE)}\qquad\begin{cases}d\mathbf{X}=\mathbf{b}(\mathbf{X},t)dt+\mathbf{B}(\mathbf{X},t)d\mathbf{W}\\ \mathbf{X}(0)=\mathbf{X}_0,\end{cases}$$

for times $0\le t\le T$.

5. We must still show $\mathbf{X}(\cdot)\in\mathbb{L}^2_n(0,T)$. We have

$$E(|\mathbf{X}^{n+1}(t)|^2)\le CE(|\mathbf{X}_0|^2)+CE\left(\left|\int_0^t\mathbf{b}(\mathbf{X}^n,s)\,ds\right|^2\right)$$
$$+CE\left(\left|\int_0^t\mathbf{B}(\mathbf{X}^n,s)\,d\mathbf{W}\right|^2\right)$$
$$\le C(1+E(|\mathbf{X}_0|^2))+C\int_0^tE(|\mathbf{X}^n|^2)\,ds,$$

where, as usual, "C" denotes various constants. By induction, therefore,

$$E(|\mathbf{X}^{n+1}(t)|^2)\le\left[C+C^2+\cdots+C^{n+2}\frac{t^{n+1}}{(n+1)!}\right](1+E(|\mathbf{X}_0|^2)).$$

Consequently

$$E(|\mathbf{X}^{n+1}(t)|^2)\le C(1+E(|\mathbf{X}_0|^2))e^{Ct}.$$

Let $n\to\infty$:

$$E(|\mathbf{X}(t)|^2)\le C(1+E(|\mathbf{X}_0|^2))e^{Ct}\quad\text{ for all }0\le t\le T;$$

and so $\mathbf{X} \in \mathbb{L}^2_n(0,T)$. □

C. PROPERTIES OF SOLUTIONS

In this section we mention, without proofs, a few properties of the solution to various SDE.

THEOREM (Estimate on higher moments of solutions). *Suppose that* $\mathbf{b}$, $\mathbf{B}$ *and* $\mathbf{X}_0$ *satisfy the hypotheses of the Existence and Uniqueness Theorem. If, in addition,*

$$E(|\mathbf{X}_0|^{2p}) < \infty \quad \text{for some integer } p > 1,$$

then the solution $\mathbf{X}(\cdot)$ *of*

$$\text{(SDE)} \qquad \begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases}$$

satisfies the estimates

$$\text{(i)} \qquad E(|\mathbf{X}(t)|^{2p}) \le C_2(1 + E(|\mathbf{X}_0|^{2p}))e^{C_1 t}$$

and

$$\text{(ii)} \qquad E(|\mathbf{X}(t) - \mathbf{X}_0|^{2p}) \le C_2(1 + E(|\mathbf{X}_0|^{2p}))t^p e^{C_2 t}$$

for certain constants C_1 *and* C_2*, depending only on* T, L, m, n.

The estimates above on the moments of $\mathbf{X}(\cdot)$ are fairly crude, but are nevertheless sometimes useful:

APPLICATION: SAMPLE PATH PROPERTIES. The possibility that $\mathbf{B} \equiv 0$ is not excluded, and consequently it could happen that the solution of our SDE is really a solution of the ODE

$$\dot{\mathbf{X}} = \mathbf{b}(\mathbf{X},t),$$

with possibly random initial data. In this case the mapping $t \mapsto \mathbf{X}(t)$ will be smooth if $\mathbf{b}$ is. On the other hand, if for some $1 \le i \le n$

$$\sum_{1 \le l \le m} |b^{il}(x,t)|^2 > 0 \quad \text{for all } x \in \mathbb{R}^n,\ 0 \le t \le T,$$

then almost every sample path $t \mapsto X^i(t)$ is nowhere differentiable for a.e. ω. We can however use estimates (i) and (ii) above to check the hypotheses of Kolmogorov's Theorem from §C in Chapter 3. It follows that for almost all sample paths,

the mapping $t \mapsto \mathbf{X}(t)$ is Hölder continuous with each exponent less than $\frac{1}{2}$,

provided $E(|\mathbf{X}_0|^{2p}) < \infty$ for each $1 \le p < \infty$. □

THEOREM (Dependence on parameters). *Suppose for $k = 1, 2, \dots$ that $\mathbf{b}^k, \mathbf{B}^k$ and $\mathbf{X}_0^k$ satisfy the hypotheses of the Existence and Uniqueness Theorem, with the same constant L. Assume further that*

$$(a) \qquad \lim_{k\to\infty} E(|\mathbf{X}_0^k - \mathbf{X}_0|^2) = 0,$$

and for each $M > 0$,

$$(b) \qquad \lim_{k\to\infty} \max_{\substack{0\le t\le T\\ |x|\le M}} (|\mathbf{b}^k(x,t) - \mathbf{b}(x,t)| + |\mathbf{B}^k(x,t) - \mathbf{B}(x,t)|) = 0.$$

Finally suppose that $\mathbf{X}^k(\cdot)$ solves

$$\begin{cases} d\mathbf{X}^k = \mathbf{b}^k(\mathbf{X}^k, t)dt + \mathbf{B}^k(\mathbf{X}^k, t)d\mathbf{W} \\ \mathbf{X}^k(0) = \mathbf{X}_0^k. \end{cases}$$

Then

$$\lim_{k\to\infty} E\left(\max_{0\le t\le T} |\mathbf{X}^k(t) - \mathbf{X}(t)|^2\right) = 0,$$

where $\mathbf{X}$ is the unique solution of

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0. \end{cases}$$

Example (Small noise limits). In particular, for almost every ω the random trajectories of the SDE

$$\begin{cases} d\mathbf{X}^\varepsilon = \mathbf{b}(\mathbf{X}^\varepsilon)dt + \varepsilon d\mathbf{W} \\ \mathbf{X}^\varepsilon(0) = x_0 \end{cases}$$

converge uniformly on $[0, T]$ as $\varepsilon \to 0$ to the deterministic trajectory of

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{b}(\mathbf{x}), \\ \mathbf{x}(0) = x_0. \end{cases}$$

□

D. LINEAR STOCHASTIC DIFFERENTIAL EQUATIONS

This section presents some fairly explicit formulas for solutions of linear SDE.

DEFINITION. The stochastic differential equation

$$d\mathbf{X} = \mathbf{b}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t)d\mathbf{W}$$

is *linear* provided the coefficients $\mathbf{b}$ and $\mathbf{B}$ have this form:

$$\mathbf{b}(x, t) := \mathbf{c}(t) + \mathbf{D}(t)x,$$

for $\mathbf{c} : [0, T] \to \mathbb{R}^n$, $\mathbf{D} : [0, T] \to \mathbb{M}^{n\times n}$, and

$$\mathbf{B}(x, t) := \mathbf{E}(t) + \mathbf{F}(t)x$$

for $\mathbf{E} : [0, T] \to \mathbb{M}^{n\times m}$, $\mathbf{F} : [0, T] \to L(\mathbb{R}^n, \mathbb{M}^{n\times m})$, the space of bounded linear mappings from $\mathbb{R}^n$ to $\mathbb{M}^{n\times m}$.

DEFINITION. A linear SDE is called *homogeneous* if $\mathbf{c} \equiv \mathbf{E} \equiv 0$ for $0 \leq t \leq T$. It is called *linear in the narrow sense* if $\mathbf{F} \equiv 0$.

Remark. If

$$\sup_{0 \leq t \leq T} [|\mathbf{c}(t)| + |\mathbf{D}(t)| + |\mathbf{E}(t)| + |\mathbf{F}(t)|] < \infty,$$

then $\mathbf{b}$ and $\mathbf{B}$ satisfy the hypotheses of the Existence and Uniqueness Theorem. Thus the linear SDE

$$\begin{cases} d\mathbf{X} = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X})dt + (\mathbf{E}(t) + \mathbf{F}(t)\mathbf{X})d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases}$$

has a unique solution, provided $E(|\mathbf{X}_0|^2) < \infty$, and $\mathbf{X}_0$ is independent of $\mathcal{W}^+(0)$.

□

FORMULAS FOR SOLUTIONS: linear equations in narrow sense

Suppose first $\mathbf{D} \equiv D$ is constant. Then the solution of

$$\begin{cases} d\mathbf{X} = (\mathbf{c}(t) + D\mathbf{X})dt + \mathbf{E}(t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases} \tag{10}$$

is

$$\mathbf{X}(t) = e^{Dt}\mathbf{X}_0 + \int_0^t e^{D(t-s)}(\mathbf{c}(s)ds + \mathbf{E}(s)\, d\mathbf{W}), \tag{11}$$

where

$$e^{Dt} := \sum_{k=0}^{\infty} \frac{D^k t^k}{k!}.$$

More generally, the solution of

$$\begin{cases} d\mathbf{X} = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X})dt + \mathbf{E}(t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases} \tag{12}$$

is

$$\mathbf{X}(t) = \Phi(t) \left(\mathbf{X}_0 + \int_0^t \Phi(s)^{-1}(\mathbf{c}(s)ds + \mathbf{E}(s)\, d\mathbf{W}) \right), \tag{13}$$

where $\Phi(\cdot)$ is the *fundamental matrix* of the nonautonomous system of ODE

$$\frac{d\Phi}{dt} = \mathbf{D}(t)\Phi, \quad \Phi(0) = I.$$

□

These assertions follow formally from standard formulas in ODE theory if we write $\mathbf{E}d\mathbf{W} = \mathbf{E}\xi dt$, ξ as usual denoting white noise, and regard $\mathbf{E}\xi$ as an inhomogeneous term driving the ODE

$$\dot{\mathbf{X}} = \mathbf{c}(t) + \mathbf{D}(t)\mathbf{X} + \mathbf{E}(t)\xi.$$

This will *not* be so if $\mathbf{F}(\cdot) \not\equiv 0$, owing to the extra term in Itô's formula.

Observe also that formula (13) shows $\mathbf{X}(t)$ to be Gaussian if $\mathbf{X}_0$ is.

FORMULAS FOR SOLUTIONS: general scalar linear equations

Suppose now $n = 1$, but $m \geq 1$ is arbitrary. Then the solution of

$$\begin{cases} dX = (c(t) + d(t)X)dt + \sum_{l=1}^{m}(e^l(t) + f^l(t)X)dW^l \\ X(0) = X_0 \end{cases} \tag{14}$$

is

$$\begin{aligned} X(t) = \Phi(t) \left(X_0 + \int_0^t \Phi(s)^{-1} \left(c(s) - \sum_{l=1}^{m} e^l(s) f^l(s) \right) ds \right) \\ + \int_0^t \sum_{l=1}^{m} \Phi(s)^{-1} e^l(s)\, dW^l, \end{aligned} \tag{15}$$

where

$$\Phi(t) := \exp\left(\int_0^t d - \sum_{l=1}^{m} \frac{(f^l)^2}{2}\, ds + \int_0^t \sum_{l=1}^{m} f^l\, dW^l \right).$$

See Arnold [**A, Chapter 8**] for more formulas for solutions of general linear equations. □

3. SOME METHODS FOR SOLVING LINEAR SDE

For practice with Itô's formula, let us derive some of the formulas stated above.

Example 1. Consider first the linear stochastic differential equation

$$\begin{cases} dX = d(t)Xdt + f(t)XdW \\ X(0) = X_0 \end{cases} \tag{16}$$

for $m = n = 1$. We will try to find a solution having the product form

$$X(t) = X_1(t)X_2(t),$$

where

$$\begin{cases} dX_1 = f(t)X_1 dW \\ X_1(0) = X_0 \end{cases} \tag{17}$$

and

$$\begin{cases} dX_2 = A(t)dt + B(t)dW \\ X_2(0) = 1, \end{cases} \tag{18}$$

where the functions A and B are to be selected. Then

$$\begin{aligned} dX &= d(X_1X_2) \\ &= X_1dX_2 + X_2dX_1 + f(t)X_1B(t)dt \\ &= f(t)XdW + (X_1dX_2 + f(t)X_1B(t)dt), \end{aligned}$$

according to (17). Now we try to choose A, B so that

$$dX_2 + f(t)B(t)dt = d(t)X_2 dt.$$

For this, $B \equiv 0$ and $A(t) = d(t)X_2(t)$ will work. Thus (18) reads

$$\begin{cases} dX_2 = d(t)X_2 dt \\ X_2(0) = 1. \end{cases}$$

This is non-random: $X_2(t) = e^{\int_0^t d(s)\,ds}$. Since the solution of (17) is

$$X_1(t) = X_0 e^{\int_0^t f(s)\,dW - \frac{1}{2}\int_0^t f^2(s)\,ds},$$

we conclude that

$$X(t) = X_1(t)X_2(t) = X_0 e^{\int_0^t f(s)\,dW + \int_0^t d(s) - \frac{1}{2}f^2(s)\,ds},$$

a formula noted earlier. □

Example 2. Consider next the general equation

$$\begin{cases} dX = (c(t) + d(t)X)dt + (e(t) + f(t)X)dW \\ X(0) = X_0, \end{cases} \tag{19}$$

again for $m = n = 1$. As above, we try for a solution of the form

$$X(t) = X_1(t)X_2(t),$$

where now

$$\begin{cases} dX_1 = d(t)X_1 dt + f(t)X_1 dW \\ X_1(0) = 1 \end{cases} \tag{20}$$

and

$$\begin{cases} dX_2 = A(t)dt + B(t)dW \\ X_2(0) = X_0, \end{cases} \tag{21}$$

the functions A, B to be chosen. Then

$$\begin{aligned} dX &= X_2 dX_1 + X_1 dX_2 + f(t)X_1 B(t)dt \\ &= d(t)Xdt + f(t)XdW \\ &\quad + X_1(A(t)dt + B(t)dW) + f(t)X_1 B(t)dt. \end{aligned}$$

We now require

$$X_1(A(t)dt + B(t)dW) + f(t)X_1 B(t)dt = c(t)dt + e(t)dW;$$

and this identity will hold if we take

$$\begin{cases} A(t) := [c(t) - f(t)e(t)](X_1(t))^{-1} \\ B(t) := e(t)(X_1(t))^{-1}. \end{cases}$$

Observe that since $X_1(t) = e^{\int_0^t f\,dW + \int_0^t d - \frac{1}{2}f^2\,ds}$, we have $X_1(t) > 0$ almost surely. Consequently

$$X_2(t) = X_0 + \int_0^t [c(s) - f(s)e(s)](X_1(s))^{-1}\,ds + \int_0^t e(s)(X_1(s))^{-1}\,dW.$$

Employing this and the expression above for X_1, we arrive at the formula, a special case of (15):

$$\begin{aligned} X(t) &= X_1(t)X_2(t) \\ &= \exp\left(\int_0^t d(s) - \frac{1}{2}f^2(s)\,ds + \int_0^t f(s)\,dW\right) \\ &\times \left(X_0 + \int_0^t \exp\left(-\int_0^r d(r) - \frac{1}{2}f^2(r)\,dr - \int_0^s f(r)\,dW\right)(c(s) - e(s)f(s))\,ds \right. \\ &+ \left.\int_0^t \exp\left(-\int_0^s d(r) - \frac{1}{2}f^2(r)\,dr - \int_0^s f(r)\,dW\right)e(s)\,dW.\right) \end{aligned}$$

□

Remark. There is great theoretical and practical interest in numerical methods for simulation of solutions to random differential equations. The paper of Higham [**H**] is a good introduction.

CHAPTER 6: APPLICATIONS

A. Stopping times
B. Applications to PDE, Feynman-Kac formula
C. Optimal stopping
D. Options pricing
E. The Stratonovich integral

This chapter is devoted to some applications and extensions of the theory developed earlier.

A. STOPPING TIMES

DEFINITIONS, BASIC PROPERTIES. Let $(\Omega, \mathcal{U}, P)$ be a probability space and $\mathcal{F}(\cdot)$ a filtration of σ–algebras, as in Chapters 4 and 5. We introduce now some random times that are well–behaved with respect to $\mathcal{F}(\cdot)$:

DEFINITION. A random variable $\tau : \Omega \to [0, \infty]$ is called a *stopping time* with respect to $\mathcal{F}(\cdot)$ provided

$$\{\tau \leq t\} \in \mathcal{F}(t) \quad \text{for all } t \geq 0.$$

This says that the set of all $\omega \in \Omega$ such that $\tau(\omega) \leq t$ is an $\mathcal{F}(t)$-measurable set. Note that τ is allowed to take on the value $+\infty$, and also that any constant $\tau \equiv t_0$ is a stopping time.

THEOREM (Properties of stopping times). *Let τ_1 and τ_2 be stopping times with respect to $\mathcal{F}(\cdot)$. Then*

(i) *$\{\tau < t\} \in \mathcal{F}(t)$, and so $\{\tau = t\} \in \mathcal{F}(t)$, for all times $t \geq 0$.*

(ii) *$\tau_1 \wedge \tau_2 := \min(\tau_1, \tau_2)$, $\tau_1 \vee \tau_2 := \max(\tau_1, \tau_2)$ are stopping times.*

PROOF. Observe that

$$\{\tau < t\} = \bigcup_{k=1}^{\infty} \underbrace{\{\tau \leq t - 1/k\}}_{\in \mathcal{F}(t-1/k) \subseteq \mathcal{F}(t)}.$$

Also, we have $\{\tau_1 \wedge \tau_2 \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\} \in \mathcal{F}(t)$, and furthermore $\{\tau_1 \vee \tau_2 \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \in \mathcal{F}(t)$. □

The notion of stopping times comes up naturally in the study of stochastic differential equations, as it allows us to investigate phenomena occuring over "random time intervals". An example will make this clearer:

Example (Hitting a set). Consider the solution $\mathbf{X}(\cdot)$ of the SDE

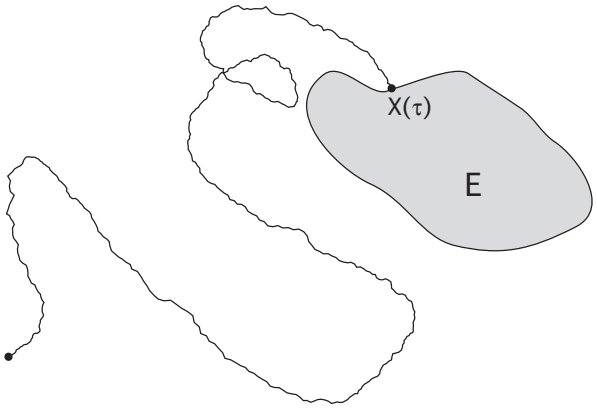
$$\begin{cases} d\mathbf{X}(t) = \mathbf{b}(t, \mathbf{X})dt + \mathbf{B}(t, \mathbf{X})d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0, \end{cases}$$

where $\mathbf{b}$, $\mathbf{B}$ and $\mathbf{X}_0$ satisfy the hypotheses of the Existence and Uniqueness Theorem.

THEOREM. *Let E be either a nonempty closed subset or a nonempty open subset of $\mathbb{R}^n$. Then*

$$\tau := \inf\{t \geq 0 \,|\, \mathbf{X}(t) \in E\}$$

is a stopping time. (We put $\tau = +\infty$ for those sample paths of $\mathbf{X}(\cdot)$ that never hit E.)



PROOF. Fix $t \geq 0$; we must show $\{\tau \leq t\} \in \mathcal{F}(t)$. Take $\{t_i\}_{i=1}^{\infty}$ to be a countable dense subset of $[0, \infty)$. First we assume that $E = U$ is an open set. Then the event

$$\{\tau \leq t\} = \bigcup_{t_i \leq t} \underbrace{\{\mathbf{X}(t_i) \in U\}}_{\in \mathcal{F}(t_i) \subseteq \mathcal{F}(t)}$$

belongs to $\mathcal{F}(t)$.

Next we assume that $E = C$ is a closed set. Set $d(x, C) := \text{dist}(x, C)$ and define the open sets

$$U_n = \{x : d(x, C) < \frac{1}{n}\}.$$

The event

$$\{\tau \leq t\} = \bigcap_{n=1}^{\infty} \bigcup_{t_i \leq t} \underbrace{\{\mathbf{X}(t_i) \in U_n)\}}_{\in \mathcal{F}(t_i) \subseteq \mathcal{F}(t)}$$

also belongs to $\mathcal{F}(t)$. □

Discussion. The random variable

$$\sigma := \sup\{t \geq 0 \,|\, \mathbf{X}(t) \in E\},$$

the last time that $\mathbf{X}(t)$ hits E, is in general *not* a stopping time. The heuristic reason is that the event $\{\sigma \leq t\}$ would depend upon the entire future history of process and thus would not in general be $\mathcal{F}(t)$-measurable. (In applications $\mathcal{F}(t)$ "contains the history of $\mathbf{X}(\cdot)$ up to and including time t, but does not contain information about the future".)

The name "stopping time" comes from the example, where we sometimes think of halting the sample path $\mathbf{X}(\cdot)$ at the first time τ that it hits a set E. But there are

many examples where we do not really stop the process at time τ. Thus "stopping time" is not a particularly good name and "Markov time" would be better. □

STOCHASTIC INTEGRALS AND STOPPING TIMES. Our next task is to consider stochastic integrals with random limits of integration and to work out an Itô formula for these.

DEFINITION. If $G \in \mathbb{L}^2(0,T)$ and τ is a stopping time with $0 \le \tau \le T$, we define

$$\int_0^\tau G\,dW := \int_0^T \chi_{\{t\le\tau\}} G\,dW.$$

LEMMA (Itô integrals with stopping times). *If $G \in \mathbb{L}^2(0,T)$ and $0 \le \tau \le T$ is a stopping time, then*

(i) $$E\left(\int_0^\tau G\,dW\right) = 0$$

(ii) $$E\left((\int_0^\tau G\,dW)^2\right) = E\left(\int_0^\tau G^2\,dt\right).$$

Proof. We have

$$E\left(\int_0^\tau G\,dW\right) = E\left(\int_0^T \underbrace{\chi_{\{t\le\tau\}}G}_{\in\mathbb{L}^2(0,T)}\,dW\right) = 0,$$

and

$$\begin{aligned} E((\int_0^\tau G\,dW)^2) &= E((\int_0^T \chi_{\{t\le\tau\}}G\,dW)^2) \\ &= E(\int_0^T (\chi_{\{t\le\tau\}}G)^2\,dt) \\ &= E(\int_0^\tau G^2\,dt). \end{aligned}$$

□

Similar formulas hold for vector–valued processes.

ITÔ'S FORMULA WITH STOPPING TIMES. As usual, let $\mathbf{W}(\cdot)$ denote m-dimensional Brownian motion. Recall next from Chapter 4 that if $d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W}$, then for each C^2 function u,

$$du(\mathbf{X},t) = \frac{\partial u}{\partial t}dt + \sum_{i=1}^n \frac{\partial u}{\partial x_i}d\mathbf{X}^i + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 u}{\partial x_i \partial x_j}\sum_{k=1}^m b^{ik}b^{jk}dt. \tag{1}$$

Written in integral form, this means:

$$u(\mathbf{X}(t),t) - u(\mathbf{X}(0),0) = \int_0^t \left(\frac{\partial u}{\partial t} + Lu\right)\, ds + \int_0^t Du \cdot \mathbf{B}\, d\mathbf{W}, \tag{2}$$

for the differential operator

$$Lu := \frac{1}{2}\sum_{i,j=1}^n a^{ij} u_{x_i x_j} + \sum_{i=1}^n b^i u_{x_i}, \quad a^{ij} = \sum_{k=1}^m b^{ik} b^{jk},$$

and

$$Du \cdot \mathbf{B}\, d\mathbf{W} = \sum_{k=1}^m \sum_{i=1}^n u_{x_i} b^{ik} dW^k.$$

The argument of u in these integrals is $(\mathbf{X}(s), s)$. We call L the *generator*.

For a fixed $\omega \in \Omega$, formula (2) holds for all $0 \le t \le T$. Thus we may set $t = \tau$, where τ is a stopping time, $0 \le \tau \le T$:

$$u(\mathbf{X}(\tau),\tau) - u(\mathbf{X}(0),0) = \int_0^\tau \left(\frac{\partial u}{\partial t} + Lu\right)\, ds + \int_0^\tau Du \cdot \mathbf{B}\, d\mathbf{W}.$$

Take expected value:

$$E(u(\mathbf{X}(\tau),\tau)) - E(u(\mathbf{X}(0),0)) = E\left(\int_0^\tau \left(\frac{\partial u}{\partial t} + Lu\right)\, ds\right). \tag{3}$$

We will see in the next section that this formula provides a very important link between stochastic differential equations and (nonrandom) partial differential equations.

BROWNIAN MOTION AND THE LAPLACIAN. The most important case is $\mathbf{X}(\cdot) = \mathbf{W}(\cdot)$, n-dimensional Brownian motion, the generator of which is

$$Lu = \frac{1}{2}\sum_{i=1}^n u_{x_i x_i} =: \frac{1}{2}\Delta u.$$

The expression Δu is called the *Laplacian* of u and occurs throughout mathematics and physics. We will demonstrate in the next section some important links with Brownian motion. □

B. APPLICATIONS TO PDE, FEYNMAN–KAC FORMULA

PROBABILISTIC FORMULAS FOR SOLUTIONS OF PDE.

Example 1 (Expected hitting time to a boundary). Let $U \subset \mathbb{R}^n$ be a bounded open set, with smooth boundary ∂U. According to standard PDE theory, there exists a smooth solution u of the equation

$$\begin{cases} -\frac{1}{2}\Delta u = 1 & \text{in } U \\ \qquad u = 0 & \text{on } \partial U. \end{cases} \tag{4}$$

Our goal is to find a probabilistic representation formula for u. For this, fix any point $x \in U$ and consider then an n-dimensional Brownian motion $\mathbf{W}(\cdot)$. Then $\mathbf{X}(\cdot) := \mathbf{W}(\cdot) + x$ represents a "Brownian motion starting at x". Define

$$\tau_x := \text{first time } \mathbf{X}(\cdot) \text{ hits } \partial U.$$

THEOREM. *We have*

$$u(x) = E(\tau_x) \quad \textit{for all } x \in U. \tag{5}$$

In particular, $u > 0$ *in* U.

PROOF. We employ formula (3), with $Lu = \frac{1}{2}\Delta u$. We have for each $n = 1, 2, \ldots$

$$E(u(\mathbf{X}(\tau_x \wedge n))) - E(u(\mathbf{X}(0))) = E\left(\int_0^{\tau_x \wedge n} \frac{1}{2}\Delta u(\mathbf{X})\, ds\right).$$

Since $\frac{1}{2}\Delta u = -1$ and u is bounded,

$$\lim_{n\to\infty} E(\tau_x \wedge n) < \infty.$$

Thus τ_x is integrable. Thus if we let $n \to \infty$ above, we get

$$u(x) - E(u(\mathbf{X}(\tau_x))) = E\left(\int_0^{\tau_x} 1\, ds\right) = E(\tau_x).$$

But $u = 0$ on ∂U, and so $u(\mathbf{X}(\tau_x)) \equiv 0$. Formula (5) follows. □

Again recall that u is bounded on U. Hence

$$E(\tau_x) < \infty, \text{ and so } \tau_x < \infty \text{ a.s., for all } x \in U.$$

This says that *Brownian sample paths starting at any point* $x \in U$ *will with probability* 1 *eventually hit* ∂U.

Example 2 (Probabilistic representation of harmonic functions). Let $U \subset \mathbb{R}^n$ be a smooth, bounded domain and $g : \partial U \to \mathbb{R}$ a given continuous function. It is known from classical PDE theory that there exists a function $u \in C^2(U) \cap C(\bar{U})$ satisfying the boundary value problem:

$$\begin{cases} \Delta u = 0 & \text{in } U \\ u = g & \text{on } \partial U. \end{cases} \tag{6}$$

We call u a *harmonic* function.

THEOREM. *We have for each point* $x \in U$

$$u(x) = E(g(\mathbf{X}(\tau_x))), \tag{7}$$

for $\mathbf{X}(\cdot) := \mathbf{W}(\cdot) + x$, *Brownian motion starting at* x.

PROOF. As shown above,

$$E(u(\mathbf{X}(\tau_x))) = E(u(\mathbf{X}(0))) + E\left(\int_0^{\tau_x} \frac{1}{2}\Delta u(\mathbf{X})\, ds\right) = E(u(\mathbf{X}(0))) = u(x),$$

the second equality valid since $\Delta u = 0$ in U. Since $u = g$ on ∂U, formula (7) follows. □

APPLICATION: In particular, if $\Delta u = 0$ in some open set containing the ball $B(x, r)$, then

$$u(x) = E(u(\mathbf{X}(\tau_x))),$$

where τ_x now denotes the hitting time of Brownian motion starting at x to $\partial B(x, r)$. Since Brownian motion is isotropic in space, we may reasonably guess that the term on the right hand side is just the average of u over the sphere $\partial B(x, r)$, with respect to surface measure. That is, we have the identity

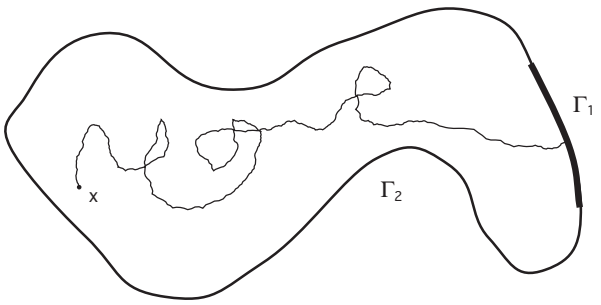
$$u(x) = \frac{1}{\text{area of } \partial B(x, r)} \int_{\partial B(x,r)} u\, dS. \tag{8}$$

This is the *mean value formula* for harmonic functions. □

Example 3 (Hitting one part of a boundary first). Assume next that we can write ∂U as the union of two disjoint parts Γ_1, Γ_2. Let u solve the PDE

$$\begin{cases} \Delta u = 0 & \text{in } U \\ u = 1 & \text{on } \Gamma_1 \\ u = 0 & \text{on } \Gamma_2. \end{cases}$$

THEOREM. *For each point $x \in U$, $u(x)$ is the probability that a Brownian motion starting at x hits Γ_1 before hitting Γ_2.*



PROOF. Apply (7) for

$$g = \begin{cases} 1 & \text{on } \Gamma_1 \\ 0 & \text{on } \Gamma_2. \end{cases}$$

Then

$$u(x) = E(g(\mathbf{X}(\tau_x))) = \text{probability of hitting } \Gamma_1 \text{ before } \Gamma_2.$$

□

FEYNMAN–KAC FORMULA. Now we extend Example #2 above to obtain a probabilistic representation for the unique solution of the PDE

$$\begin{cases} -\frac{1}{2}\Delta u + cu = f & \text{in } U \\ u = 0 & \text{on } \partial U. \end{cases}$$

We assume c, f are smooth functions, with $c \geq 0$ in U.

THEOREM (FEYNMAN–KAC FORMULA). *For each $x \in U$,*

$$u(x) = E\left(\int_0^{\tau_x} f(\mathbf{X}(t)) e^{-\int_0^t c(\mathbf{X})\, ds} dt\right)$$

where, as before, $\mathbf{X}(\cdot) := \mathbf{W}(\cdot) + x$ is a Brownian motion starting at x, and τ_x denotes the first hitting time of ∂U.

PROOF. We know $E(\tau_x) < \infty$. Since $c \geq 0$, the integral above converges.

First look at the process

$$Y(t) := e^{Z(t)},$$

for $Z(t) := -\int_0^t c(\mathbf{X})\, ds$. Then

$$dZ = -c(\mathbf{X})dt,$$

and so Itô's formula yields

$$dY = -c(\mathbf{X})Y dt.$$

Hence the Itô product rule implies

$$\begin{aligned} d\left(u(\mathbf{X})e^{-\int_0^t c(\mathbf{X})\, ds}\right) &= (du(\mathbf{X}))e^{-\int_0^t c(\mathbf{X})\, ds} \\ &\quad + u(\mathbf{X})d\left(e^{-\int_0^t c(\mathbf{X}) ds}\right) \\ &= \left(\frac{1}{2}\Delta u(\mathbf{X})dt + \sum_{i=1}^n \frac{\partial u(\mathbf{X})}{\partial x_i} dW^i\right) e^{-\int_0^t c(\mathbf{X}) ds} \\ &\quad + u(\mathbf{X})(-c(\mathbf{X})dt)e^{-\int_0^t c(\mathbf{X})\, ds}. \end{aligned}$$

We use formula (3) for $\tau = \tau_x$, and take the expected value, obtaining

$$\begin{aligned} &E\left(u(\mathbf{X}(\tau_x))e^{-\int_0^{\tau_x} c(\mathbf{X})\, ds}\right) - E(u(\mathbf{X}(0))) \\ &= E\left(\int_0^{\tau_x} \left[\frac{1}{2}\Delta u(\mathbf{X}) - c(\mathbf{X})u(\mathbf{X})\right] e^{-\int_0^t c(\mathbf{X})\, ds} dt\right). \end{aligned}$$

Since u solves (8), this simplifies to give

$$u(x) = E(u(\mathbf{X}(0))) = E\left(\int_0^{\tau_x} f(\mathbf{X}) e^{-\int_0^t c(\mathbf{X})\, ds} dt\right),$$

as claimed. □

An interpretation. We can explain this formula as describing a Brownian motion with "killing", as follows.

Suppose that the Brownian particles may disappear at a random killing time σ, for example by being absorbed into the medium within which it is moving. Assume further that the probability of its being killed in a short time interval $[t, t+h]$ is

$$c(\mathbf{X}(t))h + o(h).$$

Then the probability of the particle surviving until time t is approximately equal to

$$(1 - c(\mathbf{X}(t_1))h)(1 - c(\mathbf{X}(t_2))h) \dots (1 - c(\mathbf{X}(t_n))h),$$

where $0 = t_0 < t_1 < \cdots < t_n = t$, $h = t_{k+1} - t_k$. As $h \to 0$, this converges to $e^{-\int_0^t c(\mathbf{X})\, ds}$.

Hence it should be that

$$\begin{aligned} u(x) = &\text{ average of } f(\mathbf{X}(\cdot)) \text{ over all sample paths which survive to hit } \partial U \\ &= E\left(\int_0^{\tau_x} f(\mathbf{X}) e^{-\int_0^t c(\mathbf{X})\, ds} dt\right). \end{aligned}$$

□

Remark. If we consider in these examples the solution of the SDE

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X})dt + \mathbf{B}(\mathbf{X})d\mathbf{W} \\ \mathbf{X}(0) = x, \end{cases}$$

we can obtain similar formulas, where now

$$\tau_x = \text{hitting time of } \partial U \text{ for } \mathbf{X}(\cdot)$$

and $\frac{1}{2}\Delta u$ is replaced by the operator

$$Lu := \frac{1}{2}\sum_{i,j=1}^n a^{ij} u_{x_i x_j} + \sum_{i=1}^n b^i u_{x_i}.$$

Note, however, we need to know that the various PDE have smooth solutions. This need not always be the case for degenerate elliptic operators L. □

C. OPTIMAL STOPPING The general mathematical setting for many control theory problems is this. We are given some "system" whose state evolves in time according to a differential equation (deterministic or stochastic). Given also are certain *controls* which affect somehow the behavior of the system: these controls typically either modify some parameters in the dynamics or else stop the process, or both. Finally we are given a *cost criterion*, depending upon our choice of control and the corresponding state of the system.

The goal is to discover an optimal choice of controls, to minimize the cost criterion.

The easiest stochastic control problem of the general type outlined above occurs when we cannot directly affect the SDE controlling the evolution of $\mathbf{X}(\cdot)$ and can only decide at each instance whether or not to stop. A typical such problem follows.

STOPPING A STOCHASTIC DIFFERENTIAL EQUATION. Let $U \subset \mathbb{R}^m$ be a bounded, smooth domain. Suppose $\mathbf{b} : \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{B} : \mathbb{R}^n \to M^{n\times m}$ satisfy the usual assumptions.

Then for each $x \in U$ the stochastic differential equation

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X})dt + \mathbf{B}(\mathbf{X})d\mathbf{W} \\ X_0 = x \end{cases}$$

has a unique solution. Let $\tau = \tau_x$ denote the hitting time of ∂U. Let θ be any stopping time with respect to $\mathcal{F}(\cdot)$, and for each such θ define the *expected cost* of stopping $\mathbf{X}(\cdot)$ at time $\theta \wedge \tau$ to be

$$J_x(\theta) := E\Big(\int_0^{\theta\wedge\tau} f(\mathbf{X}(s))\,ds + g(\mathbf{X}(\theta\wedge\tau))\Big). \tag{9}$$

The idea is that if we stop at the possibly random time $\theta < \tau$, then the cost is a given function g of the current state of $\mathbf{X}(\theta)$. If instead we do not stop the process before it hits ∂U, that is, if $\theta \geq \tau$, the cost is $g(\mathbf{X}(\tau))$. In addition there is a running cost per unit time f of keeping the system in operation until time $\theta \wedge \tau$.

OPTIMAL STOPPING. The main question is this: does there exist an optimal stopping time $\theta^* = \theta_x^*$, for which

$$J_x(\theta^*) = \min_{\substack{\theta \text{ stopping} \\ \text{time}}} J_x(\theta)?$$

And if so, how can we find θ^*? It turns out to be very difficult to try to design θ^* directly. A much better idea is to turn attention to the *value function*

$$u(x) := \inf_\theta J_x(\theta), \tag{10}$$

and to try to figure out what u is as a function of $x \in U$. Note that $u(x)$ is the minimum expected cost, given we start the process at x. It turns out that once we know u, we will be then be able to construct an optimal θ^*. This approach is called *dynamic programming*.

OPTIMALITY CONDITIONS. So assume u is defined above and suppose u is smooth enough to justify the following calculations. We wish to determine the properties of this function.

First of all, notice that we could just take $\theta \equiv 0$ in the definition (10). That is, we could just stop immediately and incur the cost $g(\mathbf{X}(0)) = g(x)$. Hence

$$u(x) \leq g(x) \quad \text{for each point } x \in U. \tag{11}$$

Furthermore, $\tau \equiv 0$ if $x \in \partial U$, and so

$$u(x) = g(x) \quad \text{for each point } x \in \partial U. \tag{12}$$

Next take any point $x \in U$ and fix some small number $\delta > 0$. Now if we do not stop the system for time δ, then according to (SDE) the new state of the system

at time δ will be $\mathbf{X}(\delta)$. Then, given that we are at the point $\mathbf{X}(\delta)$, the best we can achieve in minimizing the cost thereafter must be

$$u(\mathbf{X}(\delta)).$$

So if we choose not to stop the system for time δ, and assuming we do not hit ∂U, our cost is at least

$$E(\int_0^\delta f(\mathbf{X})\,ds + u(\mathbf{X}(\delta))).$$

Since $u(x)$ is the infimum of costs over all stopping times, we therefore have

$$u(x) \le E(\int_0^\delta f(\mathbf{X})\,ds + u(\mathbf{X}(\delta))).$$

Now by Itô's formula

$$E(u(\mathbf{X}(\delta))) = u(x) + E(\int_0^\delta Lu(\mathbf{X})\,ds),$$

for

$$Lu = \frac{1}{2}\sum_{i,j=1}^n a^{ij}\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b^i \frac{\partial u}{\partial x_i}, \quad a^{ij} = \sum_{k=1}^m b^{ik}b^{jk}.$$

Hence

$$0 \le E(\int_0^\delta f(\mathbf{X}) + Lu(\mathbf{X})\,ds).$$

Divide by $\delta > 0$, and then let $\delta \to 0$:

$$0 \le f(x) + Lu(x).$$

Equivalently, we have

$$Mu \le f \quad \text{in } U, \tag{13}$$

where $Mu := -Lu$.

Finally we observe that if in (11) a strict inequality held, that is, if

$$u(x) < g(x) \text{ at some point } x \in U,$$

then it is optimal not to stop the process at once. Thus it is plausible to think that we should leave the system going, for at least some very small time δ. In this circumstance we then would have an equality in the formula above; and so

$$Mu = f \quad \text{at those points where } u < g. \tag{14}$$

In summary, we combine (11)–(14) to find that *if* the formal reasoning above is valid, then the value function u satisfies:

$$\begin{cases} \max(Mu - f, u - g) = 0 & \text{in } U \\ u = g & \text{on } \partial U \end{cases} \tag{15}$$

These are the *optimality conditions*.

SOLVING FOR THE VALUE FUNCTION. Our rigorous study of the stopping time problem now begins by showing first that there exists a unique solution u of (15) and second that this u is in fact $\min_\theta J_x(\theta)$. Then we will use u to design θ^*, an optimal stopping time.

THEOREM. *Suppose f, g are given smooth functions. There exists a unique funtion u, with bounded second derivatives, such that:*

(i) $u \leq g$ *in* U,
(ii) $Mu \leq f$ *almost everywhere in* U,
(iii) $\max(Mu - f, u - g) = 0$ *almost everywhere in* U,
(iv) $u = g$ *on* ∂U.

In general $u \notin C^2(U)$.

The idea of the proof is to approximate (15) by a *penalized problem* of this form:

$$\begin{cases} Mu^\varepsilon + \beta_\varepsilon(u^\varepsilon - g) = f & \text{in } U \\ u^\varepsilon = g & \text{on } \partial U, \end{cases}$$

where $\beta_\varepsilon : \mathbb{R} \to \mathbb{R}$ is a smooth, convex function, $\beta_\varepsilon' \geq 0$, and $\beta_\varepsilon \equiv 0$ for $x \leq 0$, $\lim_{\epsilon \to 0} \beta_\varepsilon(x) = \infty$ for $x > 0$. Then $u^\varepsilon \to u$. It will in practice be difficult to find a precise formula for u, but computers can provide accurate numerical approximations.

DESIGNING AN OPTIMAL STOPPING POLICY. Now we show that our solution of (15) is in fact the value function, and along the way we will learn how to design an optimal stopping strategy θ^*.

First note that the *stopping set*

$$S := \{x \in U \mid u(x) = g(x)\}$$

is closed. Define for each $x \in \bar{U}$,

$$\theta^* = \text{first hitting time of } S.$$

THEOREM. *Let u be the solution of* (15). *Then*

$$u(x) = J_x(\theta^*) = \inf_\theta J_x(\theta)$$

for all $x \in \bar{U}$.

This says that we should first compute the solution to (15) to find S, define θ^* as above, and then we should run $\mathbf{X}(\cdot)$ until it hits S (or else exits from U).

PROOF. 1. Define the *continuation set*

$$C := U - S = \{x \in U \mid u(x) < g(x)\}.$$

On this set $Lu = f$, and furthermore $u = g$ on ∂C. Since $\tau \wedge \theta^*$ is the exit time from C, we have for $x \in C$

$$u(x) = E(\int_0^{\tau \wedge \theta^*} f(\mathbf{X}(s))\, ds + g(\mathbf{X}(\theta^* \wedge \tau))) = J_x(\theta^*).$$

On the other hand, if $x \in S$, $\tau \wedge \theta^* = 0$; and so

$$u(x) = g(x) = J_x(\theta^*).$$

Thus for all $x \in \bar{U}$, we have $u(x) = J_x(\theta^*)$.

2. Now let θ be any other stopping time. We need to show

$$u(x) = J_x(\theta^*) \leq J_x(\theta).$$

Now by Itô's formula

$$u(x) = E(\int_0^{\tau \wedge \theta} Mu(\mathbf{X})\, ds + u(\mathbf{X}(\tau \wedge \theta)))$$

But $Mu \leq f$ and $u \leq g$ in $\bar{U}$. Hence

$$u(x) \leq E(\int_0^{\tau \wedge \theta} f(\mathbf{X})\, ds + g(\mathbf{X}(\tau \wedge \theta))) = J_x(\theta).$$

But since $u(x) = J_x(\theta^*)$, we consequently have

$$u(x) = J_x(\theta^*) = \min_\theta J_x(\theta),$$

as asserted. □

D. OPTIONS PRICING

In this section we outline an application to mathematical finance, mostly following Baxter–Rennie [**B-R**] and the class lectures of L. Goldberg. Another basic reference is Hull [**Hu**].

THE BASIC PROBLEM. Let us consider a given *security*, say a stock, whose price at time t is $S(t)$. We suppose that S evolves according to the SDE introduced in Chapter 5:

$$\begin{cases} dS = \mu S dt + \sigma S dW \\ S(0) = s_0, \end{cases} \tag{16}$$

where $\mu > 0$ is the drift and $\sigma \neq 0$ the volatility. The initial price s_0 is known.

A *derivative* is a financial instrument whose payoff depends upon (i.e., is derived from) the behavior of $S(\cdot)$. We will investigate a *European call option*, which is the right to buy one share of the stock S, at the price p at time T. The number p is called the *strike price* and $T > 0$ the *strike* (or *expiration*) *time*. The basic question is this:

What is the "proper price" at time $t = 0$ of this option?

In other words, if you run a financial firm and wish to sell your customers this call option, how much should you charge? (We are looking for the "break–even" price, for which the firm neither makes nor loses money.)

ARBITRAGE AND HEDGING. To simplify, we assume hereafter that the prevailing, no-risk interest rate is the constant $r > 0$. This means that \$1 put in a bank at time $t = 0$ becomes $\$e^{rT}$ at time $t = T$. Equivalently, \$1 at time $t = T$ is worth only $\$e^{-rT}$ at time $t = 0$.

As for the problem of pricing our call option, a first guess might be that the proper price should be

$$e^{-rT}E((S(T) - p)^+), \tag{17}$$

for $x^+ := \max(x, 0)$. The reasoning behind this guess is that if $S(T) < p$, then the option is worthless. If $S(T) > p$, we can buy a share for the price p, immediately sell at price $S(T)$, and thereby make a profit of $(S(T) - p)^+$. We average this over all sample paths and multiply by the discount factor e^{-rT}, to arrive at (17).

As reasonable as this may all seem, (17) is in fact *not* the proper price. Other forces are at work in financial markets. Indeed the fundamental factor in options pricings is *arbitrage*, meaning the possibility of risk-free profits.

We must price our option so as to create no arbitrage opportunities for others.

To convert this principle into mathematics, we introduce also the notion of *hedging*. This means somehow eliminating our risk as the seller of the call option. The exact details appear below, but the basic idea is that we can in effect "duplicate" our option by a portfolio consisting of (continually changing) holdings of a risk–free bond and of the stock on which the call is written.

A PARTIAL DIFFERENTIAL EQUATION. We demonstrate next how use these principles to convert our pricing problem into a PDE. We introduce for $s \geq 0$ and $0 \leq t \leq T$, the unknown *price function*

(18)

$u(s, t)$, denoting the proper price of the option at time t, given that $S(t) = s$.

Then $u(s_0, 0)$ is the price we are seeking.

Boundary conditions. We need to calculate u. For this, notice first that at the expiration time T, we have

$$u(s, T) = (s - p)^+ \qquad (s \geq 0). \tag{19}$$

Furthermore, if $s = 0$, then $S(t) = 0$ for all $0 \leq t \leq T$ and so

$$u(0, t) = 0 \qquad (0 \leq t \leq T). \tag{20}$$

We seek how a PDE u solves for $s > 0$, $0 \leq t \leq T$.

Duplicating an option, self-financing. To go further, define the process

$$C(t) := u(S(t), t) \qquad (0 \leq t \leq T). \tag{21}$$

Thus $C(t)$ is the current price of the option at time t, and is random since the stock price $S(t)$ is random. According to Itô's formula and (16)

$$\begin{aligned} dC &= u_t dt + u_s dS + \frac{1}{2} u_{ss} (dS)^2 \\ &= (u_t + \mu S u_s + \frac{\sigma^2}{2} S^2 u_{ss}) dt + \sigma S u_s dW. \end{aligned} \tag{22}$$

Now comes the key idea: we propose to "duplicate" C by a portfolio consisting of shares of S and of a *bond* B. More precisely, assume that B is a risk-free investment, which therefore grows at the prevailing interest rate r:

$$\begin{cases} dB = rBdt \\ B(0) = 1. \end{cases} \tag{23}$$

This just means $B(t) = e^{rt}$, of course. We will try to find processes ϕ and ψ so that

$$C = \phi S + \psi B \qquad (0 \leq t \leq T). \tag{24}$$

Discussion. The point is that if we can construct ϕ, ψ so that (24) holds, we can eliminate all risk. To see this more clearly, imagine that your financial firm sells a call option, as above. The firm thereby incurs the risk that at time T, the stock price $S(T)$ will exceed p, and so the buyer will exercise the option. But if in the meantime the firm has constructed the portfolio (24), the profits from it will exactly equal the funds needed to pay the customer. Conversely, if the option is worthless at time T, the portfolio will have no profit. □

But to make this work, the financial firm should not have to inject any new money into the hedging scheme, beyond the initial investment to set it up. We ensure this by requiring that the portfolio represented on the right-hand side of (24) be *self-financing*. This means that the changes in the value of the portfolio should depend only upon the changes in S, B. We therefore require that

$$dC = \phi dS + \psi dB \qquad (0 \leq t \leq T). \tag{25}$$

Remark (discrete version of self-financing). Roughly speaking, a portfolio is self-financing if it is financially self contained. To understand this better, let us consider a different model in which time is discrete, and the values of the stock and bond at a time t_i are given by S_i and B_i respectively. Here $\{t_i\}_{i=0}^N$ is an increasing sequence of times and we suppose that each time step $t_{i+1} - t_i$ is small. A portfolio can now be thought of as a sequence $\{(\phi_i, \psi_i)\}_{i=0}^N$, corresponding to our changing holdings of the stock S and the bond B over each time interval.

Now for a given time interval (t_i, t_{i+1}), $C_i = \phi_i S_i + \psi_i B_i$ is the opening value of the portfolio and $C_{i+1} = \phi_i S_{i+1} + \psi_i B_{i+1}$ represents the closing value. The self-financing condition means that the financing gap $C_{i+1} - C_i$ of cash (that would otherwise have to be injected to pay for our construction strategy) must be zero.This is equivalent to saying that

$$C_{i+1} - C_i = \phi_i (S_{i+1} - S_i) + \psi_i (B_{i+1} - B_i),$$

the continuous version of which is condition (25). □

Combining formulas (22), (23) and (25) provides the identity

$$(u_t + \mu S u_s + \frac{\sigma^2}{2} S^2 u_{ss})dt + \sigma S u_s dW \\ = \phi(\mu S dt + \sigma S dW) + \psi r B dt. \tag{26}$$

So if (24) holds, (26) must be valid, and we are trying to select ϕ, ψ to make all this so. We observe in particular that the terms multiplying dW on each side of (26) will match provided we take

$$\phi(t) := u_s(S(t), t) \qquad (0 \le t \le T). \tag{27}$$

Then (26) simplifies, to read

$$(u_t + \frac{\sigma^2}{2} S^2 u_{ss})dt = r\psi B dt.$$

But $\psi B = C - \phi S = u - u_s S$, according to (24), (27). Consequently,

$$(u_t + rSu_s + \frac{\sigma^2}{2} S^2 u_{ss} - ru)dt = 0. \tag{28}$$

The argument of u and its partial derivatives is $(S(t), t)$.

Consequently, to make sure that (21) is valid, we ask that the function $u = u(s,t)$ solve the *Black–Scholes–Merton PDE*

$$u_t + rsu_s + \frac{\sigma^2}{2} s^2 u_{ss} - ru = 0 \qquad (0 \le t \le T). \tag{29}$$

The main outcome of all our financial reasoning is the derivation of this partial differential equation. Observe that the parameter μ does not appear.

More on self-financing. Before going on, we return to the self-financing condition (25). The Itô product rule and (24) imply

$$dC = \phi dS + \psi dB + S d\phi + B d\psi + d\phi\, dS.$$

To ensure (25), we consequently must make sure that

$$S d\phi + B d\psi + d\phi\, dS = 0, \tag{30}$$

where we recall $\phi = u_s(S(t), t)$. Now $d\phi\, dS = \sigma^2 S^2 u_{ss} dt$. Thus (30) is valid provided

$$d\psi = -B^{-1}(S d\phi + \sigma^2 S^2 u_{ss} dt). \tag{31}$$

We can confirm this by noting that (24), (27) imply

$$\psi = B^{-1}(C - \phi S) = e^{-rt}(u(S,t) - u_s(S,t)S).$$

A direct calculation using (28) verifies (31).

SUMMARY. To price our call option, we solve the boundary-value problem

$$\begin{cases} u_t + rsu_s + \frac{\sigma^2}{2}s^2u_{ss} - ru = 0 & (s > 0, 0 \le t \le T) \\ u = (s-p)^+ & (s > 0, t = T) \\ u = 0 & (s = 0, 0 \le t \le T). \end{cases} \tag{32}$$

Remember that $u(s_0, 0)$ is the price we are trying to find. It turns out that this problem can be solved explicitly, although we omit the details here: see for instance Baxter–Rennie [**B-R**].

E. THE STRATONOVICH INTEGRAL

We next discuss the *Stratonovich stochastic calculus*, which is an alternative to Itô's approach. Most of the following material is from Arnold [**A, Chapter 10**].

1. Motivation. Let us consider first of all the *formal* random differential equation

$$\begin{cases} \dot{X} = d(t)X + f(t)X\xi \\ X(0) = X_0, \end{cases} \tag{33}$$

where $m = n = 1$ and $\xi(\cdot)$ is 1-dimensional "white noise". If we interpret this rigorously as the stochastic differential equation:

$$\begin{cases} dX = d(t)Xdt + f(t)XdW \\ X(0) = X_0, \end{cases} \tag{34}$$

we then recall from Chapter 5 that the unique solution is

$$X(t) = X_0 e^{\int_0^t d(s) - \frac{1}{2}f^2(s)\,ds + \int_0^t f(s)\,dW}. \tag{35}$$

On the other hand perhaps (33) is a proposed mathematical model of some physical process and we are not really sure whether $\xi(\cdot)$ is "really" white noise. It could perhaps be instead some process with smooth (but highly complicated) sample paths. How would this possibility change the solution?

APPROXIMATING WHITE NOISE. More precisely, suppose that $\{\xi^k(\cdot)\}_{k=1}^\infty$ is a sequence of stochastic processes satisfying:

(a) $E(\xi^k(t)) = 0$,
(b) $E(\xi^k(t)\xi^k(s)) := d^k(t-s)$,
(c) $\xi^k(t)$ is Gaussian for all $t \ge 0$,
(d) $t \mapsto \xi^k(t)$ is smooth for all ω,

where we suppose that the functions $d^k(\cdot)$ converge as $k \to \infty$ to δ_0, the Dirac measure at 0.

In light of the formal definition of the white noise $\xi(\cdot)$ as a Gaussian process with $E\xi(t) = 0$, $E(\xi(t)\xi(s)) = \delta_0(t-s)$, the $\xi^k(\cdot)$ are thus presumably smooth approximations of $\xi(\cdot)$.

LIMITS OF SOLUTIONS. Now consider the problem

$$\begin{cases} \dot{X}^k = d(t)X^k + f(t)X^k\xi^k \\ X^k(0) = X_0. \end{cases} \tag{36}$$

For each ω this is just a regular ODE, whose solution is

$$X^k(t) := X_0 e^{\int_0^t d(s)\,ds + \int_0^t f(s)\xi^k(s)\,ds}.$$

Next look at

$$Z^k(t) := \int_0^t f(s)\xi^k(s)\,ds.$$

For each time $t \geq 0$, this is a Gaussian random variable, with

$$E(Z^k(t)) = 0.$$

Furthermore,

$$\begin{aligned} E(Z^k(t)Z^k(s)) &= \int_0^t \int_0^s f(\tau)f(\sigma)d_k(\tau-\sigma)\,d\sigma d\tau \\ &\to \int_0^t \int_0^s f(\tau)f(\sigma)\delta_0(\tau-\sigma)\,d\sigma d\tau \\ &= \int_0^{t\wedge s} f^2(\tau)\,d\tau. \end{aligned}$$

Hence as $k \to \infty$, $Z^k(t)$ converges in L^2 to a process whose distributions agree with those $\int_0^t f(s)\,dW$. And therefore $X^k(t)$ converges to a process whose distributions agree with

$$\hat{X}(t) := X_0 e^{\int_0^t d(s)\,ds + \int_0^t f(s)\,dW}. \tag{37}$$

This does not agree with the solution (35)!

Discussion. Thus if we regard (33) as an Itô SDE with $\xi(\cdot)$ a "true" white noise, (35) is our solution. But if we approximate $\xi(\cdot)$ by smooth processes $\xi^k(\cdot)$, solve the approximate problems (36) and pass to limits with the approximate solutions $X^k(\cdot)$, we get a different solution. This means that (33) is *unstable* with respect to changes in the random term $\xi(\cdot)$. This conclusion has important consequences in questions of modeling, since it may be unclear experimentally whether we really have $\xi(\cdot)$ or instead $\xi^k(\cdot)$ in (33) and similar problems.

In view of all this, it is appropriate to ask if there is some way to redefine the stochastic integral so these difficulties do not come up. One answer is the *Stratonovich integral*.

2. Definition of Stratonovich integral.

A one-dimensional example. Recall that in Chapter 4 we defined for 1-dimensional Brownian motion

$$\int_0^T W\,dW := \lim_{|P^n|\to 0} \sum_{k=0}^{m_n-1} W(t_k^n)(W(t_{k+1}^n) - W(t_k^n)) = \frac{W^2(T)-T}{2},$$

where $P^n := \{0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = T\}$ is a partition of $[0,T]$. This corresponds to a sequence of Riemann sum approximations, where the integrand is evaluated at the *left-hand endpoint* of each subinterval $[t_k^n, t_{k+1}^n]$.

The corresponding *Stratonovich integral* is instead defined this way:

$$\int_0^T W \circ dW := \lim_{|P^n| \to 0} \sum_{k=0}^{m_n - 1} \left(\frac{W(t_{k+1}^n) + W(t_k^n)}{2} \right) (W(t_{k+1}^n) - W(t_k^n)) = \frac{W^2(T)}{2}.$$

(Observe the notational change: we hereafter write a small circle before the dW to signify the Stratonovich integral.) According to calculations in Chapter 4, we also have

$$\int_0^T W \circ dW = \lim_{|P^n| \to 0} \sum_{k=0}^{m_n - 1} W \left(\frac{t_{k+1}^n + t_k^n}{2} \right) (W(t_{k+1}^n) - W(t_k^n)).$$

Therefore for this case the Stratonovich integral corresponds to a Riemann sum approximation where we evaluate the integrand at the *midpoint* of each subinterval $[t_k^n, t_{k+1}^n]$. □

We generalize this example and so introduce the

DEFINITION. Let $\mathbf{W}(\cdot)$ be an n-dimensional Brownian motion and let $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n \times n}$ be a C^1 function such that

$$E\left(\int_0^T |\mathbf{B}(\mathbf{W}, t)|^2 \, dt \right) < \infty.$$

Then we define

$$\int_0^T \mathbf{B}(\mathbf{W}, t) \circ d\mathbf{W} := \lim_{|P^n| \to 0} \sum_{k=0}^{m_n - 1} \mathbf{B} \left(\frac{\mathbf{W}(t_{k+1}^n) + \mathbf{W}(t_k^n)}{2}, t_k^n \right) (\mathbf{W}(t_{k+1}^n) - \mathbf{W}(t_k^n)).$$

It can be shown that this limit exists in $L^2(\Omega)$.

A CONVERSION FORMULA. Remember that Itô's integral can be computed this way:

$$\int_0^T \mathbf{B}(\mathbf{W}, t) \, d\mathbf{W} = \lim_{|P^n| \to 0} \sum_{k=0}^{m_n - 1} \mathbf{B}(\mathbf{W}(t_k^n), t_k^n)(\mathbf{W}(t_{k+1}^n) - \mathbf{W}(t_k^n)).$$

This is in general not equal to the Stratonovich integral, but there is a *conversion formula*

$$\left[\int_0^T \mathbf{B}(\mathbf{W}, t) \circ d\mathbf{W} \right]^i = \left[\int_0^T \mathbf{B}(\mathbf{W}, t) \, d\mathbf{W} \right]^i + \frac{1}{2} \int_0^T \sum_{j=1}^n \frac{\partial b^{ij}}{\partial x_j}(\mathbf{W}, t) \, dt, \tag{38}$$

for $i = 1, \dots, n$. Here v^i means the i^{th}–component of the vector function $\mathbf{v}$. This formula is proved by noting

$$\int_0^T \mathbf{B}(\mathbf{W}, t) \circ d\mathbf{W} - \int_0^T \mathbf{B}(\mathbf{W}, t)\, d\mathbf{W}$$
$$= \lim_{|P^n| \to 0} \sum_{k=0}^{m_n - 1} \left[\mathbf{B}\left(\frac{\mathbf{W}(t_{k+1}^n) + \mathbf{W}(t_k^n)}{2}, t_k^n \right) - \mathbf{B}(\mathbf{W}(t_k^n), t_k^n) \right]$$
$$\cdot (\mathbf{W}(t_{k+1}^n) - \mathbf{W}(t_k^n))$$

and using the Mean Value Theorem plus some usual methods for evaluating the limit. We omit details. □

Special case. If $n = 1$, then

$$\int_0^T b(W, t) \circ dW = \int_0^T b(W, t)\, dW + \frac{1}{2} \int_0^T \frac{\partial b}{\partial x}(W, t)\, dt.$$

□

Assume now $\mathbf{B} : \mathbb{R}^n \times [0, T] \to \mathbb{M}^{n \times m}$ and $\mathbf{W}(\cdot)$ is an m-dimensional Brownian motion. We make this informal

DEFINITION. If $\mathbf{X}(\cdot)$ is a stochastic process with values in $\mathbb{R}^n$, we define

$$\int_0^T \mathbf{B}(\mathbf{X}, t) \circ d\mathbf{W} := \lim_{|P^n| \to 0} \sum_{k=0}^{m_n - 1} \mathbf{B}\left(\frac{\mathbf{X}(t_{k+1}^n) + \mathbf{X}(t_k^n)}{2}, t_k^n \right) (\mathbf{W}(t_{k+1}^n) - \mathbf{W}(t_k^n)),$$

provided this limit exists in $L^2(\Omega)$ for all sequences of partitions P^n, with $|P^n| \to 0$.

3. Stratonovich chain rule.

DEFINITION. Suppose that the process $\mathbf{X}(\cdot)$ solves the Stratonovich integral equation

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{b}(\mathbf{X}, s)\, ds + \int_0^t \mathbf{B}(\mathbf{X}, s) \circ d\mathbf{W} \quad (0 \le t \le T)$$

for $\mathbf{b} : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ and $\mathbf{B} : \mathbb{R}^n \times [0, T] \to \mathbb{M}^{n \times m}$. We then write

$$d\mathbf{X} = \mathbf{b}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t) \circ d\mathbf{W},$$

the second term on the right being the *Stratonovich stochastic differential*.

THEOREM (Stratonovich chain rule). *Assume*

$$d\mathbf{X} = \mathbf{b}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t) \circ d\mathbf{W}$$

and suppose $u : \mathbb{R}^n \times [0, T] \to \mathbb{R}$ *is smooth. Define*

$$Y(t) := u(\mathbf{X}(t), t).$$

Then

$$\begin{aligned} dY &= \frac{\partial u}{\partial t}dt + \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \circ d\mathbf{X}^i \\ &= \left(\frac{\partial u}{\partial t} + \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} b^i \right) dt + \sum_{i=1}^{n} \sum_{k=1}^{m} \frac{\partial u}{\partial x_i} b^{ik} \circ d\mathbf{W}^k. \end{aligned}$$

Thus the ordinary chain rule holds for Stratonovich stochastic differentials, and there is no additional term involving $\frac{\partial^2 u}{\partial x_i \partial x_j}$ as there is for Itô's formula. We omit the proof, which is similar to that for the Itô rule. The main difference is that we make use of the formula $\int_0^T W \circ dW = \frac{1}{2}W^2(T)$ in the approximations.

More discussion. Next let us return to the motivational example we began with. We have seen that if the differential equation (33) is interpreted to mean

$$\begin{cases} dX = d(t)X dt + f(t)X dW \quad \text{(Itô's sense)}, \\ X(0) = X_0, \end{cases}$$

then

$$X(t) = X_0 e^{\int_0^t d(s) - \frac{1}{2}f^2(s)\, ds + \int_0^t f(s)\, dW}.$$

However, if we interpret (33) to mean

$$\begin{cases} dX = d(t)X dt + f(t)X \circ dW \quad \text{(Stratonovich's sense)} \\ X(0) = X_0, \end{cases}$$

the solution is

$$\tilde{X}(t) = X_0 e^{\int_0^t d(s)\, ds + \int_0^t f(s)\, dW},$$

as is easily checked using the Stratonovich calculus described above.

This solution $\tilde{X}(\cdot)$ *is also the solution obtained by approximating the "white noise"* $\xi(\cdot)$ *by smooth processes* $\xi^k(\cdot)$ *and passing to limits.* This suggests that interpreting (16) and similar formal random differential equations in the Stratonovich sense will provide solutions which are stable with respect to perturbations in the random terms. This is indeed the case: See the articles [**S1-2**] by Sussmann.

Note also that these considerations clarify a bit the problems of interpreting mathematically the *formal* random differential equation (33), but do not say which interpretation is physically correct. This is a question of modeling and is not, strictly speaking, a mathematical issue.

CONVERSION RULES FOR SDE.

Let $\mathbf{W}(\cdot)$ be an m-dimensional Wiener process and suppose $\mathbf{b} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$, $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n\times m}$ satisfy the hypotheses of the basic existence and uniqueness theorem. Then $\mathbf{X}(\cdot)$ solves the *Itô* stochastic differential equation

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases}$$

if and only if $\mathbf{X}(\cdot)$ solves the *Stratonovich* stochastic differential equation

$$\begin{cases} d\mathbf{X} = \left[\mathbf{b}(\mathbf{X},t) - \frac{1}{2}\mathbf{c}(\mathbf{X},t)\right] dt + \mathbf{B}(\mathbf{X},t) \circ d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0, \end{cases}$$

for

$$c^i(x,t) = \sum_{k=1}^{m}\sum_{j=1}^{n} \frac{\partial b^{ik}}{\partial x_j}(x,t)b^{jk}(x,t) \qquad (1 \le i \le n).$$

A special case. For $m = n = 1$, this says

$$dX = b(X)dt + \sigma(X)dW$$

if and only if

$$dX = (b(X) - \frac{1}{2}\sigma'(X)\sigma(X))dt + \sigma(X) \circ dW.$$

4. Summary We conclude these lectures by summarizing the advantages of each definition of the stochastic integral:

Advantages of Itô integral

1. Simple formulas: $E\left(\int_0^t G\,dW\right) = 0$, $E\left(\left(\int_0^t G\,dW\right)^2\right) = E\left(\int_0^t G^2\,dt\right)$.
2. $I(t) = \int_0^t G\,dW$ is a martingale.

Advantages of Stratonovich integral

1. Ordinary chain rule holds.
2. Solutions of stochastic differential equations interpreted in Stratonovich sense are stable with respect to changes in random terms.

APPENDICES

Appendix A: Proof of Laplace–De Moivre Theorem (from §G in Chapter 2)

PROOF. 1. Set $S_n^* := \frac{S_n - np}{\sqrt{npq}}$, this being a random variable taking on the value $x_k = \frac{k-np}{\sqrt{npq}}$ $(k = 0, \ldots, n)$ with probability $p_n(k) = \binom{n}{k} p^k q^{n-k}$.

Look at the interval $\left[\frac{-np}{\sqrt{npq}}, \frac{nq}{\sqrt{npq}}\right]$. The points x_k divide this interval into n subintervals of length

$$h := \frac{1}{\sqrt{npq}}.$$

Now if n goes to ∞, and at the same time k changes so that $|x_k|$ is bounded, then

$$k = np + x_k\sqrt{npq} \to \infty$$

and

$$n - k = nq - x_k\sqrt{npq} \to \infty.$$

2. We recall next Stirling's formula, which says says

$$n! = e^{-n}n^n\sqrt{2\pi n}\,(1 + o(1)) \quad \text{as } n \to \infty,$$

where "$o(1)$" denotes a term which goes to 0 as $n \to \infty$. (See Mermin [**M**] for a nice discussion.) Hence as $n \to \infty$

$$\begin{aligned} p_n(k) &= \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= \frac{e^{-n}n^n\sqrt{2\pi n} p^k q^{n-k}}{e^{-k}k^k\sqrt{2\pi k}e^{-(n-k)}(n-k)^{(n-k)}\sqrt{2\pi(n-k)}}(1 + o(1)) \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}(1 + o(1)). \end{aligned} \tag{1}$$

3. Observe next that if $x = x_k = \frac{k-np}{\sqrt{npq}}$, then

$$1 + \sqrt{\frac{q}{np}}x = 1 + \sqrt{\frac{q}{np}}\left(\frac{k - np}{\sqrt{npq}}\right) = \frac{k}{np}$$

and

$$1 - \sqrt{\frac{p}{nq}}x = \frac{n-k}{nq}.$$

Note also $\log(1 \pm y) = \pm y - \frac{y^2}{2} + O(y^3)$ as $y \to 0$. Hence

$$\begin{aligned}\log\left(\frac{np}{k}\right)^k &= -k\log\left(\frac{k}{np}\right)\\ &= -k\log\left(1+\sqrt{\frac{q}{np}}x\right)\\ &= -(np + x\sqrt{npq})\left(\sqrt{\frac{q}{np}}x - \frac{q}{2np}x^2\right) + O\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

Similarly,

$$\log\left(\frac{nq}{n-k}\right)^{n-k} = -(nq - x\sqrt{npq})\left(-\sqrt{\frac{p}{nq}}x - \frac{p}{2nq}x^2\right) + O\left(n^{-\frac{1}{2}}\right).$$

Add these expressions and simplify, to discover

$$\lim_{\substack{n\to\infty \\ \frac{k-np}{\sqrt{npq}}\to x}} \log\left(\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}\right) = -\frac{x^2}{2}.$$

Consequently

$$\lim_{\substack{n\to\infty \\ \frac{k-np}{\sqrt{npq}}\to x}} \left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} = e^{-\frac{x^2}{2}}. \tag{2}$$

4. Finally, observe

$$\sqrt{\frac{n}{k(n-k)}} = \frac{1}{\sqrt{npq}}(1+o(1)) = h(1+o(1)), \tag{3}$$

since $k = np + x\sqrt{npq}, n-k = nq - x\sqrt{npq}$.

Now

$$P(a \le S_n^* \le b) = \sum_{\substack{a\le x_k\le b \\ x_k = \frac{k-np}{\sqrt{npq}}}} p_n(k)$$

for $a < b$. In view of $(1)-(3)$, the latter expression is a Riemann sum approximation as $n \to \infty$ of the integral

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

□

Appendix B: Proof of discrete martingale inequalities (from §I in Chapter 2)

PROOF. 1. Define

$$A_k := \bigcap_{j=1}^{k-1} \{X_j \le \lambda\} \cap \{X_k > \lambda\} \quad (k = 1, \dots, n).$$

Then

$$A := \left\{ \max_{1 \le k \le n} X_k > \lambda \right\} = \underbrace{\bigcup_{k=1}^{n} A_k}_{\text{disjoint union}} .$$

Since $\lambda P(A_k) \le \int_{A_k} X_k \, dP$, we have

$$\lambda P(A) = \lambda \sum_{k=1}^{n} P(A_k) \le \sum_{k=1}^{n} E(\chi_{A_k} X_k). \tag{4}$$

Therefore

$$\begin{aligned}
E(X_n^+) &\ge \sum_{k=1}^{n} E(X_n^+ \chi_{A_k}) \\
&= \sum_{k=1}^{n} E(E(X_n^+ \chi_{A_k} \mid X_1, \dots, X_k)) \\
&= \sum_{k=1}^{n} E(\chi_{A_k} E(X_n^+ \mid X_1, \dots, X_k)) \\
&\ge \sum_{k=1}^{n} E(\chi_{A_k} E(X_n \mid X_1, \dots, X_k)) \\
&\ge \sum_{k=1}^{n} E(\chi_{A_k} X_k) \quad \text{by the submartingale property} \\
&\ge \lambda P(A) \quad \text{by (4).}
\end{aligned}$$

2. Notice next that the proof above in fact demonstrates

$$\lambda P\left(\max_{1 \le k \le n} X_k > \lambda \right) \le \int_{\left\{ \max_{1 \le k \le n} X_k > \lambda \right\}} X_n^+ \, dP.$$

Apply this to the submartingale $|X_k|$:

$$\lambda P(X > \lambda) \le \int_{\{X > \lambda\}} Y \, dP, \tag{5}$$

for $X := \max_{1\le k\le n} |X_k|$, $Y := |X_n|$. Now take some $1 < p < \infty$. Then

$$\begin{aligned}
E(|X|^p) &= -\int_0^\infty \lambda^p \, dP(\lambda) \quad \text{for } P(\lambda) := P(X > \lambda) \\
&= p \int_0^\infty \lambda^{p-1} P(\lambda) \, d\lambda \\
&\le p \int_0^\infty \lambda^{p-1} \left(\frac{1}{\lambda} \int_{\{X>\lambda\}} Y \, dP \right) d\lambda \quad \text{by (5)} \\
&= p \int_\Omega Y \left(\int_0^X \lambda^{p-2} d\lambda \right) dP \\
&= \frac{p}{p-1} \int_\Omega Y X^{p-1} \, dP \\
&\le \frac{p}{p-1} \left(\int_\Omega Y^p \, dP \right)^{1/p} \left(\int_\Omega X^p \, dP \right)^{1-1/p}.
\end{aligned}$$

□

Appendix C: Proof of continuity of indefinite Itô integral (from §C in Chapter 4)

PROOF. We will assume assertion (i) of the Theorem in §C of Chapter 4, which states that the indefinite integral $I(\cdot)$ is a martingale.

There exist step processes $G^n \in \mathbb{L}^2(0,T)$, such that

$$E\left(\int_0^T (G^n - G)^2 \, dt \right) \to 0.$$

Write $I^n(t) := \int_0^t G^n \, dW$, for $0 \le t \le T$. If $G^n(s) \equiv G_k^n$ for $t_k^n \le s < t_{k+1}^n$, then

$$I^n(t) = \sum_{i=0}^{k-1} G_i^n (W(t_{i+1}^n) - W(t_i^n)) + G_k^n (W(t) - W(t_k^n))$$

for $t_k^n \le t < t_{k+1}^n$. Therefore $I^n(\cdot)$ has continuous sample paths a.s., since Brownian motion does. Since $I^n(\cdot)$ is a martingale, it follows that $|I^n - I^m|^2$ is a submartingale. The martingale inequality now implies

$$\begin{aligned}
P\left(\sup_{0\le t\le T} |I^n(t) - I^m(t)| > \varepsilon \right) &= P\left(\sup_{0\le t\le T} |I^n(t) - I^m(t)|^2 > \varepsilon^2 \right) \\
&\le \frac{1}{\varepsilon^2} E(|I^n(T) - I^m(T)|^2) \\
&= \frac{1}{\varepsilon^2} E\left(\int_0^T |G^n - G^m|^2 \, dt \right).
\end{aligned}$$

Choose $\varepsilon = \frac{1}{2^k}$. Then there exists n_k such that

$$P\left(\sup_{0\le t\le T}|I^n(t)-I^m(t)|>\frac{1}{2^k}\right)\le 2^{2k}E\left(\int_0^T|G^n(t)-G^m(t)|^2dt\right)$$
$$\le\frac{1}{k^2}\qquad\text{for } m,n\ge n_k.$$

We may assume $n_{k+1}\ge n_k\ge n_{k-1}\ge\ldots$, and $n_k\to\infty$. Let

$$A_k:=\left\{\sup_{0\le t\le T}|I^{n_k}(t)-I^{n_{k+1}}(t)|>\frac{1}{2^k}\right\}.$$

Then

$$P(A_k)\le\frac{1}{k^2}.$$

Thus by the Borel–Cantelli Lemma, $P(A_k \text{ i.o.})=0$; which is to say, for almost all ω

$$\sup_{0\le t\le T}|I^{n_k}(t,\omega)-I^{n_{k+1}}(t,\omega)|\le\frac{1}{2^k}\qquad\text{provided } k\ge k_0(\omega).$$

Hence $I^{n_k}(\cdot,\omega)$ converges uniformly on $[0,T]$ for almost every ω, and therefore $J(t,\omega):=\lim_{k\to\infty}I^{n_k}(t,\omega)$ is continuous for amost every ω. As $I^n(t)\to I(t)$ in $L^2(\Omega)$ for all $0\le t\le T$, we deduce as well that $J(t)=I(t)$ amost every for all $0\le t\le T$. In other words, $J(\cdot)$ is a version of $I(\cdot)$. Since for almost every ω, $J(\cdot,\omega)$ is the uniform limit of continuous functions, $J(\cdot)$ has continuous sample paths a.s. □

EXERCISES

(1) Show, using the formal manipulations for Itô's formula discussed in Chapter 1, that

$$Y(t) := e^{W(t)-\frac{t}{2}}$$

solves the stochastic differential equation

$$\begin{cases} dY = Y dW, \\ Y(0) = 1. \end{cases}$$

(Hint: If $X(t) := W(t) - \frac{t}{2}$, then $dX = -\frac{dt}{2} + dW$.)

(2) Show that

$$P(t) = p_0 e^{\sigma W(t) + \left(\mu - \frac{\sigma^2}{2}\right)t},$$

solves

$$\begin{cases} dP = \mu P dt + \sigma P dW, \\ P(0) = p_0. \end{cases}$$

(3) Let Ω be any set and $\mathcal{A}$ any collection of subsets of Ω. Show that there exists a unique smallest σ-algebra $\mathcal{U}$ of subsets of Ω containing $\mathcal{A}$. We call $\mathcal{U}$ the σ-algebra *generated* by $\mathcal{A}$.
(Hint: Take the intersection of all the σ-algebras containing $\mathcal{A}$.)

(4) Let $X = \sum_{i=1}^{k} a_i \chi_{A_i}$ be a simple random variable, where the real numbers a_i are distinct, the events A_i are pairwise disjoint, and $\Omega = \cup_{i=1}^{k} A_i$. Let $\mathcal{U}(X)$ be the σ-algebra generated by X.

(i) Describe precisely which sets are in $\mathcal{U}(X)$.

(ii) Suppose the random variable Y is $\mathcal{U}(X)$-measurable. Show that Y is constant on each set A_i.

(iii) Show that therefore Y can be written as a function of X.

(5) Verify:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}, \quad \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = m,$$
$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2.$$

(6) (i) Suppose A and B are independent events in some probability space. Show that A^c and B are independent. Likewise, show that A^c and B^c are independent.

(ii) Suppose that $A_1, A_2, \ldots, A_m$ are disjoint events, each of positive probability, such that $\Omega = \cup_{j=1}^{m} A_j$. Prove *Bayes' formula*:

$$P(A_k \mid B) = \frac{P(B \mid A_k) P(A_k)}{\sum_{j=1}^{m} P(B \mid A_j) P(A_j)} \quad (k = 1, \ldots, m),$$

provided $P(B) > 0$.

(7) During the Fall, 1999 semester 105 women applied to UC Sunnydale, of whom 76 were accepted, and 400 men applied, of whom 230 were accepted.

During the Spring, 2000 semester, 300 women applied, of whom 100 were accepted, and 112 men applied, of whom 21 were accepted.

Calculate numerically

a. the probability of a female applicant being accepted during the fall,
b. the probability of a male applicant being accepted during the fall,
c. the probability of a female applicant being accepted during the spring,
d. the probability of a male applicant being accepted during the spring.

Consider now the total applicant pool for both semesters together, and calculate

e. the probability of a female applicant being accepted,
f. the probability of a male applicant being accepted.

Are the University's admission policies biased towards females? or males?

(8) Let X be a real–valued, $N(0,1)$ random variable, and set $Y := X^2$. Calculate the density g of the distribution function for Y.
(Hint: You must find g so that $P(-\infty < Y \leq a) = \int_{-\infty}^{a} g\,dy$ for all a.)

(9) Take $\Omega = [0,1] \times [0,1]$, with $\mathcal{U}$ the Borel sets and P Lebesgue measure. Let $g : [0,1] \to \mathbb{R}$ be a continuous function.

Define the random variables

$$X_1(\omega) := g(x_1), \quad X_2(\omega) := g(x_2) \quad \text{for } \omega = (x_1, x_2) \in \Omega.$$

Show that X_1 and X_2 are independent and identically distributed.

(10) (i) Let $(\Omega, \mathcal{U}, P)$ be a probability space and $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \ldots$ be events. Show that

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} P(A_m).$$

(Hint: Look at the disjoint events $B_n := A_{n+1} - A_n$.)

(ii) Likewise, show that if $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \ldots$, then

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} P(A_m).$$

(11) Let $f : [0,1] \to \mathbb{R}$ be continuous and define the *Bernstein polynomial*

$$b_n(x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$

Prove that $b_n \to f$ uniformly on $[0,1]$ as $n \to \infty$, by providing the details for the following steps.

(i) Since f is uniformly continuous, for each $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that $|f(x) - f(y)| \leq \epsilon$ if $|x - y| \leq \delta(\epsilon)$.

(ii) Given $x \in [0, 1]$, take a sequence of independent random variables X_k such that $P(X_k = 1) = x, P(X_k = 0) = 1 - x$. Write $S_n = X_1 + \cdots + X_n$. Then $b_n(x) = E(f(\frac{S_n}{n}))$.

(iii) Therefore

$$\begin{aligned} |b_n(x) - f(x)| &\leq E(|f(\frac{S_n}{n}) - f(x)|) \\ &= \int_A |f(\frac{S_n}{n}) - f(x)|\, dP + \int_{A^c} |f(\frac{S_n}{n}) - f(x)|\, dP, \end{aligned}$$

for $A = \{\omega \in \Omega \mid |\frac{S_n}{n} - x| \leq \delta(\epsilon)\}$.

(iv) Then show

$$|b_n(x) - f(x)| \leq \epsilon + \frac{2M}{\delta(\epsilon)^2} V(\frac{S_n}{n}) = \epsilon + \frac{2M}{n\delta(\epsilon)^2} V(X_1),$$

for $M = \max |f|$. Conclude that $b_n \to f$ uniformly.

(12) Let X and Y be independent random variables, and suppose that f_X and f_Y are the density functions for X, Y. Show that the density function for $X + Y$ is

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\, dy.$$

(Hint: If $g : \mathbb{R} \to \mathbb{R}$, we have

$$E(g(X + Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) g(x + y)\, dxdy,$$

where $f_{X,Y}$ is the joint density function of X, Y.)

(13) Let X and Y be two independent positive random variables, each with density

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Find the density of $X + Y$.

(14) Show that

$$\lim_{n\to\infty} \int_0^1 \int_0^1 \cdots \int_0^1 f(\frac{x_1 + \ldots x_n}{n})\, dx_1 dx_2 \ldots dx_n = f(\frac{1}{2})$$

for each continuous function f.
(Hint: $P(|\frac{x_1+\ldots x_n}{n} - \frac{1}{2}| > \epsilon) \leq \frac{1}{\epsilon^2} V(\frac{x_1+\ldots x_n}{n}) = \frac{1}{12\epsilon^2 n}$.)

(15) Prove that

(i) $E(E(X \mid \mathcal{V})) = E(X)$.

(ii) $E(X) = E(X \mid \mathcal{W})$, where $\mathcal{W} = \{\emptyset, \Omega\}$ is the trivial σ-algebra.

□

(16) Let X, Y be two real–valued random variables and suppose their joint distribution function has the density $f(x,y)$. Show that

$$E(X|Y) = \Phi(Y) \quad \text{a.s.}$$

for

$$\Phi(y) = \frac{\int_{-\infty}^{\infty} xf(x,y)\,dx}{\int_{-\infty}^{\infty} f(x,y)\,dx}.$$

(Hints: $\Phi(Y)$ is a function of Y and so is $\mathcal{U}(Y)$–measurable. Therefore we must show that

$$(*) \qquad \int_A X\,dP = \int_A \Phi(Y)\,dP \quad \text{for all } A \in \mathcal{U}(Y).$$

Now $A = Y^{-1}(B)$ for some Borel subset of $\mathbb{R}$. So the left hand side of $(*)$ is

$$(**) \qquad \int_A X\,dP = \int_\Omega \chi_B(Y) X\,dP = \int_{-\infty}^{\infty}\int_B xf(x,y)\,dydx.$$

The right hand side of $(*)$ is

$$\int_A \Phi(Y)\,dP = \int_{-\infty}^{\infty}\int_B \Phi(y)f(x,y)\,dydx,$$

which equals the right hand side of $(**)$. Fill in the details.)

(17) A smooth function $\Phi : \mathbb{R} \to \mathbb{R}$ is called *convex* if $\Phi''(x) \geq 0$ for all $x \in \mathbb{R}$.
(i) Show that if Φ is convex, then

$$\Phi(y) \geq \Phi(x) + \Phi'(x)(y-x) \quad \text{for all } x, y \in \mathbb{R}.$$

(ii) Show that

$$\Phi(\frac{x+y}{2}) \leq \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(y) \quad \text{for all } x, y \in \mathbb{R}.$$

(iii) A smooth function $\Phi : \mathbb{R}^n \to \mathbb{R}$ is called *convex* if the matrix $((\Phi_{x_i x_j}))$ is nonnegative definite for all $x \in \mathbb{R}^n$. (This means that $\sum_{i,j=1}^n \Phi_{x_i x_j}\xi_i\xi_j \geq 0$ for all $\xi \in \mathbb{R}^n$.) Prove

$$\Phi(y) \geq \Phi(x) + D\Phi(x)\cdot(y-x) \text{ and } \Phi(\frac{x+y}{2}) \leq \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(y)$$

for all $x, y \in \mathbb{R}^n$. (Here "D" denotes the gradient.)

(18) (i) Prove *Jensen's inequality*:

$$\Phi(E(X)) \leq E(\Phi(X))$$

for a random variable $X : \Omega \to \mathbb{R}$, where Φ is convex. (Hint: Use assertion (iii) from the previous problem.)
(ii) Prove the *conditional Jensen's inequality*:

$$\Phi(E(X|\mathcal{V})) \leq E(\Phi(X)|\mathcal{V}).$$

(19) Let $W(\cdot)$ be a one-dimensional Brownian motion. Show

$$E(W^{2k}(t)) = \frac{(2k)!t^k}{2^k k!}.$$

(20) Show that if $\mathbf{W}(\cdot)$ is an n-dimensional Brownian motion, then so are
(i) $\mathbf{W}(t+s) - \mathbf{W}(s)$ for all $s \geq 0$,
(ii) $c\mathbf{W}(t/c^2)$ for all $c > 0$ ("Brownian scaling").

(21) Let $W(\cdot)$ be a one-dimensional Brownian motion, and define

$$\bar{W}(t) := \begin{cases} tW(\frac{1}{t}) & \text{for } t > 0 \\ 0 & \text{for } t = 0. \end{cases}$$

Show that $\bar{W}(t) - \bar{W}(s)$ is $N(0, t-s)$ for times $0 \leq s \leq t$. ($\bar{W}(\cdot)$ also has independent increments and so is a one-dimensional Brownian motion. You do not need to show this.)

(22) Define $X(t) := \int_0^t W(s)\,ds$, where $W(\cdot)$ is a one-dimensional Brownian motion. Show that

$$E(X^2(t)) = \frac{t^3}{3} \quad \text{for each } t > 0.$$

(23) Define $X(t)$ as in the previous problem. Show that

$$E(e^{\lambda X(t)}) = e^{\frac{\lambda^2 t^3}{6}} \quad \text{for each } t > 0.$$

(Hint: $X(t)$ is a Gaussian random variable, the variance of which we know from the previous homework problem.)

(24) Define $U(t) := e^{-t}W(e^{2t})$, where $W(\cdot)$ is a one-dimensional Brownian motion. Show that

$$E(U(t)U(s)) = e^{-|t-s|} \quad \text{for all } -\infty < s, t < \infty.$$

(25) Let $W(\cdot)$ be a one-dimensional Brownian motion. Show that

$$\lim_{m\to\infty} \frac{W(m)}{m} = 0 \quad \text{almost surely.}$$

(Hint: Fix $\epsilon > 0$ and define the event $A_m := \{|\frac{W(m)}{m}| \geq \epsilon\}$. Then $A_m = \{|X| \geq \sqrt{m}\epsilon\}$ for the $N(0,1)$ random variable $X = \frac{W(m)}{\sqrt{m}}$. Apply the Borel–Cantelli Lemma.)

(26) (i) Let $0 < \gamma \leq 1$. Show that if $f : [0,T] \to \mathbb{R}^n$ is uniformly Hölder continuous with exponent γ, it is also is uniformly Hölder continuous with each exponent $0 < \delta < \gamma$.
(ii) Show that $f(t) = t^\gamma$ is uniformly Hölder continuous with exponent γ on the interval $[0,1]$.

(27) Let $0 < \gamma < \frac{1}{2}$. These notes show that if $W(\cdot)$ is a one–dimensional Brownian motion, then for almost every ω there exists a constant K, *depending on* ω, such that

$$(*) \qquad |W(t,\omega) - W(s,\omega)| \le K|t-s|^\gamma \quad \text{for all } 0 \le s,t \le 1.$$

Show that there does not exist a constant K such that $(*)$ holds for almost all ω.

(28) Prove that if $G, H \in \mathbb{L}^2(0,T)$, then

$$E\left(\int_0^T G\,dW \int_0^T H\,dW\right) = E\left(\int_0^T GH\,dt\right).$$

(Hint: $2ab = (a+b)^2 - a^2 - b^2$.)

(29) Let $(\Omega, \mathcal{U}, P)$ be a probability space, and take $\mathcal{F}(\cdot)$ to be a filtration of σ–algebras. Assume X be an integrable random variable, and define $X(t) := E(X|\mathcal{F}(t))$ for times $t \ge 0$.

Show that $X(\cdot)$ is a martingale.

(30) Show directly that $I(t) := W^2(t) - t$ is a martingale.
(Hint: $W^2(t) = (W(t)-W(s))^2 - W^2(s) + 2W(t)W(s)$. Take the conditional expectation with respect to $\mathcal{W}(s)$, the history of $W(\cdot)$, and then condition with respect to the history of $I(\cdot)$.)

(31) Suppose $X(\cdot)$ is a real-valued martingale and $\Phi : \mathbb{R} \to \mathbb{R}$ is convex. Assume also $E(|\Phi(X(t))|) < \infty$ for all $t \ge 0$. Show that

$$\Phi(X(\cdot)) \text{ is a submartingale.}$$

(Hint: Use the conditional Jensen's inequality.)

(32) Use the Itô chain rule to show that $Y(t) := e^{\frac{t}{2}} \cos(W(t))$ is a martingale.

(33) Let $\mathbf{W}(\cdot) = (W^1, \ldots, W^n)$ be an n-dimensional Brownian motion, and write $Y(t) := |\mathbf{W}(t)|^2 - nt$ for times $t \ge 0$. Show that $Y(\cdot)$ is a martingale.

(Hint: Compute dY.)

(34) Show that

$$\int_0^T W^2\,dW = \frac{1}{3}W^3(T) - \int_0^T W\,dt$$

and

$$\int_0^T W^3\,dW = \frac{1}{4}W^4(T) - \frac{3}{2}\int_0^T W^2\,dt.$$

(35) Recall from the notes that

$$Y := e^{\int_0^t g\,dW - \frac{1}{2}\int_0^t g^2\,ds}$$

satisfies

$$dY = gY\,dW.$$

Use this to prove

$$E(e^{\int_0^T g\,dW}) = e^{\frac{1}{2}\int_0^T g^2\,ds}.$$

(36) Let $u = u(x,t)$ be a smooth solution of the *backwards diffusion equation*

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0,$$

and suppose $W(\cdot)$ is a one-dimensional Brownian motion.

Show that for each time $t > 0$:

$$E(u(W(t),t)) = u(0,0).$$

(37) Calculate $E(B^2(t))$ for the Brownian bridge $B(\cdot)$, and show in particular that $E(B^2(t)) \to 0$ as $t \to 1^-$.

(38) Let X solve the Langevin equation, and suppose that X_0 is an $N(0, \frac{\sigma^2}{2b})$ random variable. Show that

$$E(X(s)X(t)) = \frac{\sigma^2}{2b}e^{-b|t-s|}.$$

(39) (i) Consider the ODE

$$\begin{cases} \dot{x} = x^2 \quad (t > 0) \\ x(0) = x_0. \end{cases}$$

Show that if $x_0 > 0$, the solution "blows up to infinity" in finite time.
(ii) Next, look at the ODE

$$\begin{cases} \dot{x} = x^{\frac{1}{2}} \quad (t > 0) \\ x(0) = 0. \end{cases}$$

Show that this problem has infinitely many solutions.
(Hint: $x \equiv 0$ is a solution. Find also a solution which is positive for times $t > 0$, and then combine these solutions to find ones which are zero for some time and then become positive.)

(40) (i) Use the substituion $X = u(W)$ to solve the SDE

$$\begin{cases} dX = -\frac{1}{2}e^{-2X}dt + e^{-X}dW \\ X(0) = x_0. \end{cases}$$

(ii) Show that the solution blows up at a finite, random time.

(41) Solve the SDE $dX = -Xdt + e^{-t}dW$.

(42) Let $\mathbf{W} = (W^1, W^2, \ldots, W^n)$ be an n-dimensional Brownian motion and write

$$R := |\mathbf{W}| = \left(\sum_{i=1}^n (W^i)^2\right)^{\frac{1}{2}}.$$

Show that R solves the stochastic *Bessel equation*

$$dR = \sum_{i=1}^{n} \frac{W^i}{R} dW^i + \frac{n-1}{2R} dt.$$

(43) (i) Show that $\mathbf{X} = (\cos(W), \sin(W))$ solves the SDE system

$$\begin{cases} dX^1 = -\frac{1}{2}X^1 dt - X^2 dW \\ dX^2 = -\frac{1}{2}X^2 dt + X^1 dW \end{cases}$$

(ii) Show also that if $\mathbf{X} = (X^1, X^2)$ is any other solution, then $|\mathbf{X}|$ is constant in time.

(44) Solve the system

$$\begin{cases} dX^1 = dt + dW^1 \\ dX^2 = X^1 dW^2, \end{cases}$$

where $\mathbf{W} = (W^1, W^2)$ is a Brownian motion.

(45) Solve

$$\begin{cases} dX^1 = X^2 dt + dW^1 \\ dX^2 = X^1 dt + dW^2. \end{cases}$$

(46) Solve

$$\begin{cases} dX = \frac{1}{2}\sigma'(X)\sigma(X)dt + \sigma(X)dW \\ X(0) = 0 \end{cases}$$

where W is a one–dimensional Brownian motion and σ is a smooth, positive function.
(Hint: Let $f(x) := \int_0^x \frac{dy}{\sigma(y)}$ and set $g := f^{-1}$, the inverse function of f. Show $X := g(W)$.)

(47) Let τ be the first time a one–dimensional Brownian motion hits the half-open interval $(a, b]$. Show τ is a stopping time.

(48) Let $\mathbf{W}$ denote an n–dimensional Brownian motion, for $n \geq 3$. Write $\mathbf{X} = \mathbf{W} + x_0$, where the point x_0 lies in the region $U = \{0 < R_1 < |x| < R_2\}$ Calculate explicitly the probability that $\mathbf{X}$ will hit the outer sphere $\{|x| = R_2\}$ before hitting the inner sphere $\{|x| = R_1\}$.
(Hint: Check that

$$\Phi(x) = \frac{1}{|x|^{n-2}}$$

satisfies $\Delta\Phi = 0$ for $x \neq 0$. Modify Φ to build a function u which equals 0 on the inner sphere and 1 on the outer sphere.)

References

[A] L. Arnold, *Stochastic Differential Equations: Theory and Applications*, Wiley, 1974.

[B-R] M. Baxter and A. Rennie, *Financial Calculus: An Introduction to Derivative Pricing*, Cambridge U. Press, 1996.

[B-S] A. N. Borodin and P. Salminen, *Handbook of Brownian Motion: Facts and Formulae*, Birkhauser, 1996.

[B] L. Breiman, *Probability*, Addison–Wesley, 1968.

[Br] P. Bremaud, *An Introduction to Probabilistic Modeling*, Springer, 1988.

[C] K. L. Chung, *Elementary Probability Theory with Stochastic Processes*, Springer, 1975.

[D] M. H. A. Davis, *Linear Estimation and Stochastic Control*, Chapman and Hall.

[F] A. Friedman, *Stochastic Differential Equations and Applications, Vol. 1 and 2*, Academic Press.

[Fr] M. Freidlin, *Functional Integration and Partial Differential Equations*, Princeton U. Press, 1985.

[G] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Springer, 1983.

[G-S] I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations*, Springer, 1972.

[G] D. Gillespie, *The mathematics of Brownian motion and Johnson noise*, American J. Physics **64** (1996), 225–240.

[H] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review **43** (2001), 525–546.

[Hu] J. C. Hull, *Options, Futures and Other Derivatives (4th ed)*, Prentice Hall, 1999.

[K] N. V. Krylov, *Introduction to the Theory of Diffusion Processes*, American Math Society, 1995.

[L1] J. Lamperti, *Probability*, Benjamin.

[L2] J. Lamperti, *A simple construction of certain diffusion processes*, J. Math. Kyôto (1964), 161–170.

[Ml] A. G. Malliaris, *Itô's calculus in financial decision making*, SIAM Review **25** (1983), 481–496.

[M] D. Mermin, *Stirling's formula!*, American J. Physics **52** (1984), 362–365.

[McK] H. McKean, *Stochastic Integrals*, Academic Press, 1969.

[N] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, 1967.

[O] B. K. Oksendal, *Stochastic Differential Equations: An Introduction with Applications, 4th ed.*, Springer, 1995.

[P-W-Z] R. Paley, N. Wiener, and A. Zygmund, *Notes on random functions*, Math. Z. **37** (1959), 647–668.

[P] M. Pinsky, *Introduction to Fourier Analysis and Wavelets*, Brooks/Cole, 2002.

[S] D. Stroock, *Probability Theory: An Analytic View*, Cambridge U. Press, 1993.

[S1] H. Sussmann, *An interpretation of stochastic differential equations as ordinary differential equations which depend on the sample point.*, Bulletin AMS **83** (1977), 296–298.

[S2] H. Sussmann, *On the gap between deterministic and stochastic ordinary differential equations*, Ann. Probability **6** (1978), 19–41.