LECTURE NOTES

On Basic

STOCHASTIC CALCULUS

By Patrik Albin

18th June 2001

Preface

These notes should be accessible for mathematically interested students, with a knowledge about undergraduate probability (e.g., [17, Chapters 1-5 and 7]), together with the most basic concepts of abstract Lebesgue integration (e.g., [34, pp. 5-25]). Knowledge about discrete time martingales (e.g., [8, Sections 9.3-9.4]) helps, but makes quite little difference. Same with elementary Markov theory (e.g., [17, Chapter 6]).

Stochastic Calculus is about manoeuvring around in the subject, with the help of "proofs", and one do not get there without reading such. On the other hand, traditional "state of the art" treatments of the subject (e.g., [22], [27], [31] and [33]) are too difficult with the preparations graduate students have today, and something less harsh is needed. The purpose of the notes is to present basic theory of Stochastic Calculus, with proofs of almost all statements, and yet in an economical and accessible way, with a minimum of technical details. But well worked out and explained such, and with probabilistic rather than mathematical arguments, when there is a choice. All proofs come with this focus, and are thus (close to) elementary.

Put simple, Stochastic Calculus concerns differentiating a stochastic process $\{X(t)\}_{t\geq 0}$ wrt. another $\{Y(t)\}_{t\geq 0}$. However, $X(t)$ and $Y(t)$ need not be differentiable in the usual sense (wrt. each other), and most often are not. The kind of process that can be differentiated in Stochastic Calculus is a *semimartingale*, and is composed of a process that is differentiable in the usual sense, together with a nondifferentiable component, a *local martingale*. The latter are the fundamental objects of Stochastic Calculus, and are introduced in Lecture 12. Preceeding lectures are preparatory. Lectures 1-5 treat basic facts crucial in the sequel about variation of functions, conditional expectations, Brownian motion and other Lévy processes, continuous time martingales and Markov processes, and strong Markov property. Further such facts, especially regarding martingales, are introduced in later lectures, when needed. Stochastic integrals are constructed in Lectures 6-11, and identified as local martingales in Lecture 12. The development of Stochastic Calculus starts in Lecture 13.

Of the many state of the art treatments of Stochastic Calculus, [9], [21], [22], [27], [33] and [36] have been our main sources of information. The need for Stochastic Calculus in applications to for example Mathematical Finance (e.g., [23, Preface]), has made non-technical knowledge of the subject desired (useful even), and inspired several texts without (most of) proofs. Among such, we have been influenced by [25].

Gothenburg 14 June 2001, *Patrik Albin*

Department of Mathematics
Chalmers University of Technology and Gothenburg University
SE-412 96 Gothenburg, SWEDEN
Email: palbin@math.chalmers.se URL: http://www.math.chalmers.se/~palbin/

Contents

1 **First Lecture** **1**
- 1.1 Variation of Functions . . . 1
- 1.2 Quadratic Variation of Functions . . . 3
- 1.3 Probability and Independence . . . 4
- 1.4 Mathematical Expectations and Conditional Expectations 6
- 1.5 Stochastic Processes . . . 9

2 **Second Lecture** **11**
- 2.1 Brownian Motion . . . 11
- 2.2 Gaussian Stochastic Processes . . . 13
- 2.3 Sample Path Properties of BM . . . 15
- 2.4 Introduction to Martingale Theory . . . 18
- 2.5 Martingale Properties of BM . . . 21

3 **Third Lecture** **22**
- 3.1 Conditional Expectations (continued) . . . 22
- 3.2 Introduction to Markov Theory . . . 24
- 3.3 Markov Properties of Lévy Processes . . . 25
- 3.4 Self-Similarity and Reflection Principle . . . 26
- 3.5 Zero Crossings of BM . . . 28

4 **Fourth Lecture** **31**
- 4.1 The Poisson Process . . . 31
- 4.2 Martingale Properties of Lévy Processes . . . 32
- 4.3 Markov and Martingale Properties of PP . . . 34
- 4.4 Stopping Times . . . 35
- 4.5 Strong Markov Property and Feller Processes . . . 36

5 **Fifth Lecture** **40**
- 5.1 On the Markov Property . . . 40
- 5.2 Strong Markov Property of Lévy Processes . . . 42

6 **Sixth Lecture** **44**
- 6.1 Introduction to Stochastic Integrals wrt. BM . . . 44

7 **Seventh Lecture** **50**
7.1 Maximal Inequalities for Martingales 50
7.2 Stochastic Integration of Simple Processes 51

8 **Eigth Lecture** **56**
8.1 Stochastic Integration of Processes in E_T 56
8.2 Concepts of Measurability 57

9 **Ninth Lecture** **59**
9.1 Stochastic Integration of Processes in E_T (continued) . . . 59

10 **Tenth Lecture** **63**
10.1 Properties of Stochastic Integrals in E_T 63
10.2 Stochastic Integration of Processes in P_T 65
10.3 Approximating Sums for Itô Integrals of Processes in E_T 67

11 **Eleventh Lecture** **68**
11.1 Stochastic Integration of Processes in P_T (continued) . . 68
11.2 Properties of Stochastic Integrals in P_T 70
11.3 Uniform Integrability and Regular Martingales 71
11.4 Approximating Sums for Itô Integrals of Processes in P_T 72

12 **Twelfth Lecture** **73**
12.1 Optional Stopping 73
12.2 Local Martingales 74
12.3 Properties of Stochastic Integrals in P_T (continued) . . . 75

13 **Thirteenth Lecture** **77**
13.1 Examples of Stochastic Integration 77
13.2 Introduction to Quadratic Variation of the Itô Integral . . 77
13.3 Quadratic Variation of the Itô Integral 78

14 **Fourteenth Lecture** **81**
14.1 Introduction to Itô's Formula 81
14.2 Itô's Formula 81
14.3 Examples of Stochastic Calculus 83

15 **Fifteenth Lecture** **85**
15.1 Quadratic Variation for Local Martingales 85
15.2 Itô's Formula (continued) 87
15.3 Stochastic Exponentials 88
15.4 Introduction to Diffusion Type SDE wrt. BM 88

16 **Sixteenth Lecture** **91**
16.1 Itô's Formula for Local Martingales 91
16.2 Beginning Values for SDE 92
16.3 Uniqueness of Solutions to SDE 94

17 **Seventeenth Lecture** **97**
17.1 Uniqueness of Solutions to SDE (continued) 97
17.2 Existence of Strong Solutions to SDE 98

18 **Eighteenth Lecture** **103**
18.1 Paul Lévy's Characterization of BM 103
18.2 Stratonovich Stochastic Integrals 104
18.3 Introduction to Girsanov's Theorem 105

19 **Nineteenth Lecture** **108**
19.1 Girsanov's Theorem 108
19.2 Multidimensional SDE 109
19.3 One-Dimensional Time Homogeneous SDE 111
19.4 Local Martingale Problems 113

20 **Twentieth Lecture** **117**
20.1 Martingale Problems 117
20.2 Existence of Solutions to Martingale Problems 119
20.3 PDO and PDE 122

21 **Twentyfirst Lecture** **126**
21.1 Uniqueness for Solutions to Martingale Problems 126
21.2 Feynman-Kac Formula for parabolic Cauchy Problem . . . 128
21.3 Solutions to Elliptic Dirichlet Problems 129

22 **Twentysecond Lecture** **132**
22.1 Markov Properties of SDE 132
22.2 How everything hang together 134
22.3 Multidimensional Time Homogeneous SDE 136

23 **Twentythird Lecture** **138**
23.1 What is a Stopping Time? 138
23.2 When do Smartingales Converge? 139
23.3 Optional Sampling 141
23.4 When are Local Martingales Martingales? 145
23.5 Burkholder-Davis-Gundy Inequalities 147

24 **Twentyfourth Lecture** **150**
24.1 Change of Time for Local Martingales 150
24.2 Semimartingales 152

25 **Twentyfifth Lecture** **157**
25.1 Application of Martingale Problems 157
25.2 Stationary One-Dimensional Diffusion Processes 159

26 **Twentysixth Lecture** **163**
26.1 Transformation of One-Dimensional Diffusion Processes 163
26.2 Transition Densities for One-Dimensional Diffusions 166

A ***Appendix. An Elementary Construction of BM** **169**

B ***Appendix. An Elementary Construction of PP** **172**

C **Appendix. Solution to Exercise 64** **174**

D ***Appendix. Selected further Martingale Proofs** **175**

References **176**

Index **178**

1 First Lecture

1.1 Variation of Functions

Definition 1.1 *A function* $g:[a,b]\to\mathbb{R}$ *has* <u>*finite variation over* $[a,b]$</u> $\subseteq\mathbb{R}$, *if*

$$\underline{V_g([a,b])} \equiv \sup\Big\{\sum_{i=1}^{n} |g(t_i)-g(t_{i-1})| : a=t_0<t_1<\ldots<t_n=b,\ n\in\mathbb{N}\Big\} < \infty.$$

A function $g:[0,\infty)\to\mathbb{R}$ *has* <u>*finite variation*</u> *if* $\underline{V_g(t)} \equiv V_g([0,t]) < \infty$ *for* $t>0$.
A function $g:[0,\infty)\to\mathbb{R}$ *has* <u>*bounded variation*</u> *if* $\lim_{t\to\infty} V_g(t) < \infty$.

*Remark 1.2[1] The space of functions g with $V_g([a,b])<\infty$, equipped with the norm $\|g\|=V_g([a,b])$, is the Banach space $BV([a,b])$. By the *Riesz Representation Theorem*, it can be indentified as the dual space to the space of continuous functions $g:[a,b]\to\mathbb{R}$ equipped with the supremum norm (e.g., [12, pp. 14-18]). #

Exercise 1 Show that

$$V_g([a,b]) = \lim_{\delta\downarrow 0}\sup\Big\{\sum_{i=1}^{n}|g(t_i)-g(t_{i-1})| : \begin{array}{c} a=t_0<t_1<\ldots<t_n=b \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}=\delta \end{array}\Big\}.$$

Example 1.3 For a continuously differentiable function $g:[a,b]\to\mathbb{R}$, we have

$$g(t_i)-g(t_{i-1}) = \int_{t_{i-1}}^{t_i} g'(t)\,dt = g'(\tau_i)\,(t_i-t_{i-1}) \qquad \text{for some} \quad \tau_i\in[t_{i-1},t_i].$$

By the theory for the Riemann integral (and Riemann sums), it follows that

$$\sum_{i=1}^{n}|g(t_i)-g(t_{i-1})| = \sum_{i=1}^{n}|g'(\tau_i)|\,(t_i-t_{i-1}) \to \int_a^b |g'(t)|\,dt$$

as $\max_{1\leq i\leq n} t_i-t_{i-1}\to 0$. Hence Exercise 1 gives $V_g([a,b])=\int_a^b|g'(t)|\,dt<\infty$. #

Theorem 1.4 *A function has finite variation over a closed finite interval iff. it can be expressed as the difference between two increasing[2] functions. A function has finite variation iff. it can be expressed as the difference between two increasing functions.*

Proof. The implication to the right follows writing $g(t)=V_g(t)-(V_g(t)-g(t))$. □

Exercise 2 Complete the proof of Theorem 1.4 by proving that V_g-g is increasing (clearly V_g is), and by giving the argument for the implication to the left.

[1] Material that is not required for the understanding of subsequent material is marked with an asterisk *. Exercises marked in this way are often more difficult than others as well.

[2] We use "positive" and "negative" in the non-strict sense. Same with "increasing" and "decreasing".

For a function f with finite variation over $[a,b]$ [finite variation], that is right-continuous say, the Lebesgue-Stieltjes integral $\int_{x\in[a,b]} g(x)\,df(x)$ $[\int_{x\in[0,\infty)} g(x)\,df(x)]$ is well-defined. Here $df(x)$ is the signed Stieltjes measure $df(x) = df_1(x) - df_2(x)$, obtained by representing $f = f_1 - f_2$ as the difference between two increasing functions, that correspond to positive Stieltjes measures $df_1(x)$ and $df_2(x)$, respectively.

Corollary 1.5 *A right-continuous function has finite variation over a closed finite interval iff. it can be expressed as the difference between two right-continuous increasing functions. A right-continuous function has finite variation iff. it can be expressed as the difference between two right-continuous increasing functions.*

Exercise 3 Show how Corollary 1.5 follows from Theorem 1.4.

Definition 1.6 *A function $g:(a,b)\to\mathbb{R}$, $-\infty\le a<b\le\infty$, has a jump discontinuity at $t_0\in(a,b)$ if the limits $\lim_{t\uparrow t_0} g(t)$ and $\lim_{t\downarrow t_0} g(t)$ exist, but are not both equal to $g(t_0)$.*
A function $g:[a,b]\to\mathbb{R}$ has a jump discontinuity at a [b] if the limit $\lim_{t\downarrow a} g(t)$ [$\lim_{t\uparrow b} g(t)$] exists, but is not equal to $g(a)$ [$g(b)$].

Theorem 1.7 *A function that has finite variation over a closed finite interval has at most countable many discontinuities, all of which are jumps. A function that has finite variation has at most countable many discontinuities, all of which are jumps.*

Proof. The second statement follows from the first one, by considering subsequent restrictions of the function to $[0,n]$ for $n=1,2,\ldots$. To prove the first statement, notice that, since a function g with finite variation over $[a,b]$ is the difference between two increasing functions, it has limits from the left (except at a), and from the right (except at b). Thus all discontinuities are jumps. These jumps are contained in

$$\bigcup_{n=1}^{\infty}\Big\{t\in[a,b] : g \text{ has a jump at } t \text{ with } |g(t^+)-g(t)|+|g(t)-g(t^-)|\ge 1/n\Big\}.$$

Since the sets that form this union are finite (because otherwise g would not have finite variation), the union itself is a countable set. □

***Exercise 4** Give an example of a continuous function that does not have finite variation over $[0,1]$. (**Hint:** Build a sequence of continuous functions that converges to a continuous function, but for which the variation over $[0,1]$ goes to ∞.)

1.2 Quadratic Variation of Functions

Definition 1.8 *The quadratic variation over $[a,b]\subseteq\mathbb{R}$ of a function $g:[a,b]\to\mathbb{R}$, is given by the following limit, provided that the limit exists,*

$$\underline{[g]([a,b])}\equiv\lim\left\{\sum_{i=1}^{n}(g(t_i)-g(t_{i-1}))^2 : \begin{array}{c} a=t_0<t_1<\ldots<t_n=b \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0\end{array}\right\}.$$

***Remark 1.9** The limit in Definition 1.8 exists iff., for each choice of a family $\{\{t_i^{(N)}\}_{i=1}^{n_N}\}_{N=1}^{\infty}$ of partitions $a=t_0^{(N)}<t_1^{(N)}<\ldots<t_{n_N}^{(N)}=b$ of $[a,b]$, that satisfies $\lim_{N\to\infty}\max_{1\leq i\leq n_N} t_i^{(N)}-t_{i-1}^{(N)}=0$, we have convergence

$$\lim_{N\to\infty}\sum_{i=1}^{n_N}(g(t_i^{(N)})-g(t_{i-1}^{(N)}))^2=[g]([a,b]).$$

Equivalently, the limit exists, with value $[g]([a,b])$, iff. we have

$$\lim_{\delta\downarrow 0}\sup\left\{\sum_{i=1}^{n}(g(t_i)-g(t_{i-1}))^2 : \begin{array}{c} a=t_0<t_1<\ldots<t_n=b \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}=\delta\end{array}\right\}=[g]([a,b])$$

and

$$\lim_{\delta\downarrow 0}\inf\left\{\sum_{i=1}^{n}(g(t_i)-g(t_{i-1}))^2 : \begin{array}{c} a=t_0<t_1<\ldots<t_n=b \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}=\delta\end{array}\right\}=[g]([a,b]).\quad \#$$

Definition 1.10 *The covariation over $[a,b]\subseteq\mathbb{R}$ of the functions $f,g:[a,b]\to\mathbb{R}$, is given by the following limit, provided that the limit exists,*

$$\underline{[f,g]([a,b])}\equiv\lim\left\{\sum_{i=1}^{n}(f(t_i)-f(t_{i-1}))\,(g(t_i)-g(t_{i-1})) : \begin{array}{c} a=t_0<t_1<\ldots<t_n=b \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0\end{array}\right\}.$$

Notice the trivial fact that $[g]([a,b])=[g,g]([a,b])$.

We use the notation $\underline{[g](t)}\equiv[g]([0,t])$ and $\underline{[f,g](t)}\equiv[f,g]([0,t])$.

Theorem 1.11 *For* $f\in\underline{\mathbb{C}([a,b])}\equiv\{(\tilde f:[a,b]\to\mathbb{R}) : \tilde f$ is continuous$\}$ *and* g *with finite variation over* $[a,b]$, *we have* $[f,g]([a,b])=0$.

Exercise 5 Prove Theorem 1.11.

Theorem 1.12 (Polarization) *If the covariations involved exist, we have*

$$[f,g]=\frac{1}{2}\Big([f+g,f+g]-[f,f]-[g,g]\Big)=\frac{1}{4}\Big([f+g,f+g]-[f-g,f-g]\Big).$$

EXERCISE 6 Prove Theorem 1.12. [Equip the space L of functions with well-defined quadratic variation over $[a,b]$, with the symmetric bilinear form given by the covariation $[\cdot,\cdot]([a,b]): L\times L\to\mathbb{R}$. What is such a space called (e.g., [7, p. V.1])?]*

1.3 Probability and Independence

A σ-algebra $\mathcal{F}$ is a non-empty family of subsets (called events) of a set Ω, that is closed under the formation of intersections, complements and countable unions. A mesurable space $(\Omega,\mathcal{F})$, is a set Ω equipped with a σ-algebra $\mathcal{F}$ (of subsets of Ω).

A probability measure $\mathbf{P}$ on a measurable space $(\Omega,\mathcal{F})$, is a positive measure on $\mathcal{F}$ such that $\mathbf{P}(\Omega)=1$. A probability space $(\Omega,\mathcal{F},\mathbf{P})$ is a measurable space $(\Omega,\mathcal{F})$, equipped with a probability measure $\mathbf{P}$ [on $(\Omega,\mathcal{F})$].

A ($\mathbb{R}$-valued) *random variable [an $\mathbb{R}^n$-valued random variable] on a probability space $(\Omega,\mathcal{F},\mathbf{P})$, is a measurable function $X:\Omega\to\mathbb{R}$ [$X:\Omega\to\mathbb{R}^n$],* that is, $X^{-1}(B) = \{\omega\in\Omega : X(\omega)\in B\}\in\mathcal{F}$ for $B\in\mathcal{B}(\mathbb{R})$ $[B\in\mathcal{B}(\mathbb{R}^n)]$ (the Borel sets in $\mathbb{R}$ $[\mathbb{R}^n]$).

In the sequel, all random variables that appear are assumed to be defined on a common probability space $(\Omega,\mathcal{F},\mathbf{P})$. All σ-algebras that appear are assumed to be contained in $\mathcal{F}$, so that their events are assigned probabilities by $\mathbf{P}$.

*It turns out that $X=(X_1,\ldots,X_n)$ is an $\mathbb{R}^n$-valued random variable, iff. each of its components $X_1,\ldots,X_n$ are random variables (e.g., [34, p. 11]).

Definition 1.13 *The σ-algebra $\sigma(X)$ generated by an $\mathbb{R}^n$-valued random variable X, is the smallest σ-algebra on Ω that makes $X:\Omega\to\mathbb{R}^n$ measurable.*
The σ-algebra $\sigma(\mathcal{G}_\alpha:\alpha\in\mathcal{A})$ generated by the σ-algebras $\{\mathcal{G}_\alpha\}_{\alpha\in\mathcal{A}}$ is the smallest σ-algebra on Ω that contains $\bigcup_{\alpha\in\mathcal{A}}\mathcal{G}_\alpha$.

Definition 1.14 *The σ-algebras $\{\mathcal{G}_\alpha\}_{\alpha\in\mathcal{A}}$ are independent, if for each choice of $n\in\mathbb{N}$ and $\alpha_1,\ldots,\alpha_n\in\mathcal{A}$, we have*

$$\mathbf{P}\{A_1\cap\ldots\cap A_n\} = \mathbf{P}\{A_1\}\ldots\mathbf{P}\{A_n\} \qquad \text{for} \quad A_1\in\mathcal{G}_{\alpha_1},\ldots,A_n\in\mathcal{G}_{\alpha_n}.$$

The random variables $\{X_\alpha\}_{\alpha\in\mathcal{A}}$ are independent if $\{\sigma(X_\alpha)\}_{\alpha\in\mathcal{A}}$ are independent.
The σ-algebra $\mathcal{G}$ [random variable X] is independent of the σ-algebras $\{\mathcal{G}_\alpha\}_{\alpha\in\mathcal{A}}$ if $\mathcal{G}$ [$\sigma(X)$] is independent of $\sigma(\mathcal{G}_\alpha:\alpha\in\mathcal{A})$.

Similarly, a σ-algebra $\mathcal{H}$ [a random variable X] is independent of a random variable Y and a σ-algebra $\mathcal{G}$, if $\mathcal{H}$ [X] is independent of $\sigma(\sigma(Y),\mathcal{G})$, etc.

The σ-algebra $\sigma(\mathcal{G}_\alpha:\alpha\in\mathcal{A})$ consists of sets built by means of performing a countable number of set operations (unions, intersections and complements), involving a countable number of members of $\bigcup_{\alpha\in\mathcal{A}}\mathcal{G}_\alpha$. This explains the following result:

Theorem 1.15 *The family of σ-algebras $\{\mathcal{G}_\alpha\}_{\alpha\in\mathfrak{A}}$ is independent of the family of σ-algebras $\{\mathcal{H}_\beta\}_{\beta\in\mathfrak{B}}$ iff. for all $n\in\mathbb{N}$, $\alpha_1,\ldots,\alpha_n\in\mathfrak{A}$ and $\beta_1,\ldots,\beta_n\in\mathfrak{B}$, we have*

$$\mathbf{P}\Big\{(A_1\cap\ldots\cap A_n)\cap(B_1\cap\ldots\cap B_n)\Big\}=\mathbf{P}\{A_1\cap\ldots\cap A_n\}\,\mathbf{P}\{B_1\cap\ldots\cap B_n\}$$
$$\text{for}\quad A_1\in\mathcal{G}_{\alpha_1},\ldots,A_n\in\mathcal{G}_{\alpha_n}\quad\text{and}\quad B_1\in\mathcal{H}_{\beta_1},\ldots,B_n\in\mathcal{H}_{\beta_n}.$$

Theorem 1.16 (Dynkin System Lemma) ([13, pp. 1-2])* *Let $\mathcal{C}$ and $\mathcal{D}$ be classes of subsets of a set S. Assume that $\mathcal{C}$ has the following two properties:*

- *$\mathcal{C}$ is closed under finite intersections;*
- *$\mathcal{C}\subseteq\mathcal{D}$.*

We have $\sigma(\mathcal{C})\subseteq\mathcal{D}$ provided that $\mathcal{D}$ is a Dynkin system, that is, provided that

- *$S\in\mathcal{D}$;*
- *$\mathcal{D}\ni A\subseteq B\in\mathcal{D}\ \Rightarrow\ B-A\in\mathcal{D}$;*
- *$\{A_n\}_{n=1}^\infty\subseteq\mathcal{D}$ are increasing $\Rightarrow\ \bigcup_{n=1}^\infty A_n\subseteq\mathcal{D}$.*

**Proof of Theorem 1.15.* Put $\mathcal{G}\equiv\sigma(\mathcal{G}_\alpha:\alpha\in\mathfrak{A})$ and $\mathcal{H}\equiv\sigma(\mathcal{H}_\beta:\beta\in\mathfrak{B})$. Let

$$\begin{cases}\mathcal{C}_1=\Big\{A_1\cap\ldots\cap A_n: A_1,\ldots,A_n\in\bigcup_{\alpha\in\mathfrak{A}}\mathcal{G}_\alpha,\ n\in\mathbb{N}\Big\}\\ \mathcal{C}_2=\Big\{B_1\cap\ldots\cap B_n: B_1,\ldots,B_n\in\bigcup_{\beta\in\mathfrak{B}}\mathcal{H}_\beta,\ n\in\mathbb{N}\Big\}\end{cases},$$

and

$$\begin{cases}\mathcal{D}_1=\Big\{A\in\mathcal{G}:\mathbf{P}\{A\cap B\}=\mathbf{P}\{A\}\,\mathbf{P}\{B\}\ \text{for all}\ B\in\mathcal{C}_2\Big\}\\ \mathcal{D}_2=\Big\{B\in\mathcal{H}:\mathbf{P}\{A\cap B\}=\mathbf{P}\{A\}\,\mathbf{P}\{B\}\ \text{for all}\ A\in\mathcal{G}\Big\}\end{cases}.$$

By construction, $\mathcal{C}_1$ and $\mathcal{C}_2$ are closed under finite intersections, while by hypothesis, $\mathcal{C}_1\subseteq\mathcal{D}_1$. Further, it is easy to check that $\mathcal{D}_1$ is a Dynkin system ($S=\Omega$). Hence the *Dynkin System Lemma* gives $\sigma(\mathcal{C}_1)\subseteq\mathcal{D}_1$. However, $\sigma(\mathcal{C}_1)=\mathcal{G}$ [since $\bigcup_{\alpha\in\mathcal{A}}\mathcal{G}_\alpha\subseteq\mathcal{C}_1$], and so $\mathbf{P}\{A\cap B\}=\mathbf{P}\{A\}\,\mathbf{P}\{B\}$ for $A\in\mathcal{G}$ and $B\in\mathcal{C}_2$, that is $\mathcal{C}_2\subseteq\mathcal{D}_2$. Since also $\mathcal{D}_2$ is a Dynkin system, the *Dynkin System Lemma* gives $\sigma(\mathcal{C}_2)\subseteq\mathcal{D}_2$, where $\sigma(\mathcal{C}_2)=\mathcal{H}$. Consequently, $\mathbf{P}\{A\cap B\}=\mathbf{P}\{A\}\,\mathbf{P}\{B\}$ for $A\in\mathcal{G}$ and $B\in\mathcal{H}$. □

***Remark 1.17** For an $\mathbb{R}^n$-valued random variable $X=(X_1,\ldots,X_n)$, we have $\sigma(X)=\sigma(\sigma(X_1),\ldots,\sigma(X_n))$. This follows from, and is equivalent with, the fact that X is an $\mathbb{R}^n$-valued random variable iff. all its components are random variables. #

1.4 Mathematical Expectations and Conditional Expectations

The expectation of a positive random variable X is given by

$$\mathbf{E}\{X\} = \int_\Omega X\,d\mathbf{P} = \int_{\omega\in\Omega} X(\omega)\,d\mathbf{P}(\omega) \qquad \textit{(which may be infinite).}$$

The expectation of a random variable X is $\mathbf{E}\{X\} = \mathbf{E}\{X^+\} - \mathbf{E}\{X^-\}$, when at least one of $\mathbf{E}\{X^+\}$ and $\mathbf{E}\{X^-\}$ are finite ($x^+ = \max\{x,0\}$ and $x^- = \max\{-x,0\}$).

Exercise 7 Construct a random variable $X(\omega)$, $\omega\in\Omega$, such that $\mathbf{E}\{X\} = \infty$.

Theorem 1.18 *Let X be a random variable with $\mathbf{E}\{|X|\} < \infty$, and $\mathcal{G}$ a σ-algebra. There exists a $\mathcal{G}$-measurable random variable Y such that*

$$\mathbf{E}\{|Y|\} < \infty \qquad \textit{and} \qquad \mathbf{E}\{I_A X\} = \int_A X\,d\mathbf{P} = \int_A Y\,d\mathbf{P} = \mathbf{E}\{I_A Y\} \quad \textit{for each } A\in\mathcal{G}.$$

Exercise 8 Derive Theorem 1.18 from the *Radon-Nikodym Theorem* (e.g., [34, pp. 129-130])*: If μ_1 and μ_2 are finite signed measures on a measurable space $(\Omega,\mathcal{G})$, such that $\mu_1(A)=0$ for all $A\in\mathcal{G}$ with $\mu_2(A)=0$, then we have

$$\mu_1(A) = \int_A \lambda\,d\mu_2 \qquad \text{for each} \quad A\in\mathcal{G}, \quad \text{for some } \mu_2\text{-integrable function} \quad \lambda:\Omega\to\mathbb{R}.$$

Definition 1.19 *Let X be a random variable with $\mathbf{E}\{|X|\} < \infty$, and $\mathcal{G}$ a σ-algebra. The conditional expectation of X wrt. $\mathcal{G}$, is a $\mathcal{G}$-measurable random variable $\mathbf{E}\{X|\mathcal{G}\}$ with $\mathbf{E}\big\{|\mathbf{E}\{X|\mathcal{G}\}|\big\} < \infty$, such that* (cf. Theorem 1.18)

$$\mathbf{E}\{I_A X\} = \int_A X\,d\mathbf{P} = \int_A \mathbf{E}\{X|\mathcal{G}\}\,d\mathbf{P} = \mathbf{E}\big\{I_A\mathbf{E}\{X|\mathcal{G}\}\big\} \qquad \textit{for each} \quad A\in\mathcal{G}.$$

Though $\mathbf{E}\{X|\mathcal{G}\}$ exists, by Theorem 1.18, it is not unique as a function of $\omega\in\Omega$. By basic measure theory (e.g., [34, p. 21])*, two versions of $\mathbf{E}\{X|\mathcal{G}\}$, that both satisfy Definition 1.19, differ on a set of ω's in Ω of probability zero. Thus they are equal almost surely (a.s.). When referring to $\mathbf{E}\{X|\mathcal{G}\}$, we mean any such version. When stating that $\mathbf{E}\{X|\mathcal{G}\} = Y$, for some random variable Y, we mean that Y is a version of $\mathbf{E}\{X|\mathcal{G}\}$, so that $\mathbf{E}\{X|\mathcal{G}\} = Y$ satisfies Definition 1.19.

Exercise 9 Let X and Y be random variables with $\mathbf{E}\{|X|\}, \mathbf{E}\{|Y|\} < \infty$, and $\mathcal{G}$ an σ-algebra. Show that

$$\boxed{\mathbf{E}\{aX+bY|\mathcal{G}\} = a\,\mathbf{E}\{X|\mathcal{G}\} + b\,\mathbf{E}\{Y|\mathcal{G}\} \qquad \text{for constants} \quad a,b\in\mathbb{R}},$$

and that

$$\boxed{X\geq 0 \quad\Rightarrow\quad \mathbf{E}\{X|\mathcal{G}\}\geq 0}.$$

EXERCISE 10 $\boxed{\mathbf{E}\{|X|\}<\infty \;\Rightarrow\; \mathbf{E}\Big\{X\,\Big|\,\{\emptyset,\Omega\}\Big\} = \mathbf{E}\{X\}}$

Definition 1.20 *For a random variable* X *with* $\mathbf{E}\{|X|\}<\infty$, *and an* $\mathbb{R}^n$*-valued random variable* Y, *we write* $\underline{\mathbf{E}\{X|Y\}} \equiv \mathbf{E}\{X|\sigma(Y)\}$.

Notice the immediate but important fact that, for an $\mathbb{R}^n$-valued random variable X, $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}^n)) = \{X^{-1}(C) : C \in \mathcal{B}(\mathbb{R}^n)\}$.

Theorem 1.21 *Let* X *and* Y *be random variables such that* $\mathbf{E}\{|X|\}<\infty$ *and* $\mathbf{E}\{|XY|\}<\infty$. *If* X *is measurable wrt. a* σ*-algebra* $\mathcal{G}$, *then we have*

$$\mathbf{E}\{XY|\mathcal{G}\} = X\,\mathbf{E}\{Y|\mathcal{G}\}.$$

Proof. Writing $XY = X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-$, Exercise 9 shows that it is enough to prove the theorem for X and Y positive. To that end, pick simple random variables $\sum_{i=1}^{N_n} b_i^{(n)} I_{B_i^{(n)}}(\omega) \uparrow X(\omega)$ as $n\to\infty$, where $b_1^{(n)},\ldots,b_{N_n}^{(n)} \geq 0$ are constants and $B_1^{(n)},\ldots,B_{N_n}^{(n)} \in \mathcal{G}$ (e.g., [34, p. 16])*. For an event $A\in\mathcal{G}$, we have

$$\begin{aligned}
\int_A X\,\mathbf{E}\{Y|\mathcal{G}\}\,d\mathbf{P} \leftarrow \int_A \sum_{i=1}^{N_n} b_i^{(n)} I_{B_i^{(n)}}\,\mathbf{E}\{Y|\mathcal{G}\}\,d\mathbf{P} &= \sum_{i=1}^{N_n} b_i^{(n)} \int_{A\cap B_i^{(n)}} \mathbf{E}\{Y|\mathcal{G}\}\,d\mathbf{P} \\
&= \sum_{i=1}^{N_n} b_i^{(n)} \int_{A\cap B_i^{(n)}} Y\,d\mathbf{P} \\
&= \int_A \sum_{i=1}^{N_n} b_i^{(n)} I_{B_i^{(n)}}\,Y\,d\mathbf{P} \\
&\to \int_A XY\,d\mathbf{P} \\
&= \int_A \mathbf{E}\{XY|\mathcal{G}\}\,d\mathbf{P} \qquad \text{as} \quad n\to\infty,
\end{aligned}$$

by the *Monotone Convergence Theorem* (e.g., [34, p. 22])*. □

EXERCISE 11 Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be σ-algebras, and X a random variable with $\mathbf{E}\{|X|\}<\infty$. Show that

$$\boxed{\mathbf{E}\Big\{\mathbf{E}\{X|\mathcal{G}_2\}\,\Big|\,\mathcal{G}_1\Big\} = \mathbf{E}\{X|\mathcal{G}_1\} \qquad \text{when} \quad \mathcal{G}_1\subseteq\mathcal{G}_2}\,.$$

EXERCISE 12 Let X be a random variable with $\mathbf{E}\{|X|\}<\infty$, and $\mathcal{G}$ an σ-algebra. Show that

$$\boxed{\mathbf{E}\Big\{\mathbf{E}\{X|\mathcal{G}\}\Big\} = \mathbf{E}\{X\}}\,.$$

Theorem 1.22 *Let X be a random variable with $\mathbf{E}\{|X|\} < \infty$, and $\mathcal{G}$ a σ-algebra. If X and $\mathcal{G}$ are independent, then we have $\mathbf{E}\{X|\mathcal{G}\} = \mathbf{E}\{X\}$.*

Proof. By Exercise 9, it is enough to consider positive X. Pick simple functions $\sum_{i=1}^{N_n} b_i^{(n)} I_{B_i^{(n)}}(\omega) \uparrow X(\omega)$ as $n \to \infty$, where $b_1^{(n)}, \ldots, b_{N_n}^{(n)} \geq 0$ and $B_1^{(n)}, \ldots, B_{N_n}^{(n)} \in \sigma(X)$. For $A \in \mathcal{G}$, we have, as $n \to \infty$, by the *Monotone Convergence Theorem*,

$$\begin{aligned}\int_A X \, d\mathbf{P} \leftarrow \int_A \sum_{i=1}^{N} b_i I_{B_i} \, d\mathbf{P} = \sum_{i=1}^{N} b_i \mathbf{P}\{A \cap B_i\} &= \mathbf{P}\{A\} \sum_{i=1}^{N} b_i \mathbf{P}\{B_i\} \\ &\to \mathbf{P}\{A\} \int_\Omega X \, d\mathbf{P} \\ &= \mathbf{P}\{A\} \, \mathbf{E}\{X\} = \int_A \mathbf{E}\{X\} \, d\mathbf{P}. \quad \square\end{aligned}$$

Theorem 1.23 *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be σ-algebras, and X a random variable with $\mathbf{E}\{|X|\} < \infty$. If $\mathcal{G}_1$ is independent of X and $\mathcal{G}_2$, then we have*

$$\mathbf{E}\{X | \sigma(\mathcal{G}_1, \mathcal{G}_2)\} = \mathbf{E}\{X | \mathcal{G}_2\}.$$

Proof. It is enough to consider positive X. Pick simple functions $0 \leq \sum_{i=1}^{M_n} b_i^{(n)} I_{B_i^{(n)}} \uparrow X$ and $0 \leq \sum_{i=1}^{N_n} c_i^{(n)} I_{C_i^{(n)}} \uparrow \mathbf{E}\{X|\mathcal{G}_2\}$ as $n \to \infty$, where $B_1^{(n)}, \ldots, B_{M_n}^{(n)} \in \sigma(X)$ and $C_1^{(n)}, \ldots, C_{N_n}^{(n)} \in \mathcal{G}_2$. For $A_1 \in \mathcal{G}_1$ and $A_2 \in \mathcal{G}_2$, we have (by monotone convergence)

$$\begin{aligned}\int_{A_1 \cap A_2} X \, d\mathbf{P} \leftarrow \int_{A_1 \cap A_2} \sum_{i=1}^{M} b_i I_{B_i} \, d\mathbf{P} &= \sum_{i=1}^{M} b_i \, \mathbf{P}\{B_i \cap A_1 \cap A_2\} \\ &= \mathbf{P}\{A_1\} \sum_{i=1}^{M} b_i \, \mathbf{P}\{B_i \cap A_2\} \\ &\to \mathbf{P}\{A_1\} \int_{A_2} X \, d\mathbf{P} \\ &= \mathbf{P}\{A_1\} \int_{A_2} \mathbf{E}\{X|\mathcal{G}_2\} \, d\mathbf{P} \\ &\leftarrow \mathbf{P}\{A_1\} \int_{A_2} \sum_{i=1}^{N} c_i I_{C_i} \, d\mathbf{P} \\ &= \sum_{i=1}^{N} c_i \, \mathbf{P}\{C_i \cap A_1 \cap A_2\} \\ &= \int_{A_1 \cap A_2} \sum_{i=1}^{N} c_i I_{C_i} \, d\mathbf{P} \to \int_{A_1 \cap A_2} \mathbf{E}\{X|\mathcal{G}_2\} \, d\mathbf{P}\end{aligned}$$

as $n \to \infty$. Now use the *Dynkin System Lemma*, with $S = \Omega$, $\mathcal{C} = \{A_1 \cap A_2 : A_1 \in \mathcal{G}_1, \, A_2 \in \mathcal{G}_2\}$ and $\mathcal{D} = \{A \in \sigma(\mathcal{G}_1, \mathcal{G}_2) : \int_A X \, d\mathbf{P} = \int_A \mathbf{E}\{X|\mathcal{G}_2\} \, d\mathbf{P}\}$, to show that $\int_A X \, d\mathbf{P} = \int_A \mathbf{E}\{X|\mathcal{G}_2\} \, d\mathbf{P}$ for each $A \in \sigma(\mathcal{G}_1, \mathcal{G}_2)$ (and not only for $A \in \mathcal{C}$). $\square$

Theorem 1.24 (Jensen's Inequality) *For a random variable X and a convex function $g : \mathbb{R} \to \mathbb{R}$, with $\mathbf{E}\{|X|\}, \mathbf{E}\{|g(X)|\} < \infty$, we have $g(\mathbf{E}\{X\}) \leq \mathbf{E}\{g(X)\}$.*

Recall that a function $g:\mathbb{R}\to\mathbb{R}$ is convex if

$$g(\lambda x+(1-\lambda)y) \leq \lambda g(x)+(1-\lambda)g(y) \qquad \text{for} \quad x,y\in\mathbb{R} \quad \text{and} \quad \lambda\in[0,1].$$

Convex functions are continuous (e.g., [34, p. 63])*, and thus measurable.

EXERCISE 13 Prove Jensen's Inequality. (**Hint:** Apply convexity to $g(\sum_i \lambda_i x_i)$.)

EXERCISE 14 Let X be a random variable and $g:\mathbb{R}\to\mathbb{R}$ a (measurable) function such that $\mathbf{E}\{|X|\}, \mathbf{E}\{|g(X)|\}<\infty$. Show that

$$\boxed{g(\mathbf{E}\{X|\mathcal{G}\}) \leq \mathbf{E}\{g(X)|\mathcal{G}\} \qquad \text{when } g \text{ is convex}}\,.$$

Theorem 1.25 (FATOU'S LEMMA) *Let $\mathcal{G}$ be a σ-algebra, and $X_1,X_2,\ldots$ positive random variables with $\mathbf{E}\{X_n\}<\infty$ and $\mathbf{E}\{\liminf_{n\to\infty} X_n\}<\infty$. We have*

$$\mathbf{E}\{\liminf_{n\to\infty} X_n|\mathcal{G}\} \leq \liminf_{n\to\infty} \mathbf{E}\{X_n|\mathcal{G}\}.$$

Theorem 1.26 (DOMINATED CONVERGENCE THEOREM) *Let $\mathcal{G}$ be a σ-algebra, and $X_1,X_2,\ldots$ a sequence of random variables that converges a.s. If $|X_n|\leq Y$ for some positive random variable Y with $\mathbf{E}\{Y\}<\infty$, then we have*

$$\mathbf{E}\{\lim_{n\to\infty} X_n|\mathcal{G}\} = \lim_{n\to\infty} \mathbf{E}\{X_n|\mathcal{G}\}.$$

EXERCISE 15 Prove Theorems 1.25 and 1.26 (cf. e.g., [34, p. 24 and pp. 27-28])*.

Definition 1.27 *The conditional probability wrt. a σ-algebra $\mathcal{G}$ is given by*

$$\underline{\mathbf{P}\{A|\mathcal{G}\}} \equiv \mathbf{E}\{I_A|\mathcal{G}\} \qquad \text{for} \quad A\in\mathcal{F}.$$

1.5 Stochastic Processes

Definition 1.28 *A stochastic process $X=\{X(t)\}_{t\in T}$ with parameter set T, is a function $X:\Omega\times T\to\mathbb{R}$ such that $X(\cdot,t):\Omega\to\mathbb{R}$ is a random variable for $t\in T$.*

The dependence of $\omega\in\Omega$ for a stochastic process X is often suppressed in the notation, so that we write $X(t)$ or $\{X(t)\}_{t\in T}$ instead of $X(\omega,t)$ or $\{X(\omega,t)\}_{(\omega,t)\in\Omega\times T}$.

Definition 1.29 *The finite dimensional distributions (`fidi's`) $\{F_{X(t_1),\ldots,X(t_n)} : t_1,\ldots,t_n\in T,\ n\in\mathbb{N}\}$ of a stochastic process $X=\{X(t)\}_{t\in T}$, are given by*

$$F_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n) = \mathbf{P}\{X(t_1)\leq x_1,\ldots,X(t_n)\leq x_n\} \qquad \text{for} \quad x_1,\ldots,x_n\in\mathbb{R}.$$

It is tempting to belive that a stochastic process is more or less "determined" by its univariate marginal distributions $F_{X(t)}(x) = \mathbf{P}\{X(t) \le x\}$ *for* $x \in \mathbb{R}$, *for each* $t \in T$. *This is not true at all* (cf. Exercise 16 below). In fact, in general, not even all the `fidi's` are enough for that purpose (cf. Exercise 17 below).

EXERCISE 16 Consider the stochastic process $X(t) \equiv \xi$ for $t \in \mathbb{R}$, where ξ is a single $\mathrm{N}(0,1)$-distributed random variable. Let $Y(t)$ be a process that is $\mathrm{N}(0,1)$-distributed at each $t \in \mathbb{R}$, but with the random values of the process at different times independent of each other. Find the univariate marginal distributions $F_{X(t)}$ and $F_{Y(t)}$. Plot a likely appearance of the graphs (so called realisations) $\mathbb{R} \ni t \to X(t) = X(\omega, t) \in \mathbb{R}$ and $\mathbb{R} \ni t \to Y(t) = Y(\omega, t) \in \mathbb{R}$, for a "typical" $\omega \in \Omega$.

Definition 1.30 *The stochastic processes* $\{X(t)\}_{t\in T}$ *and* $\{Y(t)\}_{t\in T}$ *are versions of each other if* $\mathbf{P}\{X(t) = Y(t)\} = 1$ *for each* $t \in T$.

Probabilities of events for processes X *and* Y *that are versions of each other need not be equal* (cf. Exercise 17 below). However, usually there is no need (desire) to regard processes that are versions of each other, but not equal, as really different, but rather as different expressions of one single process.

Definition 1.31 *A stochastic processes* $\{X(t)\}_{t\in T}$, $T \subseteq \mathbb{R}$, *is separable (strongly separable), if there exists a countable set* $S \subseteq T$ *(a separator), such that*

$$\mathbf{P}\Big\{\text{to each } t \in T \text{ there exists } \{s_n\}_{n=1}^{\infty} \subseteq S \text{ such that } s_n \to t \text{ and } X(s_n) \to X(t)\Big\} = 1.$$

The important feature of a separable process, is that probabilities of "interesting events" are determined by the `fidi's`, *evaluated at the times in the separator.*

Example 1.32 For $\{X(t)\}_{t\in\mathbb{R}}$ separable with separator $S = \{s_n\}_{n=1}^{\infty}$, we have

$$\mathbf{P}\Big\{\sup_{t\in I} X(t) > x\Big\} = \lim_{n\to\infty} \mathbf{P}\Big\{\sup_{t\in I\cap\{s_1,\ldots,s_n\}} X(t) > x\Big\} \qquad \text{for } x \in \mathbb{R} \text{ and open } I \subseteq \mathbb{R}. \ \#$$

Theorem 1.33 (DOOB, 1953) (e.g., [35, Section 9.2])* *A stochastic processes* $\{X(t)\}_{t\in T}$, $T \subseteq \mathbb{R}$, *has a separable version.*

EXERCISE 17 Find two stochastic processes $\{X(t)\}_{t\in[0,1]}$ and $\{Y(t)\}_{t\in[0,1]}$, with X separable, that have common `fidi's` and are versions of each other, but satisfy

$$\mathbf{P}\{X(t) \ne Y(t) \text{ for some } t \in [0,1]\} = 1.$$

2 Second Lecture

2.1 Brownian Motion

Definition 2.1 *A stochastic process* $\{B(t)\}_{t\geq 0}$ *is* *Brownian motion (BM)* *(also called* *Wiener process*), *if it has the following properties:*

- (Continuity) $[0,\infty)\ni t\to B(\omega,t)\in\mathbb{R}$ *is continuous for all (almost all)* $\omega\in\Omega$*;*
- (Independent Increments) $B(t)-B(s)$ *is independent of* $\{B(r)\}_{r\in[0,s]}$ *for* $0\leq s<t$*;*
- (Stationary Gaussian Increments) $B(t)-B(s)$ *is* $\mathrm{N}(0,t-s)$*-distributed for* $0\leq s\leq t$*.*

A random variable ξ is $\mathrm{N}(m,\sigma^2)$-distributed if $\mathbf{P}\{\xi\leq x\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\,\sigma}\,\mathrm{e}^{-y^2/(2\sigma^2)}\,dy.$

*Remark 2.2 The existence of BM is not trivial. See Appendix A on an elementary and quite economical proof of that existence.

Definition 2.1 of BM in stochastic calculus differs from that in much other probability, by not making a specific requirement about the value of $B(0)$.

The literature is often sloppy in the use of the definition of BM, and results that require something about $B(0)$ [e.g., $B(0)=0$], are mixed with results that do not, leaving it to the reader to decide what is required and when. We explain why later. #

Exercise 18 Explain why it is possible for BM $B(t)$, at a certain time $t\geq 0$, to have a quite arbitrary (albeit not completely so) univariate probability distribution.

Exercise 19 Explain without calculations, why the `fidi's` for BM

$$F_{B(t_1),\ldots,B(t_n)}(x_1,\ldots,x_n)=\mathbf{P}\Big\{B(t_1)\leq x_1,\ldots,B(t_n)\leq x_n\Big\}$$

become determined, under the additional requirement (to those in the definition of BM) that $B(0)=x$, for some constant $x\in\mathbb{R}$.

Definition 2.3 *B^x denotes BM with $B(0)=B^x(0)=x$, for a constant* $x\in\mathbb{R}$.

Exercise 20 Show without calculations that $\{B^x(t)\}_{t\geq 0}=_{\text{same fidi's}}\{B^0(t)+x\}_{t\geq 0}$.

Definition 2.4 *For the* *probability density function of $B^x(t)$* *we use the notation*

$$p_t(x,y)=f_{B^x(t)}(y)=\frac{d}{dy}\mathbf{P}\{B^x(t)\leq y\}=\frac{1}{\sqrt{2\pi t}}\,\exp\bigg\{-\frac{(y-x)^2}{2\,t}\bigg\}.$$

Theorem 2.5 *For a measurable function $g:\mathbb{R}^n\to\mathbb{R}$, we have*

$$\mathbf{E}\Big\{g(B^x(t_1),\dots,B^x(t_n))\Big\} = \int_{\mathbb{R}^n} g(y)\, p_{t_1}(x,y_1) \prod_{i=2}^{n} p_{t_i-t_{i-1}}(y_{i-1},y_i)\, dy.$$

(The left-hand side and the right-hand side are well-defined simultaneously, and are equal when they are well-defined.)

EXERCISE 21 Prove Theorem 2.5.

We have the following elementary formulas for BM:

$$\begin{cases} \mathbf{E}\{B(t)-B(s)\} &= 0 \\ \mathbf{E}\Big\{(B(t)-B(s))^2\Big\} &= t-s \\ \mathbf{Var}\Big\{(B(t)-B(s))^2\Big\} &= 2\,(t-s)^2 \end{cases}. \tag{2.1}$$

By tradition in stochastic calculus, one uses the notation

$$\mathbf{P}_x\Big\{(B(t_1),\dots,B(t_n))\in A\Big\} = \mathbf{P}\Big\{(B(t_1),\dots,B(t_n))\in A \,\Big|\, B(0)=x\Big\}$$

to denote the probability

$$\mathbf{P}\Big\{(B^x(t_1),\dots,B^x(t_n))\in A\Big\} \qquad \text{for} \quad A\in\mathcal{B}(\mathbb{R}^n) \quad \text{and} \quad x\in\mathbb{R}.$$

Since BM is not completely determined by Definition 2.1 (Exercise 18), the conditional probability $\mathbf{P}\{\cdot\,|\,B(0)=x\}$ is not well-defined. Rather, it expresses the fact that the probability for the event $\{(B(t_1),\dots,B(t_n))\in A\}$ is determined when the starting value $B(0)=x$ of BM is specified (Exercise 19).

Theorem 2.6 (SPACE HOMOGENEITY) *For $A\in\mathcal{B}(\mathbb{R}^n)$ and $x\in\mathbb{R}$, we have*

$$\mathbf{P}\Big\{(B(t_1),\dots,B(t_n))\in A+x \,\Big|\, B(0)=x\Big\} = \mathbf{P}\Big\{(B(t_1),\dots,B(t_n))\in A \,\Big|\, B(0)=0\Big\}.$$

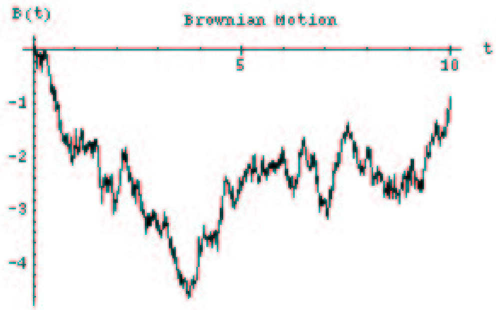
EXERCISE 22 Prove Theorem 2.6.

We shall plot two sample paths of BM $B^0(t)$ for $t\in[0,10]$. Thus we plot $[0,10]\ni t\to B^0(t) = B^0(\omega,t)$ for two different ω in the probability space $\Omega=(\Omega,\mathcal{F},\mathbf{P})$.

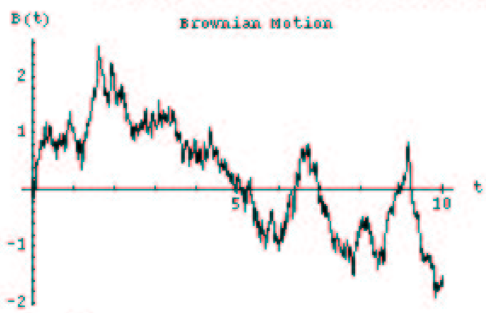
We cannot plot all values $\{B^0(t)\}_{t\in[0,10]}$, and thus decide to plot $\{B^0(\frac{k}{100})\}_{k=1}^{1000}$. Since the increments $\{B^0(\frac{k}{100})-B^0(\frac{k-1}{100})\}_{k=1}^{1000}$ are independent $\mathrm{N}(0,\frac{1}{100})$-distributed, we first create 1000 independent $\mathrm{N}(0,\frac{1}{100})$-distributed increments $\{\mathrm{Incr}(k)\}_{k=1}^{1000}$, and then compute $\{B^0(\frac{k}{100})\}_{k=1}^{1000}$, by adding the incerements:

$$\{B^0(\tfrac{k}{100})\}_{k=1}^{1000} = \Big\{\sum_{i=1}^{k}(B^0(\tfrac{i}{100})-B^0(\tfrac{i-1}{100}))\Big\}_{k=1}^{1000} =_{\text{same fidi's}} \Big\{\sum_{i=1}^{k}\mathrm{Incr}(i)\Big\}_{k=1}^{1000}.$$

```
In[1]:= << Statistics`ContinuousDistributions`

In[2]:= Incr = N[Table[Random[NormalDistribution[0, 0.1]], {1000}]];

In[3]:= B = {Incr[[1]]}; For[k = 2, k ≤ 1000, k++ , B = Join[B, {B[[k - 1]] + Incr[[k]]}]];

In[4]:= ListPlot[B, Ticks → {{{100, ""}, {200, ""}, {300, ""}, {400, ""}, {500, "5", 0.02},
            {600, ""}, {700, ""}, {800, ""}, {900, ""}, {1000, "10", 0.02}}, Automatic},
            AxesLabel → {"t", "B(t)"}, PlotLabel → "    Brownian Motion", PlotJoined → True]
```




```
Out[4]= - Graphics -

In[5]:= Incr = N[Table[Random[NormalDistribution[0, 0.1]], {1000}]];

In[6]:= B = {Incr[[1]]}; For[k = 2, k ≤ 1000, k++ , B = Join[B, {B[[k - 1]] + Incr[[k]]}]];

In[7]:= ListPlot[B, Ticks → {{{100, ""}, {200, ""}, {300, ""}, {400, ""}, {500, "5", 0.02},
            {600, ""}, {700, ""}, {800, ""}, {900, ""}, {1000, "10", 0.02}}, Automatic},
            AxesLabel → {"t", "B(t)"}, PlotLabel → "    Brownian Motion", PlotJoined → True]
```




```
Out[7]= - Graphics -
```

*Remark 2.7 `Mathematica` uses the notation $\mathrm{N}(0,\sigma)$ instead of $\mathrm{N}(0,\sigma^2)$, so that $\mathrm{N}(0,\frac{1}{10})$ means a Gaussian random variable, with expected value 0 and variance $\frac{1}{100}$. Both notations $\mathrm{N}(0,\sigma)$ and $\mathrm{N}(0,\sigma^2)$ are used in the literature, with the latter one standard in non-elementary texts: Beware of mistakes caused by this! #

2.2 Gaussian Stochastic Processes

Definition 2.8 *A* *stochastic process* $\{X(t)\}_{t\in T}$ *is Gaussian* *(or* *normal*), *if for each choice of constants* $a_1,\ldots,a_n\in\mathbb{R}$, *parameters* $t_1,\ldots,t_n\in T$, *and* $n\in\mathbb{N}$,

the linear combination $\displaystyle\sum_{i=1}^{n} a_i X(t_i)$ *is univariate Gaussian distributed.*

An *$\mathbb{R}^n$-valued random variable* $X=(X_1,\ldots,X_n)$ *is Gaussian*, *if* $\{X_i\}_{i\in\{1,\ldots,n\}}$ *is a Gaussian process.*

Definition 2.9 *The* <u>*covariance function* $r : T \times T \to \mathbb{R}$</u> *for a (Gaussian) stochastic process* $\{X(t)\}_{t\in T}$ *is defined*

$$r(s,t) = \mathbf{Cov}\{X(s), X(t)\} = \mathbf{E}\{X(s)X(t)\} - \mathbf{E}\{X(s)\}\mathbf{E}\{X(t)\}.$$

EXERCISE 23 Show that a Gaussian process $\{X(t)\}_{t\in T}$, with covariance function r and mean function $m(t) = \mathbf{E}\{X(t)\}$, has Laplace-Stieltjes transform

$$\mathbf{E}\Big\{\mathrm{e}^{\lambda_1 X(t_1)+\ldots+\lambda_n X(t_n)}\Big\} = \exp\left\{\begin{pmatrix}\lambda_1 & \ldots & \lambda_n\end{pmatrix}\begin{pmatrix} m(t_1)\\ \vdots \\ m(t_n)\end{pmatrix} + \frac{1}{2}\begin{pmatrix}\lambda_1 & \ldots & \lambda_n\end{pmatrix}\Big(r(t_i,t_j)\Big)\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}\right\}.$$

Theorem 2.10 *The* `fidi's` *of a Gaussian process* $\{X(t)\}_{t\in T}$ *are determined by the covariance function together with the mean function* $m(t) = \mathbf{E}\{X(t)\}$.

Proof. By Exercise 23, r and m determine the Laplace-Stieltjes transform

$$\mathbf{E}\Big\{\mathrm{e}^{\lambda_1 X(t_1)+\ldots+\lambda_n X(t_n)}\Big\} = \int_{\mathbb{R}^n} \mathrm{e}^{\lambda_1 x_1+\ldots+\lambda_n x_n}\, dF_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n). \quad \square$$

EXERCISE 24 B^x is Gaussian with $\mathbf{Cov}\{B^x(s), B^x(t)\} = \underline{s\wedge t} = \min\{s,t\}$.

Corollary 2.11 *If* $\{X(t)\}_{t\in T}$ *is a Gaussian process, and* $R, S \subseteq T$, $\{X(t)\}_{t\in R}$ *and* $\{X(t)\}_{t\in S}$ *are independent iff.* $\mathbf{Cov}\{X(r), X(s)\} = 0$ *for all* $r\in R$ *and* $s \in S$.

**Proof.* The implication to the right is immediate. For the implication to the left, notice that X takes non-random values on $R\cap S$, since variances are zero on $R\cap S$, by the condition on covariances. Hence it is enough to prove that $\{X(t)\}_{t\in R\setminus S}$ and $\{X(t)\}_{t\in S\setminus R}$ are independent. For this, by Theorem 1.15, it is enough to prove that $\{X(r_i)\}_{i\in\{1,\ldots,n\}}$ and $\{X(s_i)\}_{i\in\{1,\ldots,n\}}$ are independent for $r_1,\ldots,r_n \in R\setminus S$ and $s_1,\ldots,s_n \in S\setminus R$. Let $\{\hat{X}(r_i)\}_{i\in\{1,\ldots,n\}}$ and $\{\hat{X}(s_i)\}_{i\in\{1,\ldots,n\}}$ be independent, with $\{\hat{X}(r_i)\}_{i\in\{1,\ldots,n\}} =_{\text{same fidi's}} \{X(r_i)\}_{i\in\{1,\ldots,n\}}$ and $\{\hat{X}(s_i)\}_{i\in\{1,\ldots,n\}} =_{\text{same fidi's}} \{X(s_i)\}_{i\in\{1,\ldots,n\}}$. Since $\{\hat{X}(t)\}_{t\in\{r_1,\ldots,r_n\}\cup\{s_1,\ldots,s_n\}}$ has the same mean and covariance function as $\{X(t)\}_{t\in\{r_1,\ldots,r_n\}\cup\{s_1,\ldots,s_n\}}$, Theorem 2.10 shows that the processes have common `fidi's`. Hence $\{X(r_i)\}_{i\in\{1,\ldots,n\}}$ and $\{X(s_i)\}_{i\in\{1,\ldots,n\}}$ are independent. $\square$

Example 2.12 The process $X(t) = \int_0^t B^0(\tau)\,d\tau$, $t\geq 0$, is Gaussian, since

$$\sum_{i=1}^n a_i X(t_i) = \lim\left\{\sum_{i=1}^n a_i \sum_{\tau_j\leq t_i} B^0(\tau_j)\,(\tau_j-\tau_{j-1}) : \begin{array}{c} 0=\tau_0<\tau_1<\ldots<\tau_k=t_{\max}\\ k\in\mathbb{N},\ \max_{1\leq j\leq k}\tau_j-\tau_{j-1}\to 0\end{array}\right\},$$

and Gaussian random variables can only converge to Gaussian limits (e.g., by inspection of Exercise 23). The mean function is $m(t) = \mathbf{E}\{X(t)\} = 0$ by symmetry, while by *Fubini's Theorem* (e.g., [34, Chapter 7])*, the covariance function is

$$r(s,t) = \mathbf{E}\{X(s)X(t)\} = \int_0^s \int_0^t \mathbf{E}\{B^0(u)B^0(v)\}\,dudv = \int_0^s \int_0^t u \wedge v\,dudv = \ldots\ . \quad \#$$

2.3 Sample Path Properties of BM

Theorem 2.13 *BM has quadratic variation $[B](t) = t$, in the sense of convergence in mean-square.*

Proof. By independence of increments and (2.1), we have

$$\mathbf{Var}\left\{\sum_{i=1}^n (B(t_i)-B(t_{i-1}))^2\right\} = \sum_{i=1}^n \mathbf{Var}\left\{(B(t_i)-B(t_{i-1}))^2\right\} = \sum_{i=1}^n 2\,(t_i - t_{i-1})^2$$
$$\leq 2\,t \max_{1\leq i\leq n} t_i - t_{i-1}$$

which goes to zero, as we consider partitions $0 = t_0 < t_1 < \ldots < t_n = t$ of $[0,t]$, and send $\max_{1\leq i\leq n} t_i - t_{i-1} \rightarrow 0$. This means that the sum converges in mean-square to a constant, which must be equal to the limit of the mean of the sum [see (2.1)]

$$\mathbf{E}\left\{\sum_{i=1}^n (B(t_i)-B(t_{i-1}))^2\right\} = \sum_{i=1}^n \mathbf{E}\left\{(B(t_i)-B(t_{i-1}))^2\right\} = \sum_{i=1}^n t_i - t_{i-1} = t. \quad \Box$$

Theorem 2.14 *BM has infinite variation.*

Proof. Since BM is continuous, it cannot have finite variation with non-zero probability, since this would make the quadratic variation zero, with non-zero probability, by Theorem 1.11, which contradicts Theorem 2.13. □

Exercise 25 Prove that BM is not continuously differentiable with non-zero probability (without the help of Corollary 2.18 below).

Definition 2.15 *A stochastic process $\{X(t)\}_{t\geq 0}$ has independent increments, if*

$$X(t)-X(s) \quad \textit{is independent of} \quad \{X(r)\}_{r\in[0,s]} \quad \textit{for} \quad 0 \leq s < t.$$

Definition 2.16 *A stochastic process $\{X(t)\}_{t\geq 0}$ has stationary increments, if*

$$\{X(t+h)-X(h)\}_{t\in\mathbb{R}} =_{\text{same fidi's}} \{X(t)-X(0)\}_{t\in\mathbb{R}} \qquad \textit{for each constant} \quad h \geq 0.$$

*Theorem 2.17 *Let $\{X(t)\}_{t\geq 0}$ be a stochastic process, that has independent and stationary increments, such that, for some constant $\varepsilon>0$,*

$$\lim_{n\to\infty} n^\varepsilon\,\mathbf{P}\Big\{|X(1/n)-X(0)|\leq K/n\Big\} = 0 \qquad \textit{for each constant} \quad K>0. \tag{2.2}$$

The process X is not differentiable anywhere, with probability one.

**Proof.* The following proof is adapted from one for BM in [19, p. 18] (who in turn got their proof from earlier sources): Pick an $N\in\mathbb{N}$ such that $N\varepsilon\geq 2$. Notice that, if X is differentiable at some $s\geq 0$, then we have

$$|X(t)-X(s)|\leq \ell(t-s)\leq \ell(N+2)/n \qquad \text{for} \quad t\in(s,s+(N+2)/n),$$

for all sufficiently large $\ell,n\in\mathbb{N}$. Chosing $k\in\mathbb{N}$ such that $k/n,\ldots,(k+N)/n\in(s,s+(N+2)/n)$, this gives

$$|X((i+1)/n)-X(i/n)|\leq|X((i+1)/n)-X(s)|+|X(s)-X(i/n)|\leq 2\ell(N+2)/n$$

for $i=k,\ldots,k+N-1$. It follows that the event that BM is differentiable somewhere is contained in the event

$$\bigcup_{\ell=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\bigcup_{k=1}^{[n^2]}\bigcap_{i=k}^{k+N-1}\Big\{|X((i+1)/n)-X(i/n)|\leq\frac{2\ell(N+2)}{n}\Big\}.$$

Hence it is enough to prove that

$$\mathbf{P}\Bigg\{\bigcap_{n=m}^{\infty}\bigcup_{k=1}^{[n^2]}\bigcap_{i=k}^{k+N-1}\Big\{|X((i+1)/n)-X(i/n)|\leq\frac{2\ell(N+2)}{n}\Big\}\Bigg\}=0,$$

which in turn will follow if

$$\mathbf{P}\Bigg\{\bigcup_{k=1}^{[n^2]}\bigcap_{i=k}^{k+N-1}\Big\{|X((i+1)/n)-X(i/n)|\leq\frac{2\ell(N+2)}{n}\Big\}\Bigg\}\to 0 \qquad \text{as} \quad n\to\infty.$$

However, by independence and stationarity of increments, together with (2.2), the probability on the left-hand side is at most

$$\begin{aligned}
&\sum_{k=1}^{[n^2]}\prod_{i=k}^{k+N-1}\mathbf{P}\Big\{|X((i+1)/n)-X(i/n)|\leq\frac{2\ell(N+2)}{n}\Big\}\\
&\qquad=[n^2]\bigg(\mathbf{P}\Big\{|X(1/n)-X(0)|\leq\frac{2\ell(N+2)}{n}\Big\}\bigg)^N\\
&\qquad=n^{-N\varepsilon}[n^2]\bigg(n^\varepsilon\,\mathbf{P}\Big\{|X(1/n)-X(0)|\leq\frac{2\ell(N+2)}{n}\Big\}\bigg)^N\to 0 \qquad \text{as} \quad n\to\infty. \quad \square
\end{aligned}$$

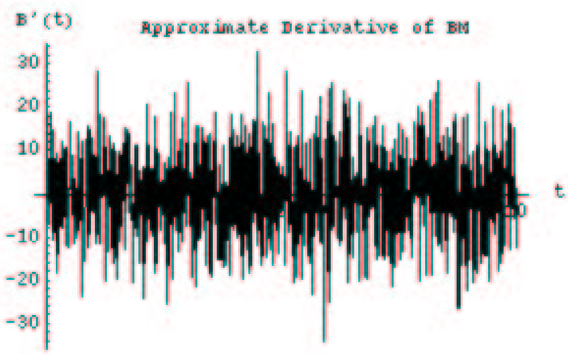
Corollary 2.18 *BM is not differentiable anywhere, with probability one.*

**Proof.* The definition of BM gives the hypothesis of Theorem 2.17, except (2.2) (see also Exercise 39 below). However, we get (2.2) from observing that

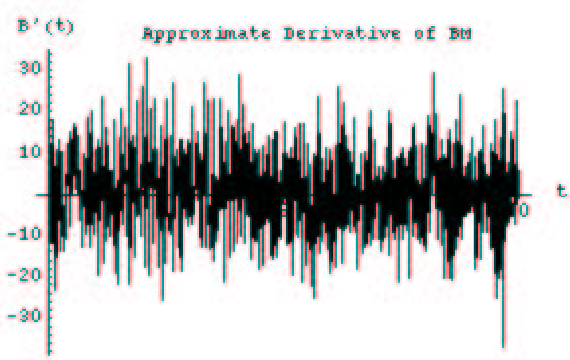
$$\mathbf{P}\left\{|X(1/n)-X(0)|\le\frac{K}{n}\right\}=\mathbf{P}\left\{|\mathrm{N}(0,1/n)|\le\frac{K}{n}\right\}=\mathbf{P}\left\{|\mathrm{N}(0,1)|\le\frac{K}{\sqrt{n}}\right\}\le\frac{2\,K}{\sqrt{2\pi\,n}}.\quad\Box$$

It is illuminating to plot the approximative derivative process $\{(B(t+h)-B(t))/h\}_{t\geq0}$, for a small value of $h>0$, to illustrate the non-differentiability of BM:

```
In[1]:= << Statistics`ContinuousDistributions`

In[2]:= Incr = N[Table[Random[NormalDistribution[0, 0.1]], {1001}]];

In[3]:= B = {Incr[[1]]}; For[k = 2, k ≤ 1001, k++, B = Join[B, {B[[k - 1]] + Incr[[k]]}]];

In[4]:= Bprime = Table[(B[[k + 1]] - B[[k]]) / (1 / 100), {k, 1, 1000}];

In[5]:= ListPlot[Bprime, Ticks → {{{100, ""}, {200, ""}, {300, ""}, {400, ""},
            {500, "5", 0.02}, {600, ""}, {700, ""}, {800, ""}, {900, ""},
            {1000, "10", 0.02}}, Automatic}, AxesLabel → {"t", "B'(t)"},
            PlotLabel → "      Approximate Derivative of BM", PlotJoined → True]
```




```
Out[5]= - Graphics -

In[6]:= Incr = N[Table[Random[NormalDistribution[0, 0.1]], {1001}]];

In[7]:= B = {Incr[[1]]}; For[k = 2, k ≤ 1001, k++, B = Join[B, {B[[k - 1]] + Incr[[k]]}]];

In[8]:= Bprime = Table[(B[[k + 1]] - B[[k]]) / (1 / 100), {k, 1, 1000}];

In[9]:= ListPlot[Bprime, Ticks → {{{100, ""}, {200, ""}, {300, ""}, {400, ""},
            {500, "5", 0.02}, {600, ""}, {700, ""}, {800, ""}, {900, ""},
            {1000, "10", 0.02}}, Automatic}, AxesLabel → {"t", "B'(t)"},
            PlotLabel → "      Approximate Derivative of BM", PlotJoined → True]
```




```
Out[9]= - Graphics -
```

⋆Remark 2.19 The hypothesis of Theorem 2.17 is satisfied by, for example, processes X with stationary and independent increments, that are self-similar with index $1/\alpha \in [1/2, 1)$ (see Definition 3.16 below), such that $X(1)$ has a density function that is bounded in a neighbourhood of zero. Indeed, this is how the argument goes in the proof of Corollary 2.18, with $1/\alpha = 1/2$. These requirements in turn are defining properties for (non-zero) α-stable Lévy processes with $\alpha \in (1, 2]$ (e.g. [4, Chapter VIII] and [35, Section 7.5]), where $\alpha = 2$ is BM (multiplied by a constant). #

2.4 Introduction to Martingale Theory

Smartingales is the class of all martingales, submartingales and supermartingales. These are very important in stochastic calculus, both as probabilistic tools in proofs, and as noise processes in stochastic differential equations (SDE):

Definition 2.20 *A family $\mathbb{F} = \{\mathcal{F}_t\}_{t \in T}$ of σ-algebras is a filtration, if*

$$\mathcal{F}_s \subseteq \mathcal{F}_t \qquad \text{for} \quad s, t \in T \quad \text{with} \quad s < t.$$

A filtration is augmented if $(\Omega, \mathcal{F}, \mathbf{P})$ is complete (e.g., [34, p. 29])⋆, *and each $\mathcal{F}_t$ contains all $\mathbf{P}$-null-sets of $\mathcal{F}$.*
A stochastic process $\{X(t)\}_{t \in T}$ is adapted to $\mathbb{F}$, if $X(t)$ is $\mathcal{F}_t$-measurable for $t \in T$.

Definition 2.21 *Let $\mathbb{F} = \{\mathcal{F}_t\}_{t \in T}$ be a filtration and $X = \{X(t)\}_{t \in T}$ a stochastic process that is adapted to $\mathbb{F}$. Assume that*

$$\mathbf{E}\{|X(t)|\} < \infty \qquad \text{for} \quad t \in T.$$

The process X is a

$$\begin{cases} \text{martingale wrt. } \mathbb{F} \text{ if} & \mathbf{E}\{X(t)|\mathcal{F}_s\} = X(s) \\ \text{submartingale wrt. } \mathbb{F} \text{ if} & \mathbf{E}\{X(t)|\mathcal{F}_s\} \geq X(s) \\ \text{supermartingale wrt. } \mathbb{F} \text{ if} & \mathbf{E}\{X(t)|\mathcal{F}_s\} \leq X(s) \end{cases} \qquad \text{for} \quad s, t \in T \quad \text{with} \quad s < t.$$

Definition 2.22 *A stochastic process $X = \{X(t)\}_{t \in T}$ is a smartingale wrt. itself, if X is a smartingale wrt. $\mathbb{F} = \{\sigma(X(s) : s \in T,\, s \leq t)\}_{t \in T}$.*

Example 2.23 If $\{\xi_i\}_{i=1}^{\infty}$ are independent random variables, with finite and positive (negative/zero) expected values, then $\{\sum_{i=1}^{n} \xi_i\}_{n \in \mathbb{N}}$ is a submartingale (supermartingale/martingale) wrt. itself. #

A standard reference for discrete time martingales is [8, Chapter 9]*. Continuous time martingale theory is a lot more similar to the discrete theory, than can be expected in general when turning from discrete time to continuous. Nevertheless, there are many differences and new difficulties (albeit often only of a technical nature).

EXERCISE 26 Let $g:\mathbb{R}\to\mathbb{R}$ be a convex increasing function, and X a submartingale. Show that $g(X)$ is a submartingale, provided that $\mathbf{E}\{|g(X)|\}<\infty$.

Standard texts on continuous time smartingales (e.g., [9])*, usually begin with a handful results, showing that it is not a real restriction to assume that a smartingale is right-continuous, with limits from the left, (called càdlàg= *"continu à droit avec des limites à gauche"*), or possibly, left-continuous, with limits from the right (càglàd ?), since much subsequent theory require this.

Theorem 2.24 (e.g., [9, Theorem 1.4.1]) *Let $\{X(t)\}_{t\geq 0}$ be a smartingale and S a countable dense subset of $[0,\infty)$* (e.g., [34, p. 59])*. *Left and right limits over S*

$$\lim_{s\in S,\, s\to t,\, s<t} X(s) \qquad \text{and} \qquad \lim_{s\in S,\, s\to t,\, s>t} X(s)$$

exists and are finite for each $t\geq 0$, with probability one.

EXERCISE 27 Show that a version of a smartingale is a smartingale (of the same kind), when the filtration is augmented.

Theorem 2.25 *A smartingale $\{X(t)\}_{t\geq 0}$, with an augmented filtration, has a version that is a smartingale (of the same kind), and that possesses finite limits from the left and from the right everywhere, with probability one.*

**Proof.* By Theorem 1.33, X has a separable version $\{\hat{X}(t)\}_{t\geq 0}$, that is a smartingale of the same kind, by Exercise 27. Let S be a separator of $\hat{X}$ (Definition 1.31). Pick a $t\geq 0$ and a sequence $\{t_n\}_{n=1}^{\infty}$ such that $t_n\to t$ from the left, or from the right, as $n\to\infty$. The claim is that $\lim_{n\to\infty} X(t_n)$ exists, and only depends on t and if the limit is from the left or right, but not on the particular sequence $\{t_n\}_{n=1}^{\infty}$.

By separability, there is a $s_n\in S$, located to the left or right of t, respectively, such that $|X(t_n)-X(s_n)|\leq 1/n$. Since $X(t_n)-X(s_n)\to 0$, it is enough to prove that $\lim_{n\to\infty} X(s_n)$ exists, and only depends .... This follows from Theorem 2.24. □

Definition 2.26 *For a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, we define $\mathcal{F}_{t+}=\bigcap_{s>t}\mathcal{F}_s$ for $t\geq 0$, and say that $\{\mathcal{F}_t\}_{t\geq 0}$ is right-continuous if $\mathcal{F}_{t+}=\mathcal{F}_t$ for $t\geq 0$.*

EXERCISE 28 Show that $\mathbb{F}^+ = \{\mathcal{F}_{t^+}\}_{t\geq 0}$ is a right-continuous filtration, for every filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$.

Many results for smartingales $(X,\mathbb{F})$ require the filtration $\mathbb{F}$ to be right-continuous. (The requirement that $\mathbb{F}$ is augmented and right-continuous is sometimes called the usual conditions.) Thus it is natural to replace $\mathbb{F}$ with $\mathbb{F}^+$ (Exercise 28).

Theorem 2.27 (e.g., [9, Theorem 1.4.2])* *Let $(X,\mathbb{F})$ be a smartingale. If X has limits from the right, with probability one, then $(X(\cdot^+),\mathbb{F}^+)$ is a right-continuous smartingale (of the same kind).*

The essential steps to prove Theorem 2.27 are the same as those to prove Theorem 2.28 below. We do the latter proof, leaving the former one as a stared exercise.

EXERCISE 29 If X is a smartingale, what can be said about $-X$?

Theorem 2.28 *Let $(X,\mathbb{F})$ be a smartingale. Let $\mathbb{F}$ and $\mathbf{E}\{X(\cdot)\}$ be right-continuous, and $\mathbb{F}$ augmented. If X has limits from the right with probability one, then $X(\cdot^+)$ is a right-continuous smartingale (of the same kind), and a version of X.*

(Recall that X has a version with limits from the right, by Theorem 2.25.) To prove Theorem 2.28, we need the following standard result for discrete martingales:

***Lemma 2.29** (e.g., [8, Theorem 9.4.7]) *If $\{X_n\}_{n\leq 0}$ is a smartingale, such that $\lim_{n\to-\infty}\mathbf{E}\{X_n\}$ exists and is finite, then there exists a random variable Y, with finite mean, such that $\lim_{n\to-\infty}\mathbf{E}\{|X_n-Y|\}=0$.*

**Proof of Theorem 2.28* (after [9, Theorem 1.4.3]). Since $X(\cdot^+)$ is right-continuous, it is enough to prove that $X(\cdot^+)$ is a version of X (recall Exercise 27). Pick a $t\geq 0$ and $t_1>t_2>\ldots>t$ such that $t_n\downarrow t$. Since $\{X(t_n)\}_{n\in\{\ldots,3,2,1\}}$ is a smartingale, and $\lim_{n\to\infty}\mathbf{E}\{X(t_n)\} = \mathbf{E}\{X(t)\}$ exists, by right-continuity of $\mathbf{E}\{X(\cdot)\}$, Lemma 2.29 gives $\lim_{n\to\infty}\mathbf{E}\{|X(t_n)-Y|\}=0$. And so $\lim_{n\to\infty}\mathbf{E}\{|X(t_n)-X(t^+)|\}=0$, since

$$\mathbf{E}\{|X(t^+)-Y|\} = \mathbf{E}\{\liminf_{n\to\infty}|X(t_n)-Y|\} \leq \liminf_{n\to\infty}\mathbf{E}\{|X(t_n)-Y|\}=0,$$

by *Fatou's Lemma*. Hence *Jensen's Inequality* shows that

$$\Big|\mathbf{E}\{X(t_n)\}-\mathbf{E}\{X(t^+)\}\Big| \leq \mathbf{E}\{|X(t_n)\}-X(t^+)|\}\to 0 \qquad \text{as} \quad n\to\infty,$$

so that $\lim_{n\to\infty}\mathbf{E}\{X(t_n)\}=\mathbf{E}\{X(t^+)\}$. And so right-continuity of $\mathbf{E}\{X(\cdot)\}$ gives

$$\mathbf{E}\{X(t)\} = \lim_{n\to\infty}\mathbf{E}\{X(t_n)\} = \mathbf{E}\{X(t^+)\}.$$

From the fact that $\mathbf{E}\{|X(t_n)-X(t^+)|\}\to 0$ as $n\to\infty$, we get $\mathbf{E}\Big\{\Big|\mathbf{E}\{X(t_n)|\mathcal{F}_t\}-\mathbf{E}\{X(t^+)|\mathcal{F}_t\}\Big|\Big\}\to 0$, by Exercise 68. In particular, $\mathbf{E}\{X(t_n)|\mathcal{F}_t\}\to\mathbf{E}\{X(t^+)|\mathcal{F}_t\}$ in probability (convergence in measure). It follows that $\mathbf{E}\{X(t_{n_k})|\mathcal{F}_t\}\to\mathbf{E}\{X(t^+)|\mathcal{F}_t\}$ a.s. (for $\mathbf{P}$-almost all $\omega\in\Omega$) as $k\to\infty$, for some subsequence $\{t_{n_k}\}_{k=1}^\infty\subseteq\{t_n\}_{n=1}^\infty$.

Now notice that $X(t^+)$ is adapted to $\mathcal{F}_t=\mathcal{F}_{t^+}$, by Theorem 2.27. In the case when $(X,\mathbb{F})$ is a supermartingale, so that $X(t)\geq\mathbf{E}\{X(t_{n_k})|\mathcal{F}_t\}$, this gives

$$X(t)\geq\lim_{k\to\infty}\mathbf{E}\{X(t_{n_k})|\mathcal{F}_t\}=\mathbf{E}\{X(t^+)|\mathcal{F}_t\}=X(t^+),$$

while instead $X(t)\leq X(t^+)$ when X is a submartingale. Hence $X(t)-X(t^+)$ is a random variable that either is negative a.s., or positive a.s., and has zero mean (as seen above). And so $X(t)=X(t^+)$ a.s., so that $X(\cdot^+)$ is a version of X. □

Many arguments in *Stochastic Calculus* require finite second moments. This makes *square-integrable smartingales* important:

Definition 2.30 *A stochastic process $\{X(t)\}_{t\in T}$ is square integrable [has bounded second moments] if $\mathbf{E}\{X(t)^2\}<\infty$ for each $t\in T$ [if $\sup_{t\in T}\mathbf{E}\{X(t)^2\}<\infty$].*

⋆Remark 2.31 There are different opinions in the literature, on whether square-integrable, which some authors call *locally square-integrable*, or $\mathbb{L}^2$-*process*, should mean "just square-integrable", or rather bounded second moments, which these authors then call square-integrable (sick, in my opinion).

Processes that are *integrable* [has bounded first moments] are defined by obvious changes in Definition 2.30, and enjoy a simliar confusion of terminology. #

2.5 Martingale Properties of BM

Corollary 2.32 (Martingale Property of BM) *BM B is a martingale wrt. itself.*

Proof. This follows from Exercise 41 below. □

Corollary 2.33 *$\{B(t)^2-t\}_{t\geq 0}$ is a martingale wrt. $\{\sigma(B(s):0\leq s\leq t)\}_{t\geq 0}$.*

Proof. This follows from Theorem 4.8 below. □

Exercise 30 $\{e^{cB(t)-c^2t/2}\}_{t\geq 0}$ is a martingale wrt. $\{\sigma(B(s):0\leq s\leq t)\}_{t\geq 0}$

3 Third Lecture

3.1 Conditional Expectations (continued)

EXERCISE 31 Let X be an $\mathbb{R}^n$-valued random variable. Show that

$$\int_{\omega\in\Omega} g(X(\omega))\,d\mathbf{P}(\omega) = \int_{x\in\mathbb{R}^n} g(x)\,dF_X(x) \qquad \text{for measurable functions} \quad g:\mathbb{R}^n\to\mathbb{R}.$$

(The integrals are well-defined simultaneously, and coincide when well-defined).

Recall that the distribution of an $\mathbb{R}^n$-valued random variable X is the probability measure $dF_X(\cdot) = (P\circ X^{-1})(\cdot)$ on $\mathbb{R}^n$ $[\mathcal{B}(\mathbb{R}^n)]$. Equivalently, dF_X is the Stieltjes measure associated with the distribution function

$$F_X(x_1,\ldots,x_n) = \mathbf{P}\{X_1\le x_1,\ldots,X_n\le x_n\} = (\mathbf{P}\circ X^{-1})((-\infty,x_1]\times\ldots\times(-\infty,x_n]).$$

Example 3.1 (ABSOLUTE CONTINUITY) An $\mathbb{R}^n$-valued random variable X is absolutely continuous, if there exists an integrable function $f_X:\mathbb{R}^n\to\mathbb{R}$, such that

$$\mathbf{P}\{X\in A\} = \int_A f_X(x)\,dx \qquad \text{for} \quad A\in\mathcal{B}(\mathbb{R}^n).$$

The density function f_X must be positive a.e., and can be chosen positive everywhere (since integrals of the density are not affected if we change its values on a null-set).

Given an $\mathbb{R}^n$-valued absolutely continuous random variable X, we have

$$\mathbf{E}\{g(X)\} = \int_{\mathbb{R}^n} g(x)f_X(x)\,dx \qquad \text{for measurable functions} \;\; g:\mathbb{R}^n\to\mathbb{R} \tag{3.1}$$

(both sides are well-defined simultaneously, and coincide when well-defined). This follows from approximating g^+ and g^- with increasing sequences of simple functions.

For a real absolutely continuous random variable X, the density function is the derivative of the distribution function $f_X(x) = F_X'(x) = \frac{d}{dx}\mathbf{P}\{X\le x\}$ a.e. (e.g., [8, pp. 10-11])⋆. [Since every distribution function is differentiable a.e. (e.g., [8, p. 11])⋆, that property on its own does not imply that X is absolutely continuous!] #

EXERCISE 32 Let X and Y be random variables, with values in $\mathbb{R}$ and $\mathbb{R}^n$, respectively, and $\mathbf{E}\{|X|\}<\infty$. Show that there exists a measurable function $\varphi:\mathbb{R}^n\to\mathbb{R}$ such that $\mathbf{E}\{X|Y\} = \varphi(Y)$.

Definition 3.2 (POINTWISE CONDITIONAL EXPECTATION) *For random variables X and Y, with values in $\mathbb{R}$ and $\mathbb{R}^n$, respectively, and $\mathbf{E}\{|X|\}<\infty$, we write*

$$\mathbf{E}\{X|Y{=}y\} \equiv \varphi(y) \qquad \textit{where} \qquad \varphi(Y) = \mathbf{E}\{X|Y\} \qquad \text{(cf. Exercise 32).}$$

Notice that the values of φ outside $Y(\Omega)$ (the possible values of Y) are unimportant, and do no affect the validity of $\mathbf{E}\{X|Y\} = \varphi(Y)$ (cf. Remark 3.8 below).

Theorem 3.3 *For random variables X and Y, with $\mathbf{E}\{|X|\} < \infty$, we have*

$$\int_{y\in\mathbb{R}} \mathbf{E}\{X|Y=y\}\, dF_Y(y) = \mathbf{E}\{X\}.$$

Proof. Writing $g(Y) = \mathbf{E}\{X|Y\}$ (Exercise 32), so that $\mathbf{E}\{X|Y=y\} = g(y)$ (see Definition 3.3), Exercises 12 and 31 show that

$$\int_{\mathbb{R}} \mathbf{E}\{X|Y=y\}\, dF_Y(y) = \int_{\mathbb{R}} g(y)\, dF_Y(y) = \int_{\Omega} g(Y)\, d\mathbf{P} = \mathbf{E}\Big\{\mathbf{E}\{X|Y\}\Big\} = \mathbf{E}\{X\}. \quad \square$$

Example 3.4 (Elementary Conditional Expectations) Let (X,Y) be an $\mathbb{R}^n\times\mathbb{R}$-valued random variable, that is absolutely continuous with a positive density function $f_{X,Y}$. By Example 3.1 together with *Fubini's Theorem*, we have

$$\mathbf{P}\{Y\in B\} = \mathbf{P}\{X\in\mathbb{R}^n,\, Y\in B\} = \int_{\mathbb{R}^n\times B} f_{X,Y}(x,y)\, dxdy = \int_B \left(\int_{\mathbb{R}^n} f_{X,Y}(x,y)\, dx\right) dy,$$

so that also Y is absolutely continuous, with density function

$$f_Y(y) = \int_{\mathbb{R}^n} f_{X,Y}(x,y)\, dx \qquad \text{for} \quad y\in\mathbb{R}.$$

Here $f_Y(y)=0$ implies $f_{X,Y}(x,y)=0$ for almost all $x\in\mathbb{R}^n$ (since $f_{X,Y}$ is positive).

For a measurable function $g:\mathbb{R}^n\to\mathbb{R}$, with $\mathbf{E}\{|g(X)|\}<\infty$, (3.1) shows that

$$\mathbf{E}\{g(X)\} = \int_{\mathbb{R}^n\times\mathbb{R}} g(x)\, f_{X,Y}(x,y)\, dxdy \qquad \text{is well-defined}$$

[taking $g(x,y)=g(x)$]. By application of *Fubini's Theorem*, it follows that

$$E(g(X),y) \equiv \int_{\mathbb{R}^n} g(x) f_{X,Y}(x,y)\, dx \qquad \text{is an integrable function of } y\in\mathbb{R}.$$

Since $f_Y(y)\neq 0$ when $E(g(X),y)\neq 0$, we may define

$$E(g(X)|y) \equiv \begin{cases} E(g(X),y)/f_Y(y) & \text{when} \quad E(g(X),y)\neq 0 \\ 0 & \text{when} \quad E(g(X),y)=0 \end{cases}.$$

We claim that

$$\mathbf{E}\{g(X)|Y=y\} = E(g(X)|y) = \frac{1}{f_Y(y)}\int_{\mathbb{R}^n} g(x) f_{X,Y}(x,y)\, dxdy \qquad (0/0=0).$$

Proof. This follows from observing that, for $\Lambda\in\sigma(Y)$, we have $\Lambda = Y^{-1}(A) = \{\omega\in\Omega : Y(\omega)\in A\}$ for some $A\in\mathcal{B}(\mathbb{R})$ (see Lecture 1), so that [using (3.1) twice]

$$\int_\Lambda g(X)\, d\mathbf{P} = \int_{\mathbb{R}^n\times A} g(x) f_{X,Y}(x,y)\, dxdy = \int_A E(g(X)|y) f_Y(y)\, dy = \int_\Lambda E(g(X)|Y)\, d\mathbf{P}.$$

This gives $\mathbf{E}\{g(X)|Y\} = E(g(X)|Y)$, so that $\mathbf{E}\{g(X)|Y=y\} = E(g(X)|y)$. #

EXERCISE 33 Let X and Y be random variables, and $f:\mathbb{R}^2\to\mathbb{R}$ a measurable function with $\mathbf{E}\{|f(X,Y)|\}<\infty$. Further, let $\mathcal{G}_1\subseteq\mathcal{G}_2$ be σ-algebras, such that

- X is adapted to $\mathcal{G}_1$ (and thus to $\mathcal{G}_2$);
- Y is independent of $\mathcal{G}_2$ (and thus of $\mathcal{G}_1$);

Show that $\mathbf{E}\{f(X,Y)|\mathcal{G}_2\} = \mathbf{E}\{f(X,Y)|\mathcal{G}_1\}$.

Theorem 3.5 *If X and Y are random variables with values in $\mathbb{R}^n$ and $\mathbb{R}$, respectively, and $\phi:\mathbb{R}^{n+1}\to\mathbb{R}$ is measurable with $\mathbf{E}\{|\phi(X,Y)|\}<\infty$, we have*

$$\mathbf{E}\{\phi(X,Y)\} = \int_{\mathbb{R}} \mathbf{E}\{\phi(X,Y)|Y=y\}\,dF_Y(y).$$

If in addition X and Y are independent, we further have

$$\mathbf{E}\{\phi(X,Y)|Y=y\} = \mathbf{E}\{\phi(X,y)\}.$$

EXERCISE 34 Prove Theorem 3.5.

3.2 Introduction to Markov Theory

Definition 3.6 *A stochastic process $X=\{X(t)\}_{t\geq 0}$, adapted to a filtration $\mathbb{F}=\{\mathcal{F}_t\}_{t\geq 0}$, is a Markov process wrt. $\mathbb{F}$ (has the Markov property), if*

$$\mathbf{P}\{X(t)\in\cdot\,|\mathcal{F}_s\} = \mathbf{P}\{X(t)\in\cdot\,|X(s)\} \qquad \textit{for} \quad 0\leq s<t.$$

Definition 3.7 *Let X be a Markov process. A function $P:\mathcal{B}(\mathbb{R})\times(0,\infty)\times\mathbb{R}\times[0,\infty)\to[0,1]$ is a transition probability (for X), if*

$$\mathcal{B}(\mathbb{R})\times(0,\infty)\times\mathbb{R}\times[0,\infty)\ni(A,t,x,s)\to P(A,t,x,s) = \mathbf{P}\{X(t+s)\in A|X(s)=x\},$$

while $P:\mathbb{R}\times(0,\infty)\times\mathbb{R}\times[0,\infty)\to[0,1]$ is a transition distribution function, if

$$\mathbb{R}\times(0,\infty)\times\mathbb{R}\times[0,\infty)\ni(y,t,x,s)\to P(y,t,x,s) = \mathbf{P}\{X(t+s)\leq y|X(s)=x\}.$$

A function $p:\mathbb{R}\times(0,\infty)\times\mathbb{R}\times[0,\infty)\to[0,1]$ is a transition density function, if

$$\mathbf{P}\{X(t+s)\in\cdot|X(s)=x\} = \int_{y\in\cdot} p(y,t,x,s)\,dy \qquad \textit{for} \quad (t,x,s)\in(0,\infty)\times\mathbb{R}\times[0,\infty).$$

*Remark 3.8 Transition probabilities are not unique in general, because only $x\in X(\Omega,s)$ [the possible values of $X(s)$] affect the validity of $P(\cdot,t,x,s)=\mathbf{P}\{X(t+s)$

$\cdot\,|\,X(s)=x\}$, and $P(\cdot,t,x,s)$ may thus be chosen arbitrarily for $x\in\mathbb{R}\backslash X(\Omega,s)$.

If, for example, X is a Markov process started at $X(0)=0$, both

$$P^{(1)}(\cdot,t,x,0) = \mathbf{P}\{X(t)\in\cdot\} \qquad \text{and} \qquad P^{(2)}(\cdot,t,x,0) = \mathbf{P}\{X(t)\in\cdot - x\}$$

are transition probabilities, because, since $\sigma(X(0)) = \{\emptyset,\Omega\}$, we have (Exercise 10)

$$\mathbf{P}\{X(t)\in\cdot\,|\,X(0)\} = \mathbf{P}\{X(t)\in\cdot\} = P^{(1)}(\cdot,t,X(0),0) = P^{(2)}(\cdot,t,X(0),0). \quad \#$$

Theorem 3.9 *For a Markov process X with transition probability P, we have*

$$\mathbf{P}\{X(t+s)\in\cdot\} = \int_{\mathbb{R}} P(\cdot,t,x,s)\,dF_{X(s)}(x) \qquad \text{for} \quad (t,x,s)\in(0,\infty)\times\mathbb{R}\times[0,\infty).$$

Proof. By Definition 3.2 and Theorem 3.3 (together with Exercise 12), we have

$$\int_{\mathbb{R}} P(\cdot,t,x,s)\,dF_{X(s)}(x) = \mathbf{E}\Big\{P(\cdot,t,X(s),s)\Big\} = \mathbf{E}\Big\{\mathbf{P}\{X(t+s)\in\cdot\,|\,X(s)\}\Big\}$$
$$= \mathbf{P}\{X(t+s)\in\cdot\}. \quad \square$$

Definition 3.10 *A transition probability $P(\cdot,\cdot,\cdot,s)$ for a Markov process is time homogeneous if it does not depend on the last argument $s\in[0,\infty)$. A Markov process is time homogeneous if it has a time homogeneous transition probability.*

By tradition, much Markov theory is done only under time homogeneity. In the setting of *Stochastic Calculus* and *diffusion theory*, we have found little reason for this.

Definition 3.11 *A stochastic process $\{X(t)\}_{t\geq 0}$ is a Markov process wrt. itself, if it is a Markov process wrt. $\sigma(X(s): 0\leq s\leq t)\}_{t\geq 0}$.*

It is less common to deal with the Markov property wrt. other filtrations than the cannonical one (itself ...), than it is to deal with smartingales wrt. such filtrations.

3.3 Markov Properties of Lévy Processes

Definition 3.12 *A stochastic process $\{X(t)\}_{t\geq 0}$, that has independent and stationary increments, is called a Lévy process.*

Theorem 3.13 *A Lévy process X is a time homogeneous Markov process wrt. itself, and has transition probability*

$$P(\cdot,t,x,s) = \mathbf{P}\{X(t)-X(0)\in\cdot - x\} \qquad \text{for} \quad (t,x,s)\in(0,\infty)\times\mathbb{R}\times[0,\infty).$$

Proof. Since $X(t)-X(s)$ is independent of $\mathcal{F}_s = \sigma(X(r): 0 \le r \le s)$ for $0 \le s < t$, and $X(s)$ is adpated to $\sigma(X(s)) \subseteq \mathcal{F}_s$, Exercise 33 gives the Markov property

$$\begin{aligned}\mathbf{P}\{X(t+s) \in \cdot | \mathcal{F}_s\} &= \mathbf{P}\{X(t+s)-X(s) \in \cdot - X(s) | \mathcal{F}_s\} \\ &= \mathbf{E}\Big\{\mathbf{P}\{X(t+s)-X(s) \in \cdot - X(s) | X(s)\}\Big\} = \mathbf{P}\{X(t+s) \in \cdot | X(s)\}.\end{aligned}$$

By stationarity and independence of increments, Theorem 3.5 further gives

$$\begin{aligned}\mathbf{P}\{X(t+s) \in \cdot | X(s)=x\} &= \mathbf{P}\Big\{X(t+s)-X(s) \in \cdot - X(s) \Big| X(s)=x\Big\} \\ &= \mathbf{P}\{X(t+s)-X(s) \in \cdot - x\} \\ &= \mathbf{P}\{X(t)-X(0) \in \cdot - x\}. \quad \square\end{aligned}$$

Corollary 3.14 (Markov Property of BM) *BM is a time homogeneous Markov process wrt. itself, and has transition probability*

$$P(\cdot, t, x, s) = \mathbf{P}\{\mathrm{N}(0,t) \in \cdot - x\} \qquad \text{for} \quad (t,x,s) \in (0,\infty) \times \mathbb{R} \times [0,\infty),$$

and transition density (recall Definition 2.4)

$$p(y,t,x,s) = p_t(x,y) = \frac{1}{\sqrt{2\pi t}}\, \mathrm{e}^{-\frac{(y-x)^2}{2t}} \qquad \text{for} \quad (y,t,x,s) \in \mathbb{R} \times (0,\infty) \times \mathbb{R} \times [0,\infty).$$

Proof. Theorem 3.13 gives everything except the claim about the transition density. That in turn, is an elementary consequence of the form of $P(\cdot,\cdot,\cdot,\cdot)$. $\square$

***Remark 3.15** The reason for the (deliberate) unspecificness about the starting point $B(0)$ of BM, is that the relation (by time homogeneity)

$$\mathbf{P}\{B^0(t+s) \in A | B^0(s)=x\} = P(A,t,x,s) = P(A,t,x,0) = \mathbf{P}\{B^0(t) \in A | B^0(0)=x\}$$

does not "look right" for $x \neq 0$, since $B^0(0)=0$. (However, there is really nothing wrong with this, and we have explained what is going on here in Remark 3.8.) #

Exercise 35 Show how Exercise 30 gives the Markov property of BM. [You may need *regular conditional probabilities* (see Lemma 21.2 below) to make this rigorous.]

3.4 Self-Similarity and Reflection Principle

Definition 3.16 *A stochastic process* $\{X(t)\}_{t\ge 0}$ *is* <u>*self-similar with index* $\kappa > 0$</u> *if*

$$\{X(\lambda t)\}_{t\ge 0} =_{\text{same fidi's}} \{\lambda^\kappa X(t)\}_{t \ge 0} \qquad \text{for each choice of} \quad \lambda > 0.$$

Exercise 36 Show that BM B^0 is self-similar with index $1/2$.

*Theorem 3.17 (Reflection Principle) *For a continuous stochastic process $\{X(t)\}_{t\geq 0}$, with independent symmetrically distributed increments and $X(0)=0$,*

$$\mathbf{P}\Big\{\sup_{s\in[0,t]} X(s)>x\Big\} = 2\,\mathbf{P}\{X(t)>x\} \qquad \textit{for} \quad x>0 \quad \textit{and} \quad t>0.$$

**Proof.* The following argument is inspired by [11, p. 106]: By independence and symmetry of increments, together with continuity, we get (by "reflection"), as $n\to\infty$,

$$\begin{aligned}
&\mathbf{P}\{\sup_{s\in[0,t]} X(s)>x\} - \mathbf{P}\{X(t)>x\}\\
&\leftarrow \mathbf{P}\Big\{\bigcup_{k=1}^{2^n}\{X(2^{-n}kt)>x\}\Big\} - \mathbf{P}\{X(t)>x\}\\
&= \sum_{k=1}^{2^n-1}\mathbf{P}\Big\{\bigcap_{\ell=1}^{k-1}\{X(2^{-n}\ell t)\leq x\},\ X(2^{-n}kt)>x,\ X(t)\leq x\Big\}\\
&= \sum_{k=1}^{2^n-1}\mathbf{P}\Big\{\bigcap_{\ell=1}^{k-1}\{X(2^{-n}\ell t)\leq x\},\ X(2^{-n}kt)>x,\ X(t)-X(2^{-n}kt)\leq x-X(2^{-n}kt)\Big\}\\
&=_{\text{(reflection)}} \sum_{k=1}^{2^n-1}\mathbf{P}\Big\{\bigcap_{\ell=1}^{k-1}\{X(2^{-n}\ell t)\leq x\},\ X(2^{-n}kt)>x,\ X(t)-X(2^{-n}kt)\geq X(2^{-n}kt)-x\Big\}\\
&= \sum_{k=1}^{2^n-1}\mathbf{P}\Big\{\bigcap_{\ell=1}^{k-1}\{X(2^{-n}\ell t)\leq x\},\ X(2^{-n}kt)>x,\ X(t)\geq 2X(2^{-n}kt)-x\Big\} \to \mathbf{P}\{X(t)>x\}. \quad \square
\end{aligned}$$

Corollary 3.18 (Reflection Principle for BM) *For BM B^0 we have*

$$\mathbf{P}\Big\{\sup_{s\in[0,t]} B^0(s)>x\Big\} = 2\,\mathbf{P}\{B^0(t)>x\} \qquad \textit{for} \quad x>0 \quad \textit{and} \quad t>0.$$

*Example 3.19 Let B_1 and B_2 be independent BM started at zero, and pick a $\hat{t}>0$. The processes $X_1(t) = 2\,B_1(t)$, and (more interesting) $X_2(t) = B_1(t\wedge\hat{t}) + 2\,(B_2(t\vee\hat{t})-B_2(\hat{t}))$ satisfy the hypothesis of Theorem 3.17, and are not BM. #

*Exercise 37 Writing $g(t,x)\equiv\mathbf{P}\{\sup_{s\in[0,t]} B^0(s)>x\}$ for $x,t>0$, we have

$$g(t+h,x) = \mathbf{P}\Big\{\sup_{s\in[0,h)}B^0(s)>x, \sup_{s\in[h,t+h]}B^0(s)\leq x\Big\} + \mathbf{P}\{\sup_{s\in[h,t+h]}B^0(s)>x\}.$$

Use Theorem 7.2 below, to show that the first term on the right-hand side is $o(h)$ as $h\downarrow 0$. Use this in turn, to obtain, by integration by parts,

$$\begin{aligned}
\frac{g(t+h,x)-g(t,x)}{h} &\sim \int_{\mathbb{R}} \frac{\mathbf{P}\{\sup_{s\in[h,t+h]} B^0(s)>x \mid B^0(h)=y\} - g(t,x)}{h}\, dF_{B^0(h)}(y)\\
&= \int_{\mathbb{R}} -\frac{\partial_2 g(t,x-y) + \frac{g(t,x-y)-g(t,x)}{y}}{y}\, f_{B^0(h)}(y)\,dy \qquad \text{as} \quad h\downarrow 0.
\end{aligned}$$

Deduce that $\partial_t\, g(t,x) = \frac{1}{2}(\partial_x)^2 g(t,x)$. Show that also $\tilde{g}(t,x)\equiv\mathbf{P}\{|B^0(t)|>x\}$ satisfies this PDE. What about $g(0^+,x)$ and $\tilde{g}(0^+,x)$? Any relation to Corollary 3.18?

*Remark 3.20 By e.g., [10] and [28, Section V.1], first passage times of solutions to diffusion type SDE, which includes BM, satisfy PDE closely related to the PDE for their transition densities (see Lecture 22 on the latter). This ensures the existence of the derivatives featuring in Exercise 37, as well as the general validity of such an approach to first passage problems for diffusion processes. #

*Remark 3.21 There are extensions of the *Reflection Principle* to more general processes X than those in Theorem 3.17, if it is only required to hold asymptotically

$$\mathbf{P}\Big\{\sup_{s\in[0,t]} X(s)>x\Big\} \sim \text{constant} \times \mathbf{P}\{X(t)>x\} \qquad \text{as} \quad x\to\infty.$$

See [2] on such results for quite general diffusions (solutions to SDE), and [3, Section 1] on an overview for (non-continuous and/or non-symmetric) Lévy processes. #

3.5 Zero Crossings of BM

Example 3.22 From the *Reflection Principle* we see that BM B^0 changes sign in the interval $[0,t]$, with probability one, regardless of how small $t>0$ is

$$\mathbf{P}\Big\{B^0(s)\leq 0 \text{ for } s\in[0,t]\Big\} = 1-\mathbf{P}\Big\{\sup_{s\in[0,t]} B^0(s)>0\Big\} = 1-2\,\mathbf{P}\{B^0(t)>0\} = 0. \ \#$$

Definition 3.23 *For BM B^0 we define the hitting time*

$$T_x = \inf\{t\geq 0 : B^0(t)=x\} \qquad \textit{for} \quad x\neq 0.$$

Corollary 3.24 *The hitting time T_x, $x\neq 0$, for BM B^0 has distribution function*

$$F_{T_x}(t) = \mathbf{P}\{T_x\leq t\} = 2\,\mathbf{P}\{B^0(t)>|x|\} \qquad \textit{for} \quad t>0,$$

and probability density function

$$f_{T_x}(t) = \frac{d}{dt}\mathbf{P}\{T_x\leq t\} = \frac{|x|}{\sqrt{2\pi}\,t^{3/2}}\exp\Big\{-\frac{x^2}{2t}\Big\} \qquad \textit{for} \quad t>0.$$

Proof. By symmetry together with the *Reflection Principle*, we have

$$\mathbf{P}\{T_x\leq t\} = \mathbf{P}\Big\{\sup_{s\in[0,t]} B^0(s)\geq |x|\Big\} \in \Big[2\,\mathbf{P}\{B^0(t)>|x|\},\, 2\,\mathbf{P}\{B^0(t)>|x|-\varepsilon\}\Big]$$

for each $\varepsilon>0$. Sending $\varepsilon\downarrow 0$, we thus get

$$2\,\mathbf{P}\{B^0(t)>|x|\} \leq \mathbf{P}\{T_x\leq t\} \leq 2\,\mathbf{P}\{B^0(t)\geq |x|\} = 2\,\mathbf{P}\{B^0(t)>|x|\}.$$

To get the density function we differetiate, integrate by parts, and identify

$$
\begin{aligned}
f_{T_x}(t) &= \frac{d}{dt}\, 2\,\mathbf{P}\{B^0(t) > |x|\} \\
&= \frac{d}{dt}\int_{|x|}^{\infty} \frac{2}{\sqrt{2\pi t}}\, \exp\Big\{-\frac{y^2}{2t}\Big\}\, dy \\
&= \int_{|x|}^{\infty} \frac{1}{\sqrt{2\pi t}}\left(\frac{y^2}{t^2}-\frac{1}{t}\right) \exp\Big\{-\frac{y^2}{2t}\Big\}\, dy \\
&= \left[\frac{1}{\sqrt{2\pi t}}\,\frac{-y}{t}\, \exp\Big\{-\frac{y^2}{2t}\Big\}\right]_{|x|}^{\infty} + \int_{|x|}^{\infty} \frac{1}{\sqrt{2\pi t}\,t}\, \exp\Big\{-\frac{y^2}{2t}\Big\}\, dy - \frac{\mathbf{P}\{T_x \le t\}}{2t} \\
&= \frac{|x|}{\sqrt{2\pi}\, t^{3/2}}\, \exp\Big\{-\frac{x^2}{2t}\Big\}. \quad \square
\end{aligned}
$$

Since $2\,\mathbf{P}\{B^0(t) > |x|\} \to 1$ as $t \to \infty$, the event $\{T_x < \infty\}$ has probability one, making T_x a well-defined finite random variable (if we change its value to zero say, on the event of probability zero where it is infinite). Hitting times for stochastic processes in general need not be finite a.s. In order to be able to treat them as random variables anyway, one is sometimes led to consider random variables with values in the <u>extended real line $\overline{\mathbb{R}}$</u> $\equiv \mathbb{R} \cup \{-\infty, \infty\}$ (with the obvious σ-algebra).

Corollary 3.25 *For BM B^0 we have*

$$\mathbf{P}\Big\{B^0(s) = 0 \text{ for some } s \in [a,b]\Big\} = \frac{2}{\pi}\arccos\!\Big(\sqrt{a/b}\Big) \qquad \textit{for} \quad 0 < a < b.$$

Proof. The process $\{B^0(s) - B^0(a)\}_{s\in[a,b]}$ is independent of $B^0(a)$, since $\mathbf{Cov}\{B^0(s) - B^0(a), B^0(a)\} = 0$ (recall Corollary 2.11). Hence Theorem 3.5 together with stationarity of increments and Corollary 3.24 show that

$$
\begin{aligned}
&\mathbf{P}\{B^0(s) = 0 \text{ for some } s \in [a,b]\} \\
&= \mathbf{E}\Big\{ I_{\{0 \in [\inf_{s\in[a,b]} B^0(s),\, \sup_{s\in[a,b]} B^0(s)]\}} \Big\} \\
&= \mathbf{E}\Big\{ I_{\{-B^0(a) \in [\inf_{s\in[a,b]} B^0(s)-B^0(a),\, \sup_{s\in[a,b]} B^0(s)-B^0(a)]\}} \Big\} \\
&= \int_{\mathbb{R}} \mathbf{E}\Big\{ I_{\{-B^0(a) \in [\inf_{s\in[a,b]} B^0(s)-B^0(a),\, \sup_{s\in[a,b]} B^0(s)-B^0(a)]\}} \,\Big|\, B^0(a) = y \Big\}\, dF_{B^0(a)}(y) \\
&= \int_{\mathbb{R}} \mathbf{E}\Big\{ I_{\{-y \in [\inf_{s\in[a,b]} B^0(s)-B^0(a),\, \sup_{s\in[a,b]} B^0(s)-B^0(a)]\}} \Big\}\, dF_{B^0(a)}(y) \\
&= \int_{\mathbb{R}} \mathbf{P}\Big\{ B^0(s) - B^0(a) = -y \text{ for some } s \in [a,b] \Big\}\, dF_{B^0(a)}(y) \\
&= \int_{\mathbb{R}} \mathbf{P}\Big\{ B^0(s-a) - B^0(0) = -y \text{ for some } s \in [a,b] \Big\}\, dF_{B^0(a)}(y) \\
&= \int_{\mathbb{R}} \mathbf{P}\Big\{ B^0(s) = -y \text{ for some } s \in [0,b-a] \Big\}\, dF_{B^0(a)}(y) \\
&= \int_{\mathbb{R}} \mathbf{P}\{T_{-y} \le b-a\}\, dF_{B^0(a)}(y) \\
&= \int_{\mathbb{R}} \left(\int_0^{b-a} \frac{|y|}{\sqrt{2\pi}\, t^{3/2}}\, \exp\Big\{-\frac{y^2}{2t}\Big\}\, dt \right) \frac{1}{\sqrt{2\pi a}}\, \exp\Big\{-\frac{y^2}{2a}\Big\}\, dy = \frac{2}{\pi}\arctan\!\Big(\sqrt{(b-a)/b}\Big),
\end{aligned}
$$

which by elementary trigonometry is the desired arccos-law. Here we calculated the right-hand side with the help of `Mathematica`:

```
In[1]:= Simplify[Integrate[2*y*Exp[-y^2*(1/t+1/a)/2]/(2*Pi*Sqrt[t^3*a]), {y, 0, Infinity}]]
```

$$\text{Out[1]=}\ \frac{\mathrm{If}\left[\mathrm{Re}\left[\frac{1}{a}+\frac{1}{t}\right]>0,\ \frac{a\,t}{a+t},\ \int_0^\infty e^{-\frac{1}{2}\left(\frac{1}{a}+\frac{1}{t}\right)y^2}\,y\,dy\right]}{\pi\sqrt{a\,t^3}}$$

$$\text{In[2]:=}\ \mathrm{Simplify}\left[\mathrm{Integrate}\left[\frac{\frac{a\,t}{a+t}}{\pi\sqrt{a\,t^3}},\ \{t,\,0,\,b-a\}\right]\right]$$

$$\text{Out[2]=}\ \frac{a\,\mathrm{If}\left[a<b\ \&\&\ !\left(\mathrm{Im}\left[\frac{a}{a-b}\right]==0\ \&\&\ \frac{a}{a-b}>0\ \&\&\ \frac{a}{a-b}<1\right),\ \frac{2\sqrt{\frac{1}{a}}\,\mathrm{ArcTan}\left[\sqrt{\frac{1}{a}}\sqrt{-a+b}\right]}{\sqrt{a}},\ \int_0^{-a+b}\frac{t}{\sqrt{a\,t^3}\,(a+t)}\,dt\right]}{\pi}$$

Corollary 3.26 (ArcSine Law) *For BM B^0 we have*

$$\mathbf{P}\left\{B^0(s)\neq 0 \ \text{ for all }\ s\in[a,b]\right\} = \frac{2}{\pi}\arcsin\left(\sqrt{a/b}\right) \qquad \textit{for} \quad 0<a<b.$$

Exercise 38 Derive the *Arcsine Law* from Corollary 3.25.

Exercise 39 In the proof of Corollary 3.25, for the first time, we used stationarity of increments in the sense of Definition 2.16, rather than in the seemingly weeker univariate sense, imposed in Defintion 2.1 of BM: Show that a stochastic process, that has independent increments whose univariate distributions are stationary, also has stationary increments in the sense of Definition 2.16.

4 Fourth Lecture

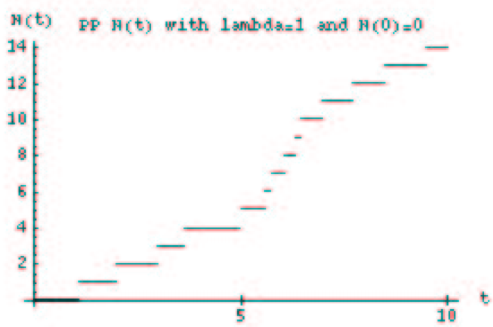
4.1 The Poisson Process

Definition 4.1 *A stochastic process* $\{N(t)\}_{t\geq 0}$ *is a* *Poisson process (PP) with intensity* $\lambda > 0$, *if it takes values in* $\mathrm{N} = \{0, 1, \ldots\}$, *with the following properties:*

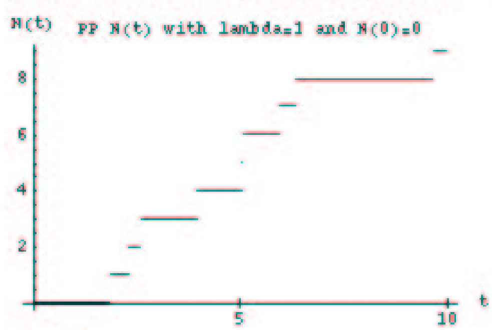
- (Right-Continuity) $[0,\infty) \ni t \to N(\omega,t) \in \mathbb{R}$ *is right-continuous for all (almost all)* $\omega \in \Omega$;
- (Independent Increments) $N(t) - N(s)$ *is independent of* $\{N(r)\}_{r\in[0,s]}$ *for* $0 \leq s < t$;
- (Stationary Poisson Increments) $N(t) - N(s)$ *is* $\mathrm{Po}(\lambda(t-s))$*-distributed for* $0 \leq s \leq t$.

A random variable ξ is $\mathrm{Po}(\lambda)$-distributed if $\mathbf{P}\{\xi = k\} = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$ for $k \in \mathbb{N}$.

```
In[1]:= << Statistics`DiscreteDistributions`;

In[2]:= Incr = N[Table[Random[PoissonDistribution[0.01]], {1000}]];

In[3]:= PP = {Incr[[1]]}; For[k = 2, k <= 1000, k++, PP = Join[PP, {PP[[k - 1]] + Incr[[k]]}]];

In[4]:= ListPlot[PP, Ticks -> {{{500, "5", 0.02}, {1000, "10", 0.02}}, Automatic},
         AxesLabel -> {"t", "N(t)"}, PlotLabel -> "  PP N(t) with lambda=1 and N(0)=0"]
```




```
Out[4]= - Graphics -

In[5]:= Incr = N[Table[Random[PoissonDistribution[0.01]], {1000}]];

In[6]:= PP = {Incr[[1]]}; For[k = 2, k <= 1000, k++, PP = Join[PP, {PP[[k - 1]] + Incr[[k]]}]];

In[7]:= ListPlot[PP, Ticks -> {{{500, "5", 0.02}, {1000, "10", 0.02}}, Automatic},
         AxesLabel -> {"t", "N(t)"}, PlotLabel -> "  PP N(t) with lambda=1 and N(0)=0"]
```




```
Out[7]= - Graphics -
```

Remark 4.2 As for BM, the `fidi's` of PP are determined when, in addition to the above requirements, the value for $N(0)$ is specified [e.g., $N(0)=0$].
As for BM, the literature is sloppy in the use of the definition of PP, so that it is often not specified whether results require something about the value of $N(0)$ or not. #

Definition 4.3 *N^x denotes PP with $N(0)=N^x(0)=x$, for a constant $x\in\mathbb{N}$.*

See Appendix B on the following description of PP, which also ensures its existence:

Theorem 4.4 *Let $\xi_1,\xi_2,\ldots$ be independent $\exp(\lambda)$-distributed random variables. The process*

$$N^x(t)\equiv x+\sup\Big\{n\in\mathbb{N}:\textstyle\sum_{i=1}^n\xi_i\le t\Big\}\qquad\text{for}\quad t\ge 0,$$

is a PP with intensity λ started at $N^x(0)=x$, for a constant $x\in\mathbb{N}$. Notice that

$$N^x(t)\ge n+x\quad\Leftrightarrow\quad \textstyle\sum_{i=1}^n\xi_i\le t,$$

so that times between "jumps" for N^x are independent $\exp(\lambda)$-distributed.

A random variable ξ is $\exp(\lambda)$-distributed if $\mathbf{P}\{\xi\le x\}=\int_0^x\lambda\,\mathrm{e}^{-\lambda y}\,dy$ for $x\ge 0$.

*EXERCISE 40 Prove Theorem 4.4. (This exercise is solved in Appendix B.)

4.2 Martingale Properties of Lévy Processes

Theorem 4.5 *For a stochastic process $\{X(t)\}_{t\ge 0}$, that has stationary increments, with means that are locally bounded, we have*

$$\mathbf{E}\{X(t)-X(s)\}=\mathbf{E}\{X(1)-X(0)\}\,(t-s)\qquad\text{for}\quad s,t\ge 0.$$

A function $f:T\to\mathbb{R}$, $T\subseteq\mathbb{R}^n$, is locally bounded if $\sup_{t\in B}|f(t)|<\infty$ for each bounded set $B\subseteq T$.

Proof. Writing $m(t)\equiv\mathbf{E}\{X(t+r)-X(r)\}$ for $r,t\ge 0$, we have

$$\begin{aligned}m(t+s)&=\mathbf{E}\{X(t+s+r)-X(r)\}\\&=\mathbf{E}\{X(t+s+r)-X(s+r)\}+\mathbf{E}\{X(s+r)-X(r)\}=m(t)+m(s).\end{aligned}$$

Since m is locally bounded, this *Cauchy functional equation* can only have solutions of the form $m(t)=Kt$ for $t\ge 0$, for some constant $K\in\mathbb{R}$ (e.g., [6, pp. 4-5])*. Hence we have $\mathbf{E}\{X(t)-X(s)\}=m(t-s)=K\,(t-s)$ for $0\le s\le t$, which by a trivial argument extends to $s,t\geq 0$. Taking $t=1$ and $s=0$, we get $K=\mathbf{E}\{X(1)-X(0)\}$. □

EXERCISE 41 Show that a stochastic process $\{X(t)\}_{t\geq 0}$, that has independent increments and constant mean, is a martingale wrt. itself. Conclude that, for a Lévy process $\{X(t)\}_{t\geq 0}$, with locally bounded means, $\big\{X(t)-\mathbf{E}\{X(1)-X(0)\}t\big\}_{t\geq 0}$ is a martingale wrt. $\{\sigma(X(s):0\leq s\leq t)\}_{t\geq 0}$.

Theorem 4.6 *A martingale $\{X(t)\}_{t\geq 0}$ has uncorrelated increments (provided that their covariances are well-defined and finite), the means of which must be zero.*

Proof. We have $\mathbf{E}\{X(t)\}=\mathbf{E}\{X(s)\}$ for $0\leq s<t$, since

$$\mathbf{E}\{X(t)-X(s)\} = \mathbf{E}\Big\{\mathbf{E}\{X(t)-X(s)\,|\,\mathcal{F}_s\}\Big\} = \mathbf{E}\Big\{\mathbf{E}\{X(t)\,|\,\mathcal{F}_s\}-X(s)\Big\} = \mathbf{E}\{X(s)-X(s)\}.$$

If $\mathbf{E}\{|X(t)-X(s)|\,|X(s)-X(r)|\}<\infty$ for $0\leq r<s<t$, we similarly get

$$\begin{aligned}\mathbf{E}\Big\{(X(t)-X(s))\,(X(s)-X(r))\Big\} &= \mathbf{E}\Big\{\mathbf{E}\Big\{(X(t)-X(s))\,(X(s)-X(r))\,\Big|\,\mathcal{F}_s\Big\}\Big\}\\ &= \mathbf{E}\Big\{\mathbf{E}\{X(t)-X(s)\,|\,\mathcal{F}_s\}\,(X(s)-X(r))\Big\}\\ &= \mathbf{E}\Big\{(X(s)-X(s))\,(X(s)-X(r))\Big\} = 0. \quad \square\end{aligned}$$

Theorem 4.7 *For a Lévy process $\{X(t)\}_{t\geq 0}$, that has increments with finite variances, we have*

$$\mathbf{Var}\{X(t)-X(s)\} = \mathbf{Var}\{X(1)-X(0)\}\,|t-s| \qquad \textit{for} \quad s,t\geq 0.$$

Proof. Writing $V(t)\equiv\mathbf{Var}\{X(t+r)-X(r)\}$ for $r,t\geq 0$, we have

$$\begin{aligned}V(t+s) &= \mathbf{Var}\{X(t+s+r)-X(r)\}\\ &= \mathbf{Var}\{X(t+s+r)-X(s+r)\}+\mathbf{Var}\{X(s+r)-X(r)\} = V(t)+V(s).\end{aligned}$$

Here V is locally bounded, since finite and increasing. This gives (cf. the proof of Theorem 4.5) $V(t)=Kt$ for $t\geq 0$, for some constant K, so that $\mathbf{Var}\{X(t)-X(s)\}$ $=K\,|t-s|=\mathbf{Var}\{X(1)-X(0)\}\,|t-s|$ for $s,t\geq 0$, by symmetry. □

Theorem 4.8 *For a Lévy process $\{X(t)\}_{t\geq 0}$, that has increments with locally bounded means and finite variances, $\Big\{\big(X(t)-\mathbf{E}\{X(1)-X(0)\}t\big)^2-\mathbf{Var}\{X(1)-X(0)\}t\Big\}_{t\geq 0}$ is a martingale wrt. $\{\sigma(X(s):0\leq s\leq t)\}_{t\geq 0}$.*

Proof. We claim that $Y(t)^2-\mathbf{Var}\{X(1)-X(0)\}t=Y(t)^2-\mathbf{Var}\{Y(1)-Y(0)\}t$ is a martingale, where $Y(t)\equiv X(t)-\mathbf{E}\{X(1)-X(0)\}t$. Here adaptedness is trivial, and means finite by assumption. Notice that Y has zero means (Theorem 4.5), and that

$$\mathbf{Var}\{Y(1)-Y(0)\}t = \mathbf{Var}\{Y(t)-Y(s)\} + \mathbf{Var}\{Y(1)-Y(0)\}s \qquad \text{for} \quad 0 \le s \le t$$

(Theorem 4.7). Since $Y(t)-Y(s)$ is independent of $\mathcal{F}_s = \sigma(X(r): 0 \le r \le s)$ [and $Y(s)$ adapted to $\mathcal{F}_s$], it follows that, for $0 \le s < t$,

$$\begin{aligned}
\mathbf{E}\Big\{Y(t)^2 - \mathbf{Var}\{Y(1)-Y(0)\}t \,\Big|\, \mathcal{F}_s\Big\} &= \mathbf{E}\Big\{(Y(t)-Y(s))^2 - \mathbf{Var}\{Y(t)-Y(s)\} \,\Big|\, \mathcal{F}_s\Big\} \\
&\quad + \mathbf{E}\Big\{2\,Y(s)\,(Y(t)-Y(s)) \,\Big|\, \mathcal{F}_s\Big\} \\
&\quad + \mathbf{E}\Big\{Y(s)^2 - \mathbf{Var}\{Y(1)-Y(0)\}s \,\Big|\, \mathcal{F}_s\Big\} \\
&= \mathbf{E}\{(Y(t)-Y(s))^2\} - \mathbf{Var}\{Y(t)-Y(s)\} \\
&\quad + 2\,Y(s)\,\mathbf{E}\{Y(t)-Y(s)\} \\
&\quad + Y(s)^2 - \mathbf{Var}\{Y(1)-Y(0)\}s \\
&= 0 + 0 + Y(s)^2 - \mathbf{Var}\{Y(1)-Y(0)\}s. \qquad \square
\end{aligned}$$

For a process $\{X(t)\}_{t \ge 0}$ we have established the following implications:

$$\left\{\begin{array}{ccc}
\boxed{\text{independent zero-mean increments}} & \Rightarrow & \boxed{\text{martingale wrt. itself}} \\
\boxed{\text{martingale with finite second moments}} & \Rightarrow & \boxed{\text{uncorrelated zero-mean increments}}
\end{array}\right. .$$

4.3 Markov and Martingale Properties of PP

Corollary 4.9 (Markov Property of PP) *A PP N with intensity λ is a time homogeneous Markov process wrt. itself, and has transition probability*

$$P(\cdot, t, x, s) = \mathbf{P}\{\mathrm{Po}(\lambda t) \in \cdot - x\} \qquad \textit{for} \quad (t, x, s) \in [0, \infty) \times \mathbb{R} \times [0, \infty).$$

Proof. This follows from Theorem 3.13. $\square$

Corollary 4.10 (Martingale Property of PP) *For a PP N with intensity λ, $\{N(t) - \lambda t\}_{t \ge 0}$ is a martingale wrt. $\{\sigma(N(s): 0 \le s \le t)\}_{t \ge 0}$.*

Proof. This follows from Exercise 41, since $\mathbf{E}\{N(1)-N(0)\} = \lambda$. $\square$

Corollary 4.11 *For a PP N with intensity λ, $\{(N(t)-\lambda t)^2 - \lambda t\}_{t \ge 0}$ is a martingale wrt. $\{\sigma(N(s): 0 \le s \le t)\}_{t \ge 0}$.*

Proof. This follows from Theorem 4.8, since $\mathbf{E}\{N(1)-N(0)\} = \mathbf{Var}\{N(1)-N(0)\} = \lambda$. $\square$

4.4 Stopping Times

Definition 4.12 *A $[0,\infty]$-valued random variable T is a* <u>stopping time</u> (see Lecture 23 for comments on variation in terminology) *wrt. a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, if*

$$\{T\leq t\} = \{\omega\in\Omega : T(\omega)\leq t\} \in \mathcal{F}_t \qquad \textit{for} \quad t\in[0,\infty).$$

In the case when $\mathcal{F}_t = \sigma(X(s): 0\leq s\leq t)$, for some stochastic process $\{X(t)\}_{t\geq 0}$, a stopping time is a random variable such that it is possible to determine whether $T\leq t$ or not, by means of the values of $\{X(s)\}_{s\in[0,t]}$. This is so, since an event in $\mathcal{F}_t$ can be "constructed" through a countable number of set operations (unions, intersections, ...) on events of type $\{X(s)\in A\}$, where $A\in\mathcal{B}(\mathbb{R})$ and $s\in[0,t]$.

Example 4.13 The hitting time T_x for BM B^0 is a stopping time wrt. the filtration $\mathcal{F}_t = \sigma(B^0(s): 0\leq s\leq t)$, since (by continuity of sample paths)

$$\{T_x\leq t\} = \Big\{\sup_{s\in[0,t]} B^0(s)\geq x\Big\} = \bigcap_{n=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcup_{k=0}^{2^m}\{B^0(2^{-m}kt) > x-1/n\} \in \mathcal{F}_t. \quad \#$$

Since hitting times may be infinite, $T=\infty$ is allowed in Definition 4.12.

Definition 4.14 *For a stopping time T wrt. a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, we define*

$$\underline{\mathcal{F}_T} \equiv \Big\{\Lambda\in\mathcal{F} : \Lambda\cap\{T\leq t\}\in\mathcal{F}_t \text{ for each } t\in[0,\infty)\Big\}.$$

Theorem 4.15 *For a stopping time T wrt. a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, we have that*

- *$T+t_0$ is a stopping time for every constant $t_0\geq 0$;*
- *$\mathcal{F}_T\subseteq\mathcal{F}_{T+t_0}$ for every constant $t_0\geq 0$;*
- *$\mathcal{F}_T$ is a σ-algebra.*

Exercise 42 Prove Theorem 4.15.

Definition 4.16 *A stochastic process $\{X(t)\}_{t\geq 0}$ [$\{X(t)\}_{t\in[0,T]}$] is* <u>measurable</u> *if the map $X:\Omega\times[0,\infty)\to\mathbb{R}$ [$X:\Omega\times[0,T]\to\mathbb{R}$] is $\mathcal{F}\times\mathcal{B}([0,\infty))$-measurable [$\mathcal{F}\times\mathcal{B}([0,T])$-measurable].*

Exercise 43 A right-continuous [left-continuous] process is measurable.

Theorem 4.17 *For a measurable stochastic process $\{X(t)\}_{t\geq 0}$, and a finite stopping time (or other positive random variable) T, $X(T)$ is a random variable.*

Proof. Recall that the composition of measurable functions is measurable. Notice that $X(T(\omega)) = X(\omega, T(\omega)) = (X \circ h)(\omega)$. Here the map

$$\Omega \ni \omega \to h(\omega) = (\omega, T(\omega)) \in \Omega \times [0,\infty) \qquad \text{is } \mathcal{F}\text{-measurable},$$

because

$$\{\omega : h(\omega) \in A \times B\} = A \cap \{\omega : T(\omega) \in B\} \in \mathcal{F} \qquad \text{for} \quad A \in \mathcal{F} \quad \text{and} \quad B \in \mathcal{B}([0,\infty)),$$

and $\{C \in \mathcal{F} \times \mathcal{B}([0,\infty)) : h^{-1}(C) \in \mathcal{F}\}$ is a σ-algebra, that thus is $\mathcal{F} \times \mathcal{B}([0,\infty))$. □

4.5 Strong Markov Property and Feller Processes

We write $\underline{\mathbb{C}_B(D)} \equiv \{(f : D \to \mathbb{R}) : f \text{ is bounded and continuous}\}$ for $D \subseteq \mathbb{R}^n$.

Definition 4.18 *A measurable and adapted stochastic* $\{X(t), \mathcal{F}_t\}_{t \geq 0}$ *is a* *<u>strong Markov process</u>* *(has the* *<u>strong Markov property</u>*) *if, for each finite stopping time* T,

$$\mathbf{E}\{f(X(t+T)) | \mathcal{F}_T\} = \mathbf{E}\{f(X(t+T)) | X(T)\} \qquad \textit{for} \quad f \in \mathbb{C}_B(\mathbb{R}) \quad \textit{and} \quad t > 0.$$

Definition 4.19 *A Markov process* $\{X(t)\}_{t \geq 0}$ *is a* *<u>Feller process</u>* *if has a transition probability* $P(\cdot,\cdot,\cdot,\cdot)$, *such that*

(A) $\displaystyle\lim_{t \downarrow 0} \int_{\mathbb{R}} f(\cdot)\, dP(\cdot, t, x, s) = f(x)$ *for* $f \in \mathbb{C}_B(\mathbb{R})$ *and* $s \geq 0$;

(B) $\displaystyle g(x,s) = \int_{\mathbb{R}} f(\cdot)\, dP(\cdot, t, x, s) \in \mathbb{C}_B(\mathbb{R} \times [0,\infty))$ *for* $f \in \mathbb{C}_B(\mathbb{R})$ *and* $t > 0$.

Exercise 44 For a Markov process X and a measurable $f : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{E}\{f(X(t)) | \mathcal{F}_s\} = \mathbf{E}\{f(X(t)) | X(s)\} \qquad \text{when} \quad \mathbf{E}\{|f(X)|\} < \infty \quad \text{and} \quad 0 \leq s < t.$$

***Exercise 45** BM together with the transition probability from Corollary 3.14 is a Feller process.

***Theorem 4.20** *Let* $\{X(t), \mathcal{F}_t\}_{t \geq 0}$ *be a right-continuous Markov process, with a transition probability* $P(\cdot,\cdot,\cdot,\cdot)$, *that satisfies condition* (B) *of a Feller process. For a finite stopping time* T *wrt.* $\{\mathcal{F}_t\}_{t \geq 0}$, *we have*

$$\mathbf{E}\{f(X(t+T)) | X(T)\} = \int_{\mathbb{R}} f(\cdot)\, dP(\cdot, t, X(T), T) \qquad \textit{for} \quad f \in \mathbb{C}_B(\mathbb{R}) \quad \textit{and} \quad t > 0.$$

**Proof.* Pick a $t > 0$ and a $B \in \sigma(X(T))$. Consider the classes

$$\mathcal{C} = \Big\{(X(T))^{-1}((-\infty, y)) : y \in \mathbb{R}\Big\} \quad \text{and} \quad \mathcal{D} = \Big\{B \in \sigma(X(T)) : B \cap \{T < t\} \in \mathcal{F}_t\Big\}.$$

Here $\mathcal{C}$ is closed under finite intersections. Further $\mathcal{D}$ is a Dynkin system. This is so, because $S\equiv\mathbb{R}\in\mathcal{D}$, since $\{T<t\}=\bigcup_{n=1}^{\infty}\{T\leq t-1/n\}\in\mathcal{F}_t$, the other two conditions for a Dynkin system being immediate. To show that $\mathcal{D}=\sigma(X(T))$, it is thus, by the *Dynkin System Lemma*, enough to show that $\mathcal{C}\subseteq\mathcal{D}$. Let $T_n=(\lfloor 2^nT\rfloor+1)/2^n$, where $\lfloor x\rfloor=k-1$ for $x\in(k-1,k]$. Since $T_n\downarrow T$ as $n\to\infty$, right-continuity gives

$$\begin{aligned}\{X(T)<y\}\cap\{T<t\} &= \bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{X(T_n)<y\}\cap\{T_n<t\}\\ &= \bigcup_{\{k:\,2^{-n}k<t\}}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{X(2^{-n}k)<y\}\cap\{T_n=2^{-n}k\}\\ &= \bigcup_{\{k:\,2^{-n}k<t\}}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{X(2^{-n}k)<y\}\cap\{T_n\in(2^{-n}(k-1),2^{-n}k]\},\end{aligned}$$

where the right-hande side belongs to $\mathcal{F}_t$. Hence we have $\mathcal{C}\subseteq\mathcal{D}$.

Now, let $T_n=([2^nT]+1)/2^n$ instead, where $[x]=k-1$ for $x\in[k-1,k)$. Take a $B\in\sigma(X(T))$, and notice that, by the investigation above,

$$B\cap\{T_n=2^{-n}k\}=B\cap\{T\in[2^{-n}(k-1),2^{-n}k)\}\in\mathcal{F}_{2^{-n}k}.$$

Since $T_n\downarrow T$ as $n\to\infty$, right-continuity and the *Dominated Convergence Theorem*, together property (B) of a Feller process and Exercise 44, give, as $n\to\infty$,

$$\begin{aligned}\int_B\int_{\mathbb{R}}f(\cdot)\,dP(\cdot,t,X(T),T)\,d\mathbf{P}\leftarrow &\int_B\int_{\mathbb{R}}f(\cdot)\,dP(\cdot,t,X(T_n),T_n)\,d\mathbf{P}\\ &=\sum_{k=0}^{\infty}\int_{B\cap\{T_n=2^{-n}k\}}\int_{\mathbb{R}}f(\cdot)\,dP(\cdot,t,X(2^{-n}k),2^{-n}k)\,d\mathbf{P}\\ &=\sum_{k=0}^{\infty}\int_{B\cap\{T_n=2^{-n}k\}}\mathbf{E}\{f(X(t+2^{-n}k))|X(2^{-n}k)\}\,d\mathbf{P}\\ &=\sum_{k=0}^{\infty}\int_{B\cap\{T_n=2^{-n}k\}}\mathbf{E}\{f(X(t+2^{-n}k))|\mathcal{F}_{2^{-n}k}\}\,d\mathbf{P}\\ &=\sum_{k=0}^{\infty}\int_{B\cap\{T_n=2^{-n}k\}}f(X(t+2^{-n}k))\,d\mathbf{P}\\ &=\int_B f(X(t+T_n))\,d\mathbf{P}\to\int_B f(X(t+T))\,d\mathbf{P}\quad\text{a.s.}\quad\square\end{aligned}$$

Theorem 4.21 *A right-continuous Markov process* $\{X(t),\mathcal{F}_t\}_{t\geq0}$, *with a transition probability* $P(\cdot,\cdot,\cdot,\cdot)$ *that satisfies condition* (B) *of a Feller process, is a strong Markov process.*

**Proof.* Let $T_n=(\lfloor 2^nT\rfloor+1)/2^n$, so that $T_n\downarrow T$ as $n\to\infty$, and

$$\Lambda\cap\{T_n=2^{-n}k\}=\Lambda\cap\{T\in(2^{-n}(k-1),2^{-n}k]\}\in\mathcal{F}_{2^{-n}k}\qquad\text{for}\quad\Lambda\in\mathcal{F}_T.$$

Using right-continuity and the *Dominated Convergence Theorem*, together property (B) of a Feller process and Exercise 44, Theorem 4.20 shows that

$$
\begin{aligned}
\int_\Lambda f(X(t+T))\,d\mathbf{P} &= \lim_{n\to\infty}\int_\Lambda f(X(t+T_n))\,d\mathbf{P}\\
&= \lim_{n\to\infty}\sum_{k=0}^{\infty}\int_{\Lambda\cap\{T_n=2^{-n}k\}} f(X(t+2^{-n}k))\,d\mathbf{P}\\
&= \lim_{n\to\infty}\sum_{k=0}^{\infty}\int_{\Lambda\cap\{T_n=2^{-n}k\}} \mathbf{E}\{f(X(t+2^{-n}k))|\mathcal{F}_{2^{-n}k}\}\,d\mathbf{P}\\
&= \lim_{n\to\infty}\sum_{k=0}^{\infty}\int_{\Lambda\cap\{T_n=2^{-n}k\}} \mathbf{E}\{f(X(t+2^{-n}k))|X(2^{-n}k)\}\,d\mathbf{P}\\
&= \lim_{n\to\infty}\sum_{k=0}^{\infty}\int_{\Lambda\cap\{T_n=2^{-n}k\}} \int_{\mathbb{R}} f(\cdot)\,dP(\cdot,t,X(2^{-n}k),2^{-n}k)\,d\mathbf{P}\\
&= \lim_{n\to\infty}\int_\Lambda\int_{\mathbb{R}} f(\cdot)\,dP(\cdot,t,X(T_n),T_n)\,d\mathbf{P}\\
&= \int_\Lambda\int_{\mathbb{R}} f(\cdot)\,dP(\cdot,t,X(T),T)\,d\mathbf{P}\\
&= \int_\Lambda \mathbf{E}\{f(X(t+T))|X(T)\}\,d\mathbf{P}. \quad \square
\end{aligned}
$$

Corollary 4.22 *A right-continuous Lévy process is a strong Markov process wrt. itself.*

Proof. It is enough to verify property (B) of a Feller process: Given an $f\in\mathbb{C}_B(\mathbb{R})$,

$$\int_{\mathbb{R}} f(\cdot)\,dP(\cdot,t,x,0) = \int_{\mathbb{R}} f(\cdot)\,d\mathbf{P}\{X(t)-X(0)\in\cdot-x\} = \int_{\mathbb{R}} f(\cdot+x)\,d\mathbf{P}\{X(t)-X(0)\in\cdot\}$$

is a (bounded) continuous function of x, by *Dominated Convergence*. □

*EXERCISE 46 Consider BM B^0 and a finite stopping time T wrt. $\{\sigma(B^0(r): 0\le r\le t)\}_{t\ge 0}$. Adapt the scheme from the proof of Theorem 4.21, to show that $\{B^0(t+T)-B^0(T)\}_{t\ge 0}$ is independent of $\mathcal{F}_T$, and has the same `fidi's` as $\{B^0(t)\}_{t\ge 0}$.

EXERCISE 47 Use Exercise 46 to explain why BM, eventually, hits any chosen point $x\in\mathbb{R}$ arbitrarily many (and thus infinitely many) times, with probability one.

*Theorem 4.23 *A right-continuous Lévy process, together with the transition probability from* Theorem 3.13, *is a Feller process.*

Proof. It is enough to show property (A) of a Feller process, since we have (B) from the proof of Corollary 4.22. By an inspection of that proof, we see that (A) holds iff.

$$\mathbf{P}\{|X(t)-X(0)|>\delta\}\to 0 \qquad \text{as}\quad t\downarrow 0, \quad \text{for each}\quad \delta>0.$$

[Recall that (A) concerns bounded and continuous functions.] This holds iff.

$$g(t;\theta)\equiv\mathbf{E}\{e^{i\theta(X(t)-X(0))}\}\to 1 \qquad \text{as}\quad t\downarrow 0, \quad \text{for each}\quad \theta>0. \tag{4.1}$$

However, by independence and stationarity of increments, we have, for $0 < k/n \in \mathbb{Q}$,

$$g(k/n;\theta) = \mathbf{E}\{e^{i\theta(X(k/n)-X(0))}\} = \prod_{i=1}^{k} \mathbf{E}\Big\{e^{i\theta(X(i/n)-X((i-1)/n))}\Big\} = g(1/n;\theta)^k.$$

Since this holds in particular for $k=n$, it follows that

$$g(k/n;\theta) = g(1/n;\theta)^k = \big(g(1;\theta)^{1/n}\big)^k = \big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\big)^{k/n}.$$

Since characteristic functions are continuous, this in turn give

$$g(t;\theta) = \big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\big)^t.$$

Now we get (4.1) by sending $t\downarrow 0$. □

*Remark 4.24 As indicated by our derivations of it, the strong Markov property is very close to the "usual" Markov property (see also Theorem 5.4). However, there exist weak Markov processes anyway, that do not have the strong Markov property. See e.g. [32, pp. 463-465] and [15, Section 2.15] for some material on this. #

5 Fifth Lecture

5.1 On the Markov Property

We write $\underline{\mathbb{L}_B(\mathbb{R})} \equiv \{(f:\mathbb{R}\to\mathbb{R}) : f \text{ is bounded and measurable}\}$.

Definition 5.1 *Let X be a Markov process. The* *σ-algebra of the future* $\underline{\mathcal{F}'_t} \equiv \sigma(X(s+t) : s\geq 0)$ *for $t\geq 0$. For a finite stopping time T,* $\underline{\mathcal{F}'_T} \equiv \sigma(X(s+T) : s\geq 0)$.

Theorem 5.2 (Markov Property) *Let $\{X(t)\}_{t\geq 0}$ be a stochastic process adapted to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. The following properties are equivalent:*

(1) *X is a Markov process wrt. $\{\mathcal{F}_t\}_{t\geq 0}$;*

(2) $\mathbf{P}\Big\{\bigcap_{i=1}^n \{X(t_i)\in\cdot\}\,\Big|\,\mathcal{F}_s\Big\} = \mathbf{P}\Big\{\bigcap_{i=1}^n \{X(t_i)\in\cdot\}\,\Big|\,X(s)\Big\}$ *for* $t_1,\ldots,t_n > s\geq 0$;

(3) $\mathbf{P}\{B|\mathcal{F}_s\} = \mathbf{P}\{B|X(s)\}$ *for* $B\in\mathcal{F}'_s$ *and* $s\geq 0$;

(4) $\mathbf{P}\{A|\mathcal{F}'_s\} = \mathbf{P}\{A|X(s)\}$ *for* $A\in\mathcal{F}_s$ *and* $s\geq 0$;

(5) $\mathbf{P}\{A\cap B|X(s)\} = \mathbf{P}\{A|X(s)\}\,\mathbf{P}\{B|X(s)\}$ *for* $A\in\mathcal{F}_s$, $B\in\mathcal{F}'_s$ *and* $s\geq 0$;

(6) $\mathbf{E}\{f(X(t))|\mathcal{F}_s\} = \mathbf{E}\{f(X(t))|X(s)\}$ *for* $f\in\mathbb{C}_B(\mathbb{R})$ *and* $t>s\geq 0$;

(7) $\mathbf{E}\{f(X(t))|\mathcal{F}_s\} = \mathbf{E}\{f(X(t))|X(s)\}$ *for* $f\in\mathbb{L}_B(\mathbb{R})$ *and* $t>s\geq 0$.

Proof (After [9, Section 1.1]). (1)⇒(5) This is Exercise 48 below.

(5)⇒(4) Given an $A\in\mathcal{F}_s$, for every $\Lambda\in\mathcal{F}'_s$, (5) gives

$$\begin{aligned}\int_\Lambda I_A\,d\mathbf{P} = \int_\Omega I_\Lambda I_A\,d\mathbf{P} = \int_\Omega \mathbf{P}\{\Lambda\cap A|X(s)\}\,d\mathbf{P} &= \int_\Omega \mathbf{P}\{\Lambda|X(s)\}\,\mathbf{P}\{A|X(s)\}\,d\mathbf{P}\\ &= \int_\Omega \mathbf{E}\Big\{I_\Lambda\mathbf{P}\{A|X(s)\}\,\Big|\,X(s)\Big\}\,d\mathbf{P}\\ &= \int_\Omega I_\Lambda\mathbf{P}\{A|X(s)\}\,d\mathbf{P}\\ &= \int_\Lambda \mathbf{P}\{A|X(s)\}\,d\mathbf{P}.\end{aligned}$$

(4)⇒(3) Given an $B\in\mathcal{F}'_s$, for every $\Lambda\in\mathcal{F}_s$, (4) gives

$$\begin{aligned}\int_\Lambda I_B\,d\mathbf{P} = \int_\Omega I_\Lambda I_B\,d\mathbf{P} = \int_\Omega \mathbf{P}\{\Lambda\cap B|\mathcal{F}'_s\}\,d\mathbf{P} &= \int_\Omega I_B\,\mathbf{P}\{\Lambda|\mathcal{F}'_s\}\,d\mathbf{P}\\ &= \int_\Omega I_B\,\mathbf{P}\{\Lambda|X(s)\}\,d\mathbf{P}\\ &= \int_\Omega \mathbf{E}\Big\{I_B\mathbf{P}\{\Lambda|X(s)\}\,\Big|\,X(s)\Big\}\,d\mathbf{P}\\ &= \int_\Omega \mathbf{P}\{B|X(s)\}\,\mathbf{P}\{\Lambda|X(s)\}\,d\mathbf{P}\\ &= \int_\Omega \mathbf{E}\Big\{I_\Lambda\mathbf{P}\{B|X(s)\}\,\Big|\,X(s)\Big\}\,d\mathbf{P}\end{aligned}$$

$$= \int_\Omega I_\Lambda \mathbf{P}\{B|X(s)\}\, d\mathbf{P}$$
$$= \int_\Lambda \mathbf{P}\{B|X(s)\}\, d\mathbf{P}.$$

(3)⇒(2)⇒(1) and (7)⇒(6) are trivial, while Exercise 44 shows that (1)⇒(7).

(6)⇒(1) By the *Dynkin System Lemma*, (1) follows if we show that $\mathbf{P}\{X(t)\in B|\mathcal{F}_s\} = \mathbf{P}\{X(t)\in B|X(s)\}$ for open $B\subseteq\mathbb{R}$ and $0\le s<t$. Now let $B_n = \{x\in B : |x-y|\ge 1/n \text{ for } y\in B^c\}$. Define $f_n(x)=1$ for $x\in B_n$, $f_n(x)=0$ for $x\in B^c$, and

$$f_n(x) = n\inf\{|x-y| : y\in B^c\} \qquad \text{for} \quad x\in B\backslash B_n.$$

Since $f_n\in\mathbb{C}_B(\mathbb{R})$, with $f_n(x)\to I_B(x)$ as $n\to\infty$, (6) gives (by Theorem 1.26)

$$\mathbf{P}\{X(t)\in B|\mathcal{F}_s\} \leftarrow \mathbf{E}\{f_n(X(t))|\mathcal{F}_s\} = \mathbf{E}\{f_n(X(t))|X(s)\} \to \mathbf{P}\{X(t)\in B|X(s)\}. \quad \square$$

Theorem 5.3 (Elementary Markov Property) *A stochastic process* $\{X(t)\}_{t\ge 0}$ *is a Markov process wrt. itself iff.*

$$\mathbf{P}\{X(t)\in\cdot\,|X(s_1),\ldots,X(s_n)\} = \mathbf{P}\{X(t)\in\cdot\,|X(s_n)\} \qquad \text{for} \quad t>s_n\ge\ldots\ge s_1\ge 0.$$

Proof. The implication to the left is a straightforward application of the *Dynkin System Lemma*. For the one to the right, notice that, for a Markov process X,

$$\begin{aligned}\mathbf{P}\{X(t)\in\cdot\,|X(s_1),\ldots,X(s_n)\} &= \mathbf{E}\Big\{\mathbf{P}\{X(t)\in\cdot\,|\mathcal{F}_{s_n}\}\,\Big|\,X(s_1),\ldots,X(s_n)\Big\}\\ &= \mathbf{E}\Big\{\mathbf{P}\{X(t)\in\cdot\,|X(s_n)\}\,\Big|\,X(s_1),\ldots,X(s_n)\Big\}\\ &= \mathbf{P}\{X(t)\in\cdot\,|X(s_n)\}. \quad \square\end{aligned}$$

Theorem 5.4 (Strong Markov Property) *Consider a measurable and adapted stochastic process* $\{X(t),\mathcal{F}_t\}_{t\ge 0}$, *together with a finite stopping time* T. *The following properties are equivalent:*

(1) $\mathbf{P}\{X(t+T)\in\cdot\,|\mathcal{F}_T\} = \mathbf{P}\{X(t+T)\in\cdot\,|X(T)\}$ *for* $t>0$;

(2) $\mathbf{P}\Big\{\bigcap_{i=1}^n\{X(t_i+T)\in\cdot\}\,\Big|\,\mathcal{F}_T\Big\} = \mathbf{P}\Big\{\bigcap_{i=1}^n\{X(t_i+T)\in\cdot\}\,\Big|\,X(T)\Big\}$ *for* $t_1,\ldots,t_n>0$;

(3) $\mathbf{P}\{B|\mathcal{F}_T\} = \mathbf{P}\{B|X(T)\}$ *for* $B\in\mathcal{F}'_T$;

(4) $\mathbf{P}\{A|\mathcal{F}'_T\} = \mathbf{P}\{A|X(T)\}$ *for* $A\in\mathcal{F}_T$;

(5) $\mathbf{P}\{A\cap B|X(T)\} = \mathbf{P}\{A|X(T)\}\,\mathbf{P}\{B|X(T)\}$ *for* $A\in\mathcal{F}_T$ *and* $B\in\mathcal{F}'_T$;

(6) $\mathbf{E}\{f(X(t+T))|\mathcal{F}_T\} = \mathbf{E}\{f(X(t+T))|X(T)\}$ *for* $f\in\mathbb{C}_B(\mathbb{R})$ *and* $t>0$;

(7) $\mathbf{E}\{f(X(t+T))|\mathcal{F}_T\} = \mathbf{E}\{f(X(t+T))|X(T)\}$ *for* $f\in\mathbb{L}_B(\mathbb{R})$ *and* $t>0$.

Proof. We have (6) from Theorem 4.21. From (6) in turn, properties (1)-(5) and (7) are established by the same arguments as those use in the proof of Theorem 5.2. □

EXERCISE 48 Prove the implication (1)⇒(5) in Theorem 5.2.

EXERCISE 49 Show how property (7) follows from (6) in Theorem 5.4.

*EXERCISE 50 Show how properties (1)-(5) follow from (7) in Theorem 5.4.

5.2 Strong Markov Property of Lévy Processes

Theorem 5.5 *Let $\{X(t)\}_{t\geq 0}$ be a right-continuous Lévy process, and T a finite stopping time wrt. $\{\mathcal{F}_t\}_{t\geq 0} = \{\sigma(X(s) : 0\leq s\leq t)\}_{t\geq 0}$. The process $\{X(t+T)-X(T)\}_{t\geq 0}$ is independent of $\mathcal{F}_T$, and has the same* `fidi's` *as $\{X(t)-X(0)\}_{t\geq 0}$.*

Proof. Given an $A\in\mathcal{F}_T$, we shall prove that

$$\mathbf{P}\Big\{A\cap\bigcap_{i=1}^n\{X(t_i+T)-X(T)\in B_i\}\Big\} = \mathbf{P}\{A\}\,\mathbf{P}\Big\{\bigcap_{i=1}^n\{X(t_i)-X(0)\in B_i\}\Big\}$$

for $B_1,\ldots,B_n\in\mathcal{B}(\mathbb{R})$, $t_1,\ldots,t_n\geq 0$ and $n\in\mathbb{N}$. For this, it is enough to prove that

$$\mathbf{P}\Big\{A\cap\bigcap_{i=1}^n\{X(t_i+T)-X(t_{i-1}+T)\in B_i\}\Big\} = \mathbf{P}\{A\}\prod_{i=1}^n\mathbf{P}\{X(t_i-t_{i-1})-X(0)\in B_i\}$$

for $B_1,\ldots,B_n\in\mathcal{B}(\mathbb{R})$ and $0=t_0<t_1<\ldots<t_n$. By right-continuity, this follows if

$$\mathbf{P}\Big\{A\cap\bigcap_{i=1}^n\{X(t_i+T)-X(s_i+T)\in B_i\}\Big\} = \mathbf{P}\{A\}\prod_{i=1}^n\mathbf{P}\{X(t_i-s_i)-X(0)\in B_i\} \qquad (5.1)$$

for $B_1,\ldots,B_n\in\mathcal{B}(\mathbb{R})$, $0=t_0<s_1<t_1<\ldots<s_n<t_n$ and $n\in\mathbb{N}$. By the *Strong Markov Property* (Theorem 5.4) and Corollary 4.22, we have

$$\begin{aligned}
&\mathbf{P}\Big\{A\cap\bigcap_{i=1}^n\{X(t_i+T)-X(s_i+T)\in B_i\}\Big\}\\
&=\mathbf{E}\Big\{\mathbf{P}\Big\{A\cap\bigcap_{i=1}^n\{X(t_i+T)-X(s_i+T)\in B_i\}\Big|\mathcal{F}_{T+s_n}\Big\}\Big\}\\
&=\mathbf{E}\Big\{\mathbf{P}\{X(t_n+T)-X(s_n+T)\in B_n|\mathcal{F}_{T+s_n}\}\,I_A\prod_{i=1}^{n-1}I_{\{X(t_i+T)-X(s_i+T)\in B_i\}}\Big\}\\
&=\mathbf{E}\Big\{\mathbf{P}\{X(t_n+T)-X(s_n+T)\in B_n|X(T+s_n)\}\,I_A\prod_{i=1}^{n-1}I_{\{X(t_i+T)-X(s_i+T)\in B_i\}}\Big\}. \qquad (5.2)
\end{aligned}$$

Here we used the facts that $A\in\mathcal{F}_T\subseteq\mathcal{F}_{T+s_n}$, by Theorem 4.15, and that $\sigma(X(\tau+T))\subseteq\mathcal{F}_{T+s_n}$ for $\tau<s_n$. This latter fact follows noticing that

$$B\cap\{\tau+T<t\}\in\mathcal{F}_t \qquad \text{for} \quad B\in\sigma(X(\tau+T)) \quad \text{and} \quad t\geq 0,$$

by the proof of Theorem 4.20, which gives $B\in\mathcal{F}_{T+s_n}$, since

$$B\cap\{T+s_n\le t\}=\bigcap_{i=I}^{\infty}B\cap\{T+s_n<t+1/i\}=\bigcap_{i=I}^{\infty}B\cap\{\tau+T<t+1/i+\tau-s_n\}\in\mathcal{F}_t$$

for $I\in\mathbb{N}$ sufficiently large, because $\mathcal{F}_{t+1/i+\tau-s_n}\subseteq\mathcal{F}_t$ (recall Definition 4.14).

By (5.2) and an induction argument, to prove (5.1), it is enough to show that

$$\mathbf{P}\{X(t+T)-X(s+T)\in B|X(T+s)\}=\mathbf{P}\{X(t-s)-X(0)\in B\}$$

for $B\in\mathcal{B}(\mathbb{R})$ and $0<s<t$. By the *Dynkin System Lemma*, this follows if

$$\mathbf{P}\{X(t+T)-X(s+T)\in B|X(T+s)\}=\mathbf{P}\{X(t-s)-X(0)\in B\}$$

for open $B\subseteq\mathbb{R}$ and $0<s<t$. This in turn follows if (cf. the proof of Theorem 5.2)

$$\mathbf{E}\{f(X(t+T)-X(s+T))|X(T+s)\}=\mathbf{E}\{f(X(t-s)-X(0))\}$$

for $f\in\mathbb{C}_B(\mathbb{R})$ and $0<s<t$. However, by Theorems 3.5, 3.13 and 4.20, we have

$$\begin{aligned}&\mathbf{E}\{f(X(t+T)-X(s+T))|X(T+s)\}\\ &\quad=\int_{\mathbb{R}}\mathbf{E}\{f(X(t+T)-y)|X(T+s)=y\}\,dF_{X(T+s)}(y)\\ &\quad=\int_{\mathbb{R}}\int_{\mathbb{R}}f(\cdot-y)\,dP(\cdot,t-s,y,0)\,dF_{X(T+s)}(y)\\ &\quad=\int_{\mathbb{R}}\int_{\mathbb{R}}f(\cdot-y)\,d\mathbf{P}\{X(t-s)-X(0)\in\cdot-y\}\,dF_{X(T+s)}(y)\\ &\quad=\int_{\mathbb{R}}\int_{\mathbb{R}}f(\cdot)\,d\mathbf{P}\{X(t-s)-X(0)\in\cdot\}\,dF_{X(T+s)}(y)=\mathbf{E}\{f(X(t-s)-X(0))\}.\quad\square\end{aligned}$$

*Remark 5.6 A "state of the art" direct proof of Theorem 5.5 uses the easily checked fact that, by the proof of Theorem 4.23, $M(t)=\mathrm{e}^{i\theta X(t)}(\mathbf{E}\{\mathrm{e}^{i\theta X(1)}\})^{-t}$ is a (complex-valued) martingale, together with the *Optional Sampling theorem* (see Example 23.11). The proof in full detail is not much shorter than that above, but less complex since from scratch, and not relying on the strong Markov property. #

Corollary 5.7 (Strong Markov Property of BM) *Consider BM B and a finite stopping time T wrt. $\{\mathcal{F}_t\}_{t\ge0}=\{\sigma(B(s):0\le s\le t)\}_{t\ge0}$. The process $\{B(t+T)-B(T)\}_{t\ge0}$ is independent of $\mathcal{F}_T$, and has the same* `fidi's` *as $\{B(t)-B(0)\}_{t\ge0}$.*

6 Sixth Lecture

6.1 Introduction to Stochastic Integrals wrt. BM

Throughout this lecture, $\{B(t)\}_{t\geq 0}$ is BM with $B(0)=0$ and filtration $\{\mathcal{F}_t\}_{t\geq 0} = \{\sigma(B(s):0\leq s\leq t)\}_{t\geq 0}$, and $\{X(t)\}_{t\geq 0}$ is a stochastic process adapted to $\{\mathcal{F}_t\}_{t\geq 0}$.

In the lecture we discuss stochastic integrals $\int_0^t X\,dB$ wrt. BM in a "soft" sense, meaning that statements and proofs, albeit correct in essence, in general require additional attention to technical details, in order to become rigorous. Consequently, exercises to the lecture must be understood and solved in that spirit.

The Itô integral $\int_0^t X\,dB$ will be defined as a Riemann-Stieltjes type limit

$$\underline{\int_0^t X(r)\,dB(r)} \equiv \lim\left\{\sum_{i=1}^n X(t_{i-1})\,(B(t_i)-B(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}.$$

When X has finite variation over $[0,t]$, we may equivalently use the definition

$$\lim\left\{\sum_{i=1}^n X(t_i)\,(B(t_i)-B(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\},$$

because the difference between this limit and the Itô integral is precisely

$$\lim\left\{\sum_{i=1}^n (X(t_i)-X(t_{i-1}))\,(B(t_i)-B(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}.$$

By Definition 1.10 and Therorem 1.11, this is $[X,B](t)=0$ (since BM is continuous).

Famous Example 6.1 The approximating sums of $\int_0^t B\,dB$ have mean

$$\mathbf{E}\left\{\sum_{i=1}^n B(t_{i-1})\,(B(t_i)-B(t_{i-1}))\right\} = \sum_{i=1}^n \mathbf{E}\{B(t_{i-1})\}\,\mathbf{E}\{B(t_i)-B(t_{i-1})\} = 0,$$

by independence of increments. But if we sample B in t_i instead of t_{i-1}, we get

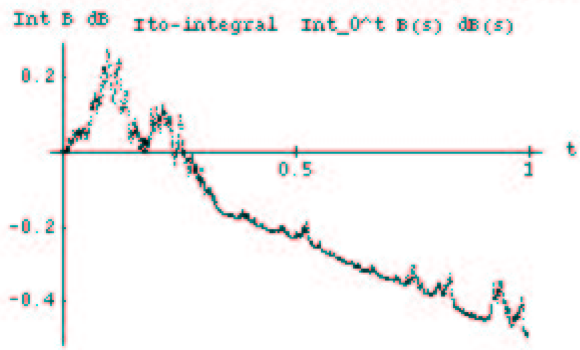
$$\begin{aligned} &\mathbf{E}\left\{\sum_{i=1}^n B(t_i)\,(B(t_i)-B(t_{i-1}))\right\} \\ &= \sum_{i=1}^n \mathbf{E}\left\{(B(t_i)-B(t_{i-1}))^2 + B(t_{i-1})\,(B(t_i)-B(t_{i-1}))\right\} = \sum_{i=1}^n (t_i-t_{i-1}) = t. \quad \# \end{aligned}$$

A third natural way to integrate is the Stratonovich (Fisk-Stratonovich) integral

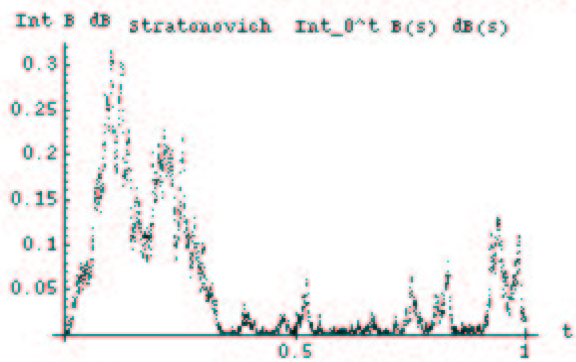
$$\begin{aligned} &\underline{\int_0^t X(r)\circ dB(r)} \\ &\equiv \lim\left\{\sum_{i=1}^n \frac{X(t_{i-1})+X(t_{i-1})}{2}\,(B(t_i)-B(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}. \end{aligned}$$

When X has finite variation over $[0,t]$, we have $\int_0^t X\,dB = \int_0^t X\circ dB$ (by the above argument), but in general, the integrals need not be equal (see Example 6.1). It turns out that there are deep differences between the two, which we will illuminate below.

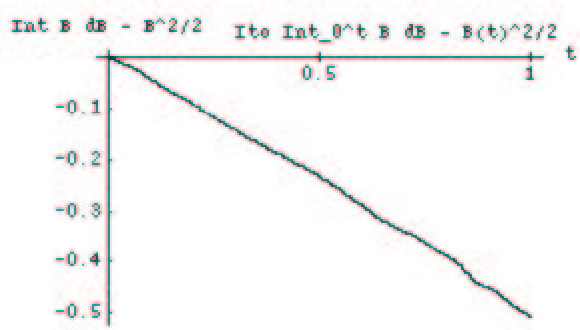
```
In[1]:= << Statistics`ContinuousDistributions`

In[2]:= Incr = N[Table[Random[NormalDistribution[0, 1/Sqrt[2000]]], {2001}]];

In[3]:= B = {Incr[[1]]}; For[k = 2, k <= 2001, k ++, B = Join[B, {B[[k - 1]] + Incr[[k]]}]];

In[4]:= ItoBdB = {0}; For[k = 2, k ≤ 2001, k ++, ItoBdB
                    = Join[ItoBdB, {ItoBdB[[k - 1]] + B[[k - 1]] * (B[[k]] - B[[k - 1]])}]];

In[5]:= StratoBdB = {B[[1]]^2/2}; For[k = 2, k ≤ 2001, k ++, StratoBdB = Join[StratoBdB,
                  {StratoBdB[[k - 1]] + (B[[k - 1]] + B[[k]]) * (B[[k]] - B[[k - 1]]) / 2}]];

In[6]:= ItoBdBminushalfBsquare = Table[ItoBdB[[k]] - B[[k]]^2/2, {k, 1, 2000}];

In[7]:= ListPlot[ItoBdB, Ticks → {{{1000, "0.5", 0.02}, {2000, "1", 0.02}}, Automatic},
         AxesLabel → {"t", "Int B dB"}, PlotLabel → "      Ito-integral  Int_0^t B(s) dB(s)"]
```




```
Out[7]= - Graphics -

In[8]:= ListPlot[StratoBdB, Ticks → {{{1000, "0.5", 0.02}, {2000, "1", 0.02}}, Automatic},
         AxesLabel → {"t", "Int B dB"}, PlotLabel → "     Stratonovich  Int_0^t B(s) dB(s)"]
```




```
Out[8]= - Graphics -

In[9]:= ListPlot[ItoBdBminushalfBsquare, Ticks → {{{1000, "0.5", 0.02}, {2000, "1", 0.02}},
          Automatic}, AxesLabel → {"t", "Int B dB - B^2/2"},
          PlotLabel → "                 Ito Int_0^t B dB - B(t)^2/2"]
```



*Remark 6.2 The Riemann-Stieltjes integral is defined as the limit

$$\lim\left\{\sum_{i=1}^{n} X(\tau_i)\,(B(t_i)-B(t_{i-1})) : \begin{array}{c}\tau_i\in(t_{i-1},t_i],\ 0=t_0<t_1<\ldots<t_n=t\\ n\in\mathbb{N},\ \max_{1\le i\le n} t_i-t_{i-1}\to 0\end{array}\right\},$$

if it exists. Hence the Itô integral is not really a Riemann-Stieltjes integral, since the choice of the time $\tau_i\in(t_{i-1},t_i]$ to sample X may matter (so that the limit does not exist). In fact, nearly all measurable and adapted processes are Itô integrable, which is a much richer class of processes than the Riemann-Stieltjes integrable ones. #

Exercise 51 Why is it not possible to define the integral

$$Y(t) = Y(\omega,t) = \int_{r=0}^{r=t} X(\omega,r)\,dB(\omega,r) \qquad \text{for} \quad t\ge 0,$$

simply by means of, for each $\omega\in\Omega$, let $Y(\omega,t)$ take the value of the Riemann-Stieltejs (Lebesgue-Stieltejs) integral on the right-hand side?

The stochastic integral $\int_0^t X\,dB$ cannot in general be understood in sample path meaning, as the Lebesgue-Stieltjes integral, or Riemann-Stieltjes integral, of X wrt. dB (Exercise 51). This gives the following relation, which every reader must digest:

$$\boxed{\Big(\int_0^t X(r)\,dB(r)\Big)(\omega) \neq \int_0^t X(\omega,r)\,dB(\omega,r)}\,.$$

We use the notation $\int_0^t X\,dB$ for the stochastic integal, and calls it an integral, because it has properties similar to those of integrals in mathematics, and is constructed in a similar way. But it has no defining relation with deterministic integrals.

Exercise 52 When X has finite variation over $[0,t]$, deterministic integration methodology can be used to define $\int_0^t X\,dB$. How?

We are going to study SDE (stochastic differential equations) of the type

$$dY(t) = a(t,Y(t))\,dt + b(t,Y(t))\,dB(t) \qquad \text{for} \quad t\in[0,T], \quad Y(0)=Y_0. \tag{6.1}$$

A solution to (6.1) is a stochastic process Y, called diffusion process, such that

$$Y(t) = Y_0 + \int_0^t a(r,Y(r))\,dr + \int_0^t b(r,Y(r))\,dB(r) \qquad \text{for} \quad t\in[0,T]. \tag{6.2}$$

The functions $a,b:\mathbb{R}^2\to\mathbb{R}$ will in general be "smooth", and the process Y continuous. Thus the first integral can be taken in the Riemann sense. It is the second integral that is "problematic", since Y will in general not have finite variation.

The solution to (6.2) should be understood in the Itô integral sense

$$Y(t) = Y_0 + \lim \sum_{t_i\le t} a(t_{i-1},Y(t_{i-1}))\,(t_i-t_{i-1}) + \sum_{t_i\le t} b(t_{i-1},Y(t_{i-1}))\,(B(t_i)-B(t_{i-1})),$$

where the limit is taken over partitions $0 = t_0 < \ldots < t_n = T$ such that $\max_{1 \le i \le n} t_i - t_{i-1} \to 0$. This means that

$$Y(t) \approx Y_0 + \sum_{t_i \le t} a(t_{i-1}, Y(t_{i-1}))\,(t_i - t_{i-1}) + \sum_{t_i \le t} b(t_{i-1}, Y(t_{i-1}))\,(B(t_i) - B(t_{i-1})),$$

with better approximation for finer partitions. In particular, we have

$$Y(t_j) \approx Y_0 + \sum_{t_i \le t_j} a(t_{i-1}, Y(t_{i-1}))\,(t_i - t_{i-1}) + \sum_{t_i \le t} b(t_{i-1}, Y(t_{i-1}))\,(B(t_i) - B(t_{i-1}))$$

for $j = 1, \ldots, n$, which can be rewritten as a recursive (*Euler iteration*) scheme

$$\boxed{Y(t_j) = Y(t_{j-1}) + a(t_{j-1}, Y(t_{j-1}))\,(t_j - t_{j-1}) + b(t_{j-1}, Y(t_{j-1}))\,(B(t_j) - B(t_{j-1}))}.$$

By recursion, this scheme gives the values $Y(t_1), \ldots, Y(t_n)$, of the solution to the SDE, on the partition. The solution Y calculated in this way, is not an exact solution to (6.2), but an approximation, obtained by "not going all the way to the limit". Nevertheless, with more attention to details, this is a useful method to solve the SDE, both to prove the existence of the solution, and to compute it numerically.

Notice that if we use the Stratonovich stochastic integral, then $Y(t_j)$ will appear both on the left-hand side and right-hand side of the recursion. This gives some insight into the theoretical superiority of the Itô integral to other alternatives.

Another important and related feature of the Itô integral, is that it is a martingale:

$$\begin{aligned}
\mathbf{E}\Big\{\int_0^t X(r)\,dB(r)\,\Big|\,\mathcal{F}_s\Big\} &= \lim \mathbf{E}\Big\{\sum_{t_i \le t} X(t_{i-1})\,(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} \\
&= \lim \mathbf{E}\Big\{\sum_{t_i \le t,\, t_{i-1} > s} X(t_{i-1})\,(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} \\
&\qquad + \lim \mathbf{E}\Big\{\sum_{t_i \le s} X(t_{i-1})\,(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} \\
&\qquad - \lim \mathbf{E}\Big\{\sum_{t_{i-1} \le s,\, t_i > s} X(t_{i-1})\,(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} \\
&= \lim \sum_{t_i \le t,\, t_{i-1} > s} \mathbf{E}\Big\{\mathbf{E}\Big\{X(t_{i-1})\,(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_{t_{i-1}}\Big\}\,\Big|\,\mathcal{F}_s\Big\} \\
&\qquad + \lim \sum_{t_i \le s} X(t_{i-1})\,(B(t_i) - B(t_{i-1})) + 0 \\
&= \lim \sum_{t_i \le t,\, t_{i-1} > s} \mathbf{E}\Big\{X(t_{i-1})\,\mathbf{E}\Big\{B(t_i) - B(t_{i-1})\,\Big|\,\mathcal{F}_{t_{i-1}}\Big\}\,\Big|\,\mathcal{F}_s\Big\} \\
&\qquad + \int_0^s X(r)\,dB(r) + 0 \\
&= 0 + \int_0^s X(r)\,dB(r) + 0.
\end{aligned}$$

Here we used that $B(t_i) - B(t_{i-1})$ is independent of $\mathcal{F}_{t_{i-1}}$, and $X(t_{i-1})$ adapted to it. Notice that the second sum, on the right-hand side of the second equality, is adapted, while the third sum has at most one term, and thus goes to zero.

The approximate recursive scheme for solving the SDE (6.1)-(6.2) indicates that a solution is a *Markov process*, since the next value of the solution only depends on "the history", through the current value. This is proven formally in Section 22.1.

Example 6.3 Let X be a solution to the Itô SDE

$$dX(t) = (1/2)\,X(t)\,dt + X(t)\,dB(t).$$

Notice that the Itô sense differential of X is

$$X(t_i)-X(t_{i-1}) = dX(t_i) = (1/2)\,X(t_{i-1})\,(t_i-t_{i-1}) + X(t_{i-1})\,(B(t_i)-B(t_{i-1})).$$

For the Stratonovich integral of this Itô solution, we therefore have

$$\begin{aligned}
&\int_0^t X(r)\circ dB(r)\\
&\quad= \lim\sum_{i=1}^n \frac{X(t_i)+X(t_{i-1})}{2}\,(B(t_i)-B(t_{i-1}))\\
&\quad= \lim\sum_{i=1}^n \frac{X(t_i)-X(t_{i-1})}{2}\,(B(t_i)-B(t_{i-1})) + \lim\sum_{i=1}^n X(t_{i-1})\,(B(t_i)-B(t_{i-1}))\\
&\quad= \lim\sum_{i=1}^n \frac{\frac{1}{2}X(t_{i-1})\,(t_i-t_{i-1}) + X(t_{i-1})\,(B(t_i)-B(t_{i-1}))}{2}\,(B(t_i)-B(t_{i-1}))\\
&\qquad\qquad + \int_0^t X(r)\,dB(r)\\
&\quad= 0 + \lim\sum_{i=1}^n \frac{X(t_{i-1})}{2}\,(B(t_i)-B(t_{i-1}))^2 + \int_0^t X(r)\,dB(r)\\
&\quad= \frac{1}{2}\int_0^t X(r)\,dr + \int_0^t X(r)\,dB(r),
\end{aligned}$$

(where $0=t_0<\ldots<t_n=t$ and $\max_{1\leq i\leq n} t_i-t_{i-1}\rightarrow 0$.) Here we used the fact that

$$\left|\sum_{i=1}^n \frac{X(t_{i-1})\,(t_i-t_{i-1})}{4}\,(B(t_i)-B(t_{i-1}))\right| \leq \sup_{s\in[0,t]}\frac{|X(s)|}{4}\max_{1\leq i\leq n}|B(t_i)-B(t_{i-1})|\,t\rightarrow 0,$$

by continuity of BM, assuming that X is continuous. Further, picking a coarser partition $0=\tilde{t}_0<\ldots<\tilde{t}_m=t$, we get, by Theorem 2.13,

$$\begin{aligned}
\sum_{i=1}^n \frac{X(t_{i-1})}{2}\,(B(t_i)-B(t_{i-1}))^2 &= \sum_{j=1}^m \sum_{t_i\in(\tilde{t}_{j-1},\tilde{t}_j]} \frac{X(t_{i-1})-X(\tilde{t}_{j-1})}{2}\,(B(t_i)-B(t_{i-1}))^2\\
&\qquad + \sum_{j=1}^m \frac{X(\tilde{t}_{j-1})}{2}\sum_{t_i\in(\tilde{t}_{j-1},\tilde{t}_j]}(B(t_i)-B(t_{i-1}))^2\\
&\sim 0 + \sum_{j=1}^m \frac{X(\tilde{t}_{j-1})}{2}\,[B]([\tilde{t}_{j-1},\tilde{t}_j])\\
&= \sum_{j=1}^m \frac{X(\tilde{t}_{j-1})}{2}\,(\tilde{t}_j-\tilde{t}_{j-1}) \rightarrow \frac{1}{2}\int_0^t X(r)\,dr,
\end{aligned}$$

as first the partition $\{t_i\}_{i=0}^n$, and then the partition $\{\tilde{t}_j\}_{j=0}^m$, becomes infinitely fine. Here we used the fact that, by uniform continuity of X together with Theorem 2.13,

$$\Big|\sum_{j=1}^{m}\sum_{t_i\in(\tilde{t}_{j-1},\tilde{t}_j]}\frac{X(t_{i-1})-X(\tilde{t}_{j-1})}{2}\,(B(t_i)-B(t_{i-1}))^2\Big|$$

$$\leq \sup_{j\in\{1,\ldots,m\}}\ \sup_{t\in(\tilde{t}_{j-1},\tilde{t}_j]}\frac{|X(t)-X(\tilde{t}_{j-1})|}{2}\sum_{i=1}^{n}(B(t_i)-B(t_{i-1}))^2 \;\rightarrow\; 0\cdot t \;=\; 0. \quad \#$$

In some applications, the Stratonovich integral is the natural one. Luckily, it turns out that the solution to the SDE (6.1), taken in the Stratonovich sense, that is,

$$Y(t) = Y_0 + \int_0^t a(r,Y(r))\,dr + \int_0^t b(r,Y(r))\circ dB(r) \qquad \text{for} \quad t\in[0,T], \tag{6.3}$$

can be obtained as the Itô sense solution to the modified equation (e.g., [37, p. 36])*

$$dY(t) = \Big(a(t,Y(t))+\tfrac{1}{2}\,b(t,Y(t))\,\partial_2 b(t,Y(t))\Big)\,dt+b(t,Y(t))\,dB(t), \quad Y(0)=Y_0. \tag{6.4}$$

EXERCISE 53 Prove (6.4). (**Hint:** Taylor expand the approximating sum.)

The major drawback with the Itô stochastic calculus, is that "natural formulas" do not hold. For example, we have the Itô differential

$$d(B(t)^2) = 2\,B(t)\,dB(t)+dt \neq 2\,B(t)\,dB(t) \tag{6.5}$$

of squared BM, while the corresponding Stratonowich differential is "natural one"

$$d(B(t)^2) = 2\,B(t)\circ dB(t).$$

[Of course, "unnatural" relations like (6.5) only occur when differentiating non-differentiable processes (with infinite variation).]

Famous Example 6.4 We calculate the Itô integral $\int_0^t B\,dB$: We have

$$\begin{aligned}
\int_0^t B(r)\,dB(r) &\leftarrow \sum_{i=1}^{n} B(t_{i-1})\,(B(t_i)-B(t_{i-1}))\\
&= \sum_{i=1}^{n}(B(t_i)^2-B(t_{i-1})^2) - \sum_{i=1}^{n} B(t_i)\,(B(t_i)-B(t_{i-1}))\\
&= B(t)^2-B(0)^2 - \sum_{i=1}^{n}(B(t_i)-B(t_{i-1}))^2 - \sum_{i=1}^{n} B(t_{i-1})\,(B(t_i)-B(t_{i-1}))\\
&\rightarrow B(t)^2 - [B](t) - \int_0^t B(r)\,dB(r)
\end{aligned}$$

as $n\rightarrow\infty$ [since $B(0)=0$]. Rearranging, Theorem 2.13 thus gives

$$\boxed{\int_0^t B(r)\,dB(r) = \frac{1}{2}\,B(t)^2 - \frac{1}{2}\,t}\,. \quad \#$$

EXERCISE 54 Prove (6.5).

7 Seventh Lecture

7.1 Maximal Inequalities for Martingales

Here we give two smartingale inequalities, which together with *Burkholder-Davis-Gundy inequalities* in Section 23.5, argubly are the most important for smartingale.

Both inequalities in this section are consequences of the following easy estimate:

Theorem 7.1 *For a right-continuous submartingale* $\{X(t)\}_{t\in[0,T]}$, *we have*

$$\mathbf{P}\Big\{\sup_{0\leq t\leq T} X(t)\geq\lambda\Big\}\leq\frac{1}{\lambda}\int_{\{\sup_{0\leq t\leq T} X(t)\geq\lambda\}} X(T)\,d\mathbf{P}\qquad \textit{for}\quad \lambda>0. \tag{7.1}$$

Proof. To prove (7.1), it is enough to show that

$$\mathbf{P}\Big\{\sup_{0\leq t\leq T} X(t)>\lambda\Big\}\leq\frac{1}{\lambda}\int_{\{\sup_{0\leq t\leq T} X(t)>\lambda\}} X(T)\,d\mathbf{P}\qquad \text{for}\quad \lambda>0, \tag{7.2}$$

because from this together with *Dominated Convergence*, we deduce that

$$\begin{aligned}\mathbf{P}\Big\{\sup_{0\leq t\leq T} X(t)\geq\lambda\Big\}=\lim_{\varepsilon\downarrow 0}\mathbf{P}\Big\{\sup_{0\leq t\leq T} X(t)>\lambda-\varepsilon\Big\}&\leq\lim_{\varepsilon\downarrow 0}\frac{1}{\lambda-\varepsilon}\int_{\{\sup_{0\leq t\leq T}X(t)>\lambda-\varepsilon\}}X(T)\,d\mathbf{P}\\ &=\frac{1}{\lambda}\int_{\{\sup_{0\leq t\leq T}X(t)\geq\lambda\}}X(T)\,d\mathbf{P},\end{aligned}$$

since $I_{\{\sup_{0\leq t\leq T}X(t)>\lambda-\varepsilon\}}X(T)\rightarrow I_{\{\sup_{0\leq t\leq T}X(t)\geq\lambda\}}X(T)$ a.s., with $|I_{\{\sup_{0\leq t\leq T}X(t)>\lambda-\varepsilon\}}X(T)|\leq|X(T)|$ (which is integrable since X is a submartingale).

To prove (7.2) in turn, it is enough to show that

$$\mathbf{P}\Big\{\sup_{0\leq k\leq n} X(kT/n)>\lambda\Big\}\leq\frac{1}{\lambda}\int_{\{\sup_{0\leq k\leq n}X(kT/n)>\lambda\}} X(T)\,d\mathbf{P}\qquad \text{for}\quad \lambda>0, \tag{7.3}$$

since by right-continuity of X, we have $I_{\{\sup_{0\leq k\leq n}X(kT/n)>\lambda\}}\rightarrow I_{\{\sup_{0\leq t\leq T}X(t)>\lambda\}}$ a.s., so that *Dominated Convergence* gives (7.2) when we send $n\rightarrow\infty$ in (7.3).

To prove (7.3), let $\tau\equiv T\wedge\min\{\ell\in\{0,\ldots,n\}:Y(\ell T/n)>\lambda\}$. Since $\{\tau=\ell T/n\}\in\mathcal{F}_{\ell T/n}$ for $\ell=0,\ldots,n$, we get (7.3) from the estimates

$$\begin{aligned}\int_{\{\sup_{0\leq k\leq n}X(kT/n)>\lambda\}}X(T)\,d\mathbf{P}&=\sum_{\ell=0}^{n}\int_{\{\sup_{0\leq k\leq n}X(kT/n)>\lambda,\,\tau=\ell T/n\}}X(T)\,d\mathbf{P}\\ &=\sum_{\ell=0}^{n}\int_{\{\sup_{0\leq k\leq \ell}X(kT/n)>\lambda,\,\tau=\ell T/n\}}X(T)\,d\mathbf{P}\\ &=\sum_{\ell=0}^{n}\int_{\{\sup_{0\leq k\leq \ell}X(kT/n)>\lambda,\,\tau=\ell T/n\}}\mathbf{E}\{X(T)|\mathcal{F}_{\ell T/n}\}\,d\mathbf{P}\\ &\geq\sum_{\ell=0}^{n}\int_{\{\sup_{0\leq k\leq \ell}X(kT/n)>\lambda,\,\tau=\ell T/n\}}X(\ell T/n)\,d\mathbf{P}\\ &\geq\int_{\{\sup_{0\leq k\leq n}X(kT/n)>\lambda\}}X(\tau)\,d\mathbf{P}\\ &\geq\lambda\int_{\{\sup_{0\leq k\leq n}X(kT/n)>\lambda\}}d\mathbf{P}=\lambda\,\mathbf{P}\Big\{\max_{0\leq k\leq n}X(kT/n)>\lambda\Big\}.\ \square\end{aligned}$$

EXERCISE 55 Let $\{X(t)\}_{t\in[0,T]}$ be a martingale, and $f:\mathbb{R}\to\mathbb{R}$ convex. Show that $\{f(X(t))\}_{t\in[0,T]}$ is a submartingale when $\mathbf{E}\{|f(X(t))|\}<\infty$ for $t\in[0,T]$.

Corollary 7.2 (DOOB-KOLMOGOROV INEQUALITY) *For a right-continuous martingale $\{X(t)\}_{t\in[0,T]}$, such that $\mathbf{E}\{|X(t)|^p\}<\infty$ for $t\in[0,T]$, where $p\geq 1$ is a constant, we have*

$$\mathbf{P}\Big\{\sup_{0\leq t\leq T}|X(t)|\geq\lambda\Big\}\leq\mathbf{E}\{|X(T)|^p\}\,/\,\lambda^p\qquad\textit{for}\quad\lambda>0.$$

Proof. Since $|\cdot|^p$ is convex, $|X|^p$ is a submartingale, by Exercise 55. Hence Theorem 7.1 gives

$$\begin{aligned}\mathbf{P}\Big\{\sup_{0\leq t\leq T}|X(t)|\geq\lambda\Big\}&=\mathbf{P}\Big\{\sup_{0\leq t\leq T}|X(t)|^p\geq\lambda^p\Big\}\\&\leq\int_{\{\sup_{0\leq t\leq T}|X(t)|^p\geq\lambda\}}|X(T)|^p\,d\mathbf{P}\Big/\lambda^p\leq\mathbf{E}\{|X(T)|^p\}\,/\,\lambda^p.\qquad\square\end{aligned}$$

Corollary 7.3 (DOOB MAXIMAL INEQUALITY) *For a right-continuous and positive submartingale $\{X(t)\}_{t\in[0,T]}$, such that $\mathbf{E}\{X(T)^p\}<\infty$, where $p>1$ is a constant, we have*

$$\mathbf{E}\Big\{\Big(\sup_{0\leq t\leq T}X(t)\Big)^p\Big\}\leq(p/(p-1))^p\,\mathbf{E}\{X(T)^p\}.$$

Proof. By Theorem 7.1, together with *Hölder's inequality* $\mathbf{E}\{|X_1X_2|\}\leq(\mathbf{E}\{|X_1|^p\})^{1/p}(\mathbf{E}\{|X_2|^{p/(p-1)}\})^{(p-1)/p}$ (e.g., [34, p. 65]), writing $X^\star\equiv\sup_{0\leq t\leq T}X(t)$, we have

$$\begin{aligned}\mathbf{E}\{(X^\star\wedge n)^p\}=\int_0^\infty\mathbf{P}\{(X^\star\wedge n)^p>\lambda\}\,d\lambda&=\int_0^{n^p}\mathbf{P}\{X^\star>\lambda^{1/p}\}\,d\lambda\\&\leq\int_0^{n^p}\Big(\lambda^{-1/p}\int_{\{X^\star\geq\lambda^{1/p}\}}X(T)\,d\mathbf{P}\Big)\,d\lambda\\&=\int_\Omega\Big(\int_0^{(X^\star\wedge n)^p}\lambda^{-1/p}\,d\lambda\Big)X(T)\,d\mathbf{P}\\&=\int_\Omega\frac{p}{p-1}\,(X^\star\wedge n)^{p-1}\,X(T)\,d\mathbf{P}\\&\leq\frac{p}{p-1}\Big(\mathbf{E}\{(X^\star\wedge n)^p\}\Big)^{(p-1)/p}(\mathbf{E}\{X(T)^p\})^{1/p}.\end{aligned}$$

Dividing by $(\mathbf{E}\{(X^\star\wedge n)^p\})^{(p-1)/p}$ on both sides, *Fatou's Lemma* thus gives

$$(\mathbf{E}\{(X^\star)^p\})^{1/p}\leq\liminf_{n\to\infty}(\mathbf{E}\{(X^\star\wedge n)^p\})^{1/p}\leq\limsup_{n\to\infty}\frac{p}{p-1}\,\mathbf{E}\{X(T)^p\}^{1/p}.\qquad\square$$

7.2 Stochastic Integration of Simple Processes

From now on, until further notice, we let $\{B(t)\}_{t\geq0}$ is BM with $B(0)=0$, and

$\{\mathcal{F}_t\}_{t\geq 0}$ its *natural filtration* $\mathcal{F}_t \sigma(B(r): 0 \leq r \leq t)$.

Definition 7.4 *A stochastic process $\{X(t)\}_{t\geq 0}$ belongs to the class $\underline{S_T}$ of simple processes on $[0,T]$, $T\in(0,\infty]$, if for some grid $0=t_0<t_1<\ldots<t_n=T$, and for some random variables $X(0), X_{t_0},\ldots,X_{t_{n-1}}$ that are adapted to $\mathcal{F}_0,\mathcal{F}_{t_0},\ldots,\mathcal{F}_{t_{n-1}}$, respectively, such that $\mathbf{E}\{X(0)^2\}, \mathbf{E}\{X_{t_0}^2\}, \ldots, \mathbf{E}\{X_{t_{n-1}}^2\} < \infty$, and such that $X_{t_{n-1}}=0$ in the case when $T=\infty$, we have*

$$X(t) = X(0)I_{\{0\}}(t) + \sum_{i=1}^{n} X_{t_{i-1}} I_{(t_{i-1},t_i]}(t) \qquad \text{for} \quad t\in[0,T].$$

A stochastic process $\{X(t)\}_{t\geq 0}$ belongs to the class $\underline{E_T}$, $T\in(0,\infty]$, if it is measurable and adapted to $\{\mathcal{F}_t\}_{t\geq 0}$, with

$$\mathbf{E}\Big\{\int_0^T X(r)^2\,dr\Big\} < \infty.$$

A stochastic process $\{X(t)\}_{t\geq 0}$ belongs to the class $\underline{P_T}$ of predictable processes on $[0,T]$, $T\in(0,\infty]$, if it is measurable and adapted to $\{\mathcal{F}_t\}_{t\geq 0}$, with

$$\mathbf{P}\Big\{\int_0^T X(r)^2\,dr<\infty\Big\} = 1.$$

Notice that the meaning of "simple" process (function) is not the same as in Lebesgue integration, and rather connects to the stepfunctions of Riemann integration.

The values $X(0)$ and X_{t_0} of a simple process $X\in S_T$ are constants, because $\mathcal{F}_0 = \mathcal{F}_{t_0} = \sigma(B(0)) = \sigma(0) = \{\emptyset,\Omega\}$, and only constants are measurable wrt. $\{\emptyset,\Omega\}$.

Definition 7.5 *The Itô integral process $\{\int_0^t X\,dB\}_{t\in[0,T]}$ of $X\in S_T$ is defined*

$$\int_0^t X(r)\,dB(r) = \sum_{i=1}^{m} X_{t_{i-1}}(B(t_i)-B(t_{i-1})) + X_{t_m}(B(t)-B(t_m)) \qquad \text{for} \quad t\in(t_m,t_{m+1}]$$

and $\int_0^t X\,dB=0$ for $t=0$. The Itô integral $\int_s^t X\,dB$ of $X\in S_T$ is defined

$$\int_s^t X(r)\,dB(r) \equiv \int_0^t X(r)\,dB(r) - \int_0^s X(r)\,dB(r) \qquad \text{for} \quad s,t\in[0,T].$$

Remark 7.6 Assume that, for a simpel $X\in S_T$, we have two representations

$$X(t) = X(0)I_{\{0\}}(t) + \sum_{i=1}^{n} X_{t_{i-1}} I_{(t_{i-1},t_i]}(t) = X(0)I_{\{0\}}(t) + \sum_{i=1}^{n'} X'_{t'_{i-1}} I_{(t'_{i-1},t'_i]}(t)$$

for $t\in[0,T]$. By introducing a third grid that contains all times of the grids $0=t_0<t_1<\ldots<t_n=T$ and $0=t'_0<t'_1<\ldots<t'_{n'}=T$, it is easy to see that the values of the Itô integral process $\{\int_0^t X\,dB\}_{t\in[0,T]}$ are the same for both representations. #

Exercise 56 Explain why the definition of $\int_s^t X\,dB$ is consistent.

We are interested in finding (defining, in fact) a solution

$$X(t) = X_0 + \int_0^t X(r)\,dB(r) \qquad \text{for} \quad t \in [0,T]$$

to the SDE (see Lecture 6)

$$dX(t) = X(t)\,dB(t) \qquad \text{for} \quad t \in (0,T], \qquad X(0) = X_0.$$

To find such a solution, we first have to define the meaning of the stochastic integral $\{\int_0^t X\,dB\}_{t\in[0,T]}$, and then find a process X that satisfies the equation.

Exercise 57 For $X, Y \in S_T$ and constants $a, b \in \mathbb{R}$, we have

$$\int_0^t \big(aX(r) + bY(r)\big)\,dB(r) = a\int_0^t X(r)\,dB(r) + b\int_0^t Y(r)\,dB(r) \qquad \text{for} \quad t \in [0,T].$$

Exercise 58 For $X \in S_T$ and $s, t \in [0,T]$, we have

$$\int_0^t X(r)\,dB(r) = \int_0^s X(r)\,dB(r) + \int_s^t X(r)\,dB(r).$$

Exercise 59 For $X \in S_T$, $\int_0^t X\,dB$ is a continuous function of $t \in [0,T]$.

Theorem 7.7 *For $X \in S_T$, $\{\int_0^t X\,dB, \mathcal{F}_t\}_{t\in[0,T]}$ is a martingale.*

Proof. By Definitons 7.4 and 7.5, $\{\int_0^t X\,dB\}_{t\in[0,T]}$ is adapted to $\{\mathcal{F}_t\}_{t\in[0,T]}$.

To show that $\mathbf{E}\{\int_0^t X\,dB \mid \mathcal{F}_s\} = \int_0^s X\,dB$ for $0 \le s < t \le T$, we can assume that $t_j = s$ and $t_{m+1} = t$ for some $t_j < t_{m+1}$ belonging to the grid $0 = t_0 < t_1 < \ldots < t_n = T$ (cf. Definition 7.4). Because otherwise the grid can be "enriched" (with the times s and/or t), without affecting the values of the process X or its Itô integral process (cf. Remark 7.6). Using that $B(t_i) - B(t_{i-1})$ is independent of $\mathcal{F}_{t_{i-1}}$, this gives

$$\begin{aligned}
&\mathbf{E}\Big\{\int_0^t X(r)\,dB(r)\,\Big|\,\mathcal{F}_s\Big\} \\
&= \mathbf{E}\Big\{\sum_{i=1}^{m+1} X_{t_{i-1}}(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} \\
&= \mathbf{E}\Big\{\sum_{i=1}^{j} X_{t_{i-1}}(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} + \mathbf{E}\Big\{\sum_{i=j+1}^{m+1} X_{t_{i-1}}(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_s\Big\} \\
&= \sum_{i=1}^{j} X_{t_{i-1}}(B(t_i) - B(t_{i-1})) + \sum_{i=j+1}^{m+1} \mathbf{E}\Big\{\mathbf{E}\Big\{X_{t_{i-1}}(B(t_i) - B(t_{i-1}))\,\Big|\,\mathcal{F}_{t_{i-1}}\Big\}\,\Big|\,\mathcal{F}_s\Big\} \\
&= \int_0^s X(r)\,dB(r) + \sum_{i=j+1}^{m+1} \mathbf{E}\Big\{X_{t_{i-1}}\mathbf{E}\Big\{B(t_i) - B(t_{i-1})\,\Big|\,\mathcal{F}_{t_{i-1}}\Big\}\,\Big|\,\mathcal{F}_s\Big\} \\
&= \int_0^s X(r)\,dB(r) + 0. \qquad \square
\end{aligned}$$

EXERCISE 60 In the proof of Theorem 7.7, we did not show that $\mathbf{E}\{|\int_0^t X\,dB|\} < \infty$ for $t \in [0,T]$, which is of course required for it to be a martingale (and for the calculations in the proof of Theorem 7.7 to make sense). For this in turn, by the definition of the Itô integral process for $X \in S_T$, it is enough to check that

$$\mathbf{E}\Big\{|X_{t_{i-1}}|\,|B(t)-B(t_{i-1})|\Big\} < \infty \qquad \text{for} \quad t \in (t_{i-1}, t_i] \quad \text{and} \quad i = 1, \ldots, n$$

(with the notation from Definition 7.4). Do this!

EXERCISE 61 $\boxed{S_T \subseteq E_T \subseteq P_T}$

Theorem 7.8 *For $X, Y \in S_T$ and $t \in [0,T]$, we have*

$$\mathbf{E}\Big\{\Big(\int_0^t X(r)\,dB(r)\Big)\Big(\int_0^t Y(\hat{r})\,dB(\hat{r})\Big)\Big\} = \int_0^t \mathbf{E}\{X(r)Y(r)\}\,dr = \mathbf{E}\Big\{\int_0^t X(r)Y(r)\,dr\Big\}.$$

Proof. We can assume that X and Y have a common grid $0 = t_0 < t_1 < \ldots < t_n = T$, and that $t = t_{m+1}$ belongs to that grid. Because otherwise we can switch to representations of X and Y by means of a third grid that contains all the times of the original grids of X and Y, together with the time t, without affecting the values of the processes X and Y, or their Itô integral processes (cf. Remark 7.6). Since $X_{t_{i-1}}Y_{t_{j-1}}(B(t_i)-B(t_{i-1}))$ is independent of $B(t_j)-B(t_{j-1})$ for $i<j$, this gives

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\int_0^t X(r)\,dB(r)\Big)\Big(\int_0^t Y(\hat{r})\,dB(\hat{r})\Big)\Big\} \\
&\quad = \mathbf{E}\Big\{\Big(\sum_{i=1}^{m+1} X_{t_{i-1}}(B(t_i)-B(t_{i-1}))\Big)\Big(\sum_{j=1}^{m+1} Y_{t_{j-1}}(B(t_j)-B(t_{j-1}))\Big)\Big\} \\
&\quad = \mathbf{E}\Big\{\sum_{i=1}^{m+1} X_{t_{i-1}}Y_{t_{i-1}}(B(t_i)-B(t_{i-1}))^2\Big\} \\
&\qquad\qquad + \mathbf{E}\Big\{\sum_{1\le i,j\le m+1,\, i\neq j} X_{t_{i-1}}Y_{t_{j-1}}(B(t_i)-B(t_{i-1}))\,(B(t_j)-B(t_{j-1}))\Big\} \\
&\quad = \sum_{i=1}^{m+1} \mathbf{E}\{X_{t_{i-1}}Y_{t_{i-1}}\}\,\mathbf{E}\Big\{(B(t_i)-B(t_{i-1}))^2\Big\} \\
&\qquad\qquad + \sum_{1\le i<j\le m+1} \mathbf{E}\Big\{B(t_j)-B(t_{j-1})\Big\}\,\mathbf{E}\Big\{X_{t_{i-1}}Y_{t_{j-1}}(B(t_i)-B(t_{i-1}))\Big\} \\
&\qquad\qquad + \sum_{1\le j<i\le m+1} \mathbf{E}\Big\{B(t_i)-B(t_{i-1})\Big\}\,\mathbf{E}\Big\{X_{t_{i-1}}Y_{t_{j-1}}(B(t_j)-B(t_{j-1}))\Big\} \\
&\quad = \sum_{i=1}^{m+1} \mathbf{E}\{X_{t_{i-1}}Y_{t_{i-1}}\}\,(t_i - t_{i-1}) + 0 + 0 \\
&\quad = \mathbf{E}\Big\{\sum_{i=1}^{m+1} X_{t_{i-1}}Y_{t_{i-1}} \int_0^t I_{(t_{i-1},t_i]}(r)\,dr\Big\} = \sum_{i=1}^{m+1}\int_0^t \mathbf{E}\Big\{X_{t_{i-1}}Y_{t_{i-1}}I_{(t_{i-1},t_i]}(r)\Big\}\,dr
\end{aligned}$$

$$= \mathbf{E}\Big\{\int_0^t \sum_{i=1}^{m+1} X_{t_{i-1}} Y_{t_{i-1}} I_{(t_{i-1},t_i]}(r)\, dr\Big\} = \int_0^t \mathbf{E}\Big\{\sum_{i=1}^{m+1} X_{t_{i-1}} Y_{t_{i-1}} I_{(t_{i-1},t_i]}(r)\Big\}\, dr$$

$$= \mathbf{E}\Big\{\int_0^t X(r)Y(r)\, dr\Big\} = \int_0^t \mathbf{E}\{X(r)Y(r)\}\, dr$$

by *Fubini's Theorem*. Here we made use of the fact that

$$\mathbf{E}\Big\{\Big|X_{t_{i-1}} Y_{t_{j-1}} (B(t_i)-B(t_{i-1}))\,(B(t_j)-B(t_{j-1}))\Big|\Big\}$$

$$\leq \frac{1}{2}\,\mathbf{E}\Big\{\Big(X_{t_{i-1}}(B(t_i)-B(t_{i-1}))\Big)^2\Big\} + \frac{1}{2}\,\mathbf{E}\Big\{\Big(Y_{t_{j-1}}(B(t_j)-B(t_{j-1}))\Big)^2\Big\}$$

$$= \frac{1}{2}\,\mathbf{E}\Big\{X_{t_{i-1}}^2\Big\}\,\mathbf{E}\Big\{(B(t_i)-B(t_{i-1}))^2\Big\} + \frac{1}{2}\,\mathbf{E}\Big\{Y_{t_{j-1}}^2\Big\}\,\mathbf{E}\Big\{(B(t_j)-B(t_{j-1}))^2\Big\} < \infty. \quad \square$$

EXERCISE 62 For $X \in S_T$, we have $\mathbf{E}\{\int_0^t X\, dB\} = 0$ for $t \in [0,T]$.

8 Eigth Lecture

8.1 Stochastic Integration of Processes in E_T

Theorem 9.1 states that, for $X \in E_T$, we have

$$\lim_{n\to\infty} \mathbf{E}\Big\{\int_0^T \big(X_n(r) - X(r)\big)^2\, dr\Big\} = 0 \qquad \text{for some sequence} \quad \{X_n\}_{n=1}^\infty \subseteq S_T. \tag{8.1}$$

Definition 8.1 *The Itô integral process* $\{\int_0^t X\, dB\}_{t\in[0,T]}$, *of* $X \in E_T$, *is defined in the sense of convergence in mean-square* (l.i.m.),

$$\int_0^t X(r)\, dB(r) = \operatorname*{l.i.m.}_{n\to\infty} \int_0^t X_n(r)\, dB(r) \qquad \textit{where} \quad \{X_n\}_{n=1}^\infty \subseteq S_T \quad \textit{satisfies } (8.1).$$

Theorem 8.2 *The Itô integral process* $\int_0^t X\, dB$ *of* $X \in E_T$ *is well-defined.*

Proof. Notice that the difference between two simple processes is simple (recall Remark 7.6). The mean-square limit in Definition 8.1 exists, since, by Theorem 7.8,

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\int_0^t X_m(r)\, dB(r) - \int_0^t X_n(r)\, dB(r)\Big)^2\Big\} \\
&= \mathbf{E}\Big\{\Big(\int_0^t (X_m(r) - X_n(r))\, dB(r)\Big)^2\Big\} \\
&= \mathbf{E}\Big\{\int_0^t (X_m(r) - X_n(r))^2\, dr\Big\} \\
&\leq 2\,\mathbf{E}\Big\{\int_0^T (X_m(r) - X(r))^2\, dr\Big\} + 2\,\mathbf{E}\Big\{\int_0^T (X(r) - X_n(r))^2\, dr\Big\} \to 0 \qquad \text{as} \quad m, n \to \infty,
\end{aligned}$$

so that $\{\int_0^t X_n\, dB\}_{n=1}^\infty$ is a Cauchy sequence, in the sense of convergence in mean-square. Further, the limit in Definition 8.1 does not depend on which particular sequence $\{X_n\}_{n=1}^\infty \subseteq S_T$ we choose that satisfies (8.1), since, again by Theorem 7.8,

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\operatorname*{l.i.m.}_{n\to\infty} \int_0^t X_n(r)\, dB(r) - \operatorname*{l.i.m.}_{n\to\infty} \int_0^t X_n'(r)\, dB(r)\Big)^2\Big\} \\
&\quad = \mathbf{E}\Big\{\Big(\operatorname*{l.i.m.}_{n\to\infty} \int_0^t (X_n(r) - X_n'(r))\, dB(r)\Big)^2\Big\} \\
&\quad = \lim_{n\to\infty} \mathbf{E}\Big\{\Big(\int_0^t (X_n(r) - X_n'(r))\, dB(r)\Big)^2\Big\} \\
&\quad = \lim_{n\to\infty} \mathbf{E}\Big\{\int_0^t (X_n(r) - X_n'(r))^2\, dr\Big\} \\
&\quad \leq \lim_{n\to\infty} 2\,\mathbf{E}\Big\{\int_0^T (X_n(r) - X(r))^2\, dr\Big\} + 2\,\mathbf{E}\Big\{\int_0^T (X(r) - X_n'(r))^2\, dr\Big\} = 0
\end{aligned}$$

when $\{X_n\}_{n=1}^\infty$, $\{X_n'\}_{n=1}^\infty \subseteq S_T$ both satisfy (8.1), so that the mean-square limits coincide. (Recall that mean-square limits commute with first and second moments.) □

Theorem 8.3 *A measurable stochastic process* $\{X(t)\}_{t\geq 0}$ *has measurable sample paths, that is,* $[0,\infty)\ni t\to X(\omega_0,t)\in\mathbb{R}$ *is a measurable function for each* $\omega_0\in\Omega$.

Lemma 8.4 *Let* $(\mathfrak{G},\mathcal{G})$ *and* $(\mathfrak{H},\mathcal{H})$ *be measurable spaces. For (the* <u>*sections*</u> *of) a set* E *in the product* σ*-algebra* $\mathcal{G}\times\mathcal{H}$ (e.g., [34, Chapter 7])*, *we have*

$$\begin{cases} E^y=\{x\in\mathfrak{G}:(x,y)\in E\}\in\mathcal{G} & \textit{for}\quad y\in\mathfrak{H} \\ E_x=\{y\in\mathfrak{H}:(x,y)\in E\}\in\mathcal{H} & \textit{for}\quad x\in\mathfrak{G}\end{cases}.$$

**Proof.* It is enough to prove the statement for E^y. Notice that the family

$$\mathcal{E}\equiv\Big\{E\in\mathcal{G}\times\mathcal{H}:E^y\in\mathcal{G}\ \text{ for all }\ y\in\mathfrak{H}\Big\}$$

is a σ-algebra, since it is easy to check the axioms of a σ-algebra. Moreover, we have

$$(G\times H)^y=\begin{cases} G & \text{if } y\in H \\ \emptyset & \text{if } y\notin H\end{cases}\in\mathcal{G}\qquad\text{for all}\quad y\in\mathfrak{H},\qquad\text{for}\quad G\times H\in\mathcal{R},$$

where $\mathcal{R}\equiv\{G\times H: G\in\mathcal{G},\, H\in\mathcal{H}\}$ are the rectangles. Hence $\mathcal{R}\subseteq\mathcal{E}$, so that $\sigma(\mathcal{R})\subseteq\sigma(\mathcal{E})=\mathcal{E}$. Since on the other hand, $\mathcal{E}\subseteq\mathcal{G}\times\mathcal{H}\equiv\sigma(\mathcal{R})$, we get $\mathcal{E}=\mathcal{G}\times\mathcal{H}$. □

Proof of Theorem 8.3. Picking an $\omega_0\in\Omega$ and a $C\in\mathcal{B}(\mathbb{R})$, Lemma 8.4 shows that

$$\Big\{t\in[0,\infty):X(\omega_0,t)\in C\Big\}=\Big\{(\omega,t)\in\Omega\times[0,\infty):X(\omega,t)\in C\Big\}_{\omega_0}\in\mathcal{B}([0,\infty)),$$

since $\{(\omega,t):X(\omega,t)\in C\}\in\mathcal{F}\times\mathcal{B}([0,\infty))$ by measurability of X. □

Definition 8.5 *A stochastic process* $\{X(t)\}_{t\geq 0}$ *[*$\{X(t)\}_{t\in[0,T]}$*] is* <u>*progressively measurable*</u>, *if*

$$X:\Omega\times[0,t]\to\mathbb{R}\qquad\textit{is}\quad \mathcal{F}_t\times\mathcal{B}([0,t])\textit{-measurable}\quad\textit{for each}\quad t\geq 0\quad[0\leq t\leq T].$$

[Recall that $\{\mathcal{F}_t\}_{t\geq 0}=\{\sigma(B(r):0\leq r\leq t)\}_{t\geq 0}$, where B is BM with $B(0)=0$.]

Theorem 8.6 *A progressively measurable process is measurable and adapted.*

Proof. For $C\in\mathcal{B}(\mathbb{R})$, the set

$$\Big\{(\omega,s)\in\Omega\times[0,\infty):X(\omega,s)\in C\Big\}=\bigcup_{n=1}^{\infty}\Big\{(\omega,s)\in\Omega\times[0,n]:X(\omega,s)\in C\Big\}$$

belongs to $\mathcal{F}\times\mathcal{B}([0,\infty))$, since the n'th member of the union belongs to $\mathcal{F}_n\times\mathcal{B}([0,$

$n]$), by progressive measurability. Picking a $t \geq 0$, Lemma 8.4 further shows that

$$\{\omega \in \Omega : X(\omega, t) \in C\} = \Big\{(\omega, s) \in \Omega \times [0,t] : X(\omega, s) \in C\Big\}^t \in \mathcal{F}_t,$$

since $\{(\omega, s) \in \Omega \times [0,t] : X(\omega, s) \in C\} \in \mathcal{F}_t \times \mathcal{B}([0,t])$, by progressive measurability. □

Theorem 8.6 has an important partial converse, famous for its long and difficult proof:

Theorem 8.7 ([29, pp. 68-ff.])$\Pi_{n=1}^{\infty}\star$ *A measurable and adapted process has a progressively measurable version.*

Theorem 8.8 *An adapted process that is left-continuous or right-continuous is progressively measurable.*

Proof. We only give a proof for left-continuous processes, since that for right-continuity is quite similar. Notice that

$$\Big\{(\omega, s) \in \Omega \times [0,t] : X(\omega, s) \in C\Big\} \in \mathcal{F}_t \times \mathcal{B}([0,t]) \qquad \text{for} \qquad C \in \mathcal{B}(\mathbb{R}) \tag{8.2}$$

holds for $t=0$, since (8.2) in that case amounts to check that

$$\{\omega \in \Omega : X(\omega, 0) \in C\} \times \{0\} \in \mathcal{F}_0 \times \mathcal{B}(\{0\}) = \mathcal{F}_0 \times \{\emptyset, \{0\}\},$$

which holds since X is adpated. For $t>0$, we consider the sampled process

$$X_n(s) \equiv X\Big([2^n(s/t)]/(2^n/t)\Big) = X(kt/2^n) \qquad \text{for} \quad s \in \Big[kt/2^n, (k+1)t/2^n\Big).$$

Since $[2^n(s/t)]/(2^n/t) \to s$ from the left as $n \to \infty$, for $s \in [0,t]$, and X is left-continuous, we have $X(s) = \lim_{n\to\infty} X_n(s)$ for $s \in [0,t]$. Since limits of measurable functions are measurable, in order to prove (8.2), it is enough to check that

$$\Big\{(\omega, s) \in \Omega \times [0,t] : X_n(\omega, s) \in C\Big\} \in \mathcal{F}_t \times \mathcal{B}([0,t]) \qquad \text{for} \qquad C \in \mathcal{B}(\mathbb{R}).$$

However, the set on the left-hand side can be expressed as the union

$$\bigcup_{k=0}^{2^n} \Big\{(\omega, s) \in \Omega \times \Big([0,t] \cap [kt/2^n, (k+1)t/2^n)\Big) : X_n(\omega, s) \in C\Big\}$$
$$= \bigcup_{k=0}^{2^n} \bigg(\Big\{\omega \in \Omega : X(\omega, kt/2^n) \in C\Big\} \times \Big([0,t] \cap [kt/2^n, (k+1)t/2^n)\Big)\bigg) \in \mathcal{F}_t \times \mathcal{B}([0,t]),$$

since, by adaptedness, $X(kt/2^n)$ is $\mathcal{F}_{kt/2^n}$-measurable, and thus $\mathcal{F}_t$-measurable. □

Exercise 63 Let $A \subseteq \Omega$ and $s \in \mathbb{R}$. Show that $A \times \{s\}$ is $\mathcal{F}_t \times \mathcal{B}([0,t])$-measurable iff. $A \in \mathcal{F}_t$ and $s \in [0,t]$.

9 Ninth Lecture

9.1 Stochastic Integration of Processes in E_T (continued)

The result of this lecture, Theorems 9.1 below, is [27, Lemma 4.4]. The ideas of the proof have the same origin, and are notable because Theorem 8.7 is not required. A simplification of this approach is sketched in [22, Problem 3.2.5], and worked out in [22, Section 3.8]. We follow the sketch, but work it out marginally different.

Theorem 9.1 *For a process* $X\in E_T$, *we have*

$$\lim_{n\to\infty}\mathbf{E}\Big\{\int_0^T\big(X_n(r)-X(r)\big)^2\,dr\Big\}=0\qquad\textit{for some sequence}\quad\{X_n\}_{n=1}^\infty\subseteq S_T.$$

Proof. It is enough to prove that, for each $T<\infty$ and $\varepsilon>0$, we have

$$\mathbf{E}\Big\{\int_0^T(\hat X(r)-X(r))^2\,dr\Big\}\le\varepsilon\qquad\text{for some}\quad \hat X\in S_T. \tag{9.1}$$

Let $X^{(N)}(r)=\max\{\min\{X(r),N\},-N\}$ for $N\in\mathbb{N}$, so that $X^{(N)}(r)\to X(r)$ as $N\to\infty$. Since $(X^{(N)}(r){-}X(r))^2\le X(r)^2$, which is integrable over $(\omega,r)\in\Omega{\times}[0,T]$ (since $X\in E_T$), the *Dominated Convergence Theorem* gives

$$\mathbf{E}\Big\{\int_0^T(X^{(N)}(r)-X(r))^2\,dr\Big\}\to 0\qquad\text{as}\quad N\to\infty. \tag{9.2}$$

It follows that it is enough to prove that, to each $\varepsilon>0$ and $N\in\mathbb{N}$, we have

$$\mathbf{E}\Big\{\int_0^T(\hat X(r)-X^{(N)}(r))^2\,dr\Big\}\le\frac{\varepsilon}{4}\qquad\text{for some}\quad \hat X\in S_T, \tag{9.3}$$

because, taking $N\in\mathbb{N}$ so that the left hand side of (9.2) is at most $\varepsilon/4$, (9.3) gives

$$\begin{aligned}&\mathbf{E}\Big\{\int_0^T(\hat X(r)-X(r))^2\,dr\Big\}\\ &\qquad\le 2\,\mathbf{E}\Big\{\int_0^T(\hat X(r)-X^{(N)}(r))^2\,dr\Big\}+2\,\mathbf{E}\Big\{\int_0^T(X^{(N)}(r)-X(r))^2\,dr\Big\}\le\varepsilon.\end{aligned}$$

Now recall that $\lfloor x\rfloor=k$ for $x\in(k,k{+}1]$, and define $\{X_n^{(N,\tau)}(r)\}_{r\in[0,T]}$ by

$$X_n^{(N,\tau)}(r)\equiv\begin{cases}X^{(N)}(2^{-n}\lfloor 2^n(r-\tau)\rfloor+\tau) & \text{for}\quad 2^{-n}\lfloor 2^n(r-\tau)\rfloor+\tau\in[0,T]\\ 0 & \text{for}\quad 2^{-n}\lfloor 2^n(r-\tau)\rfloor+\tau\notin[0,T]\end{cases},$$

where $\tau\in[0,1]$ is a constant to be specified later. Observe that

$$X_n^{(N,\tau)}(r)=\begin{cases}X^{(N)}(2^{-n}k+\tau) & \text{for}\quad r-\tau\in(2^{-n}k,2^{-n}(k{+}1)]\quad\text{and}\quad 2^{-n}k+\tau\ge 0\\ 0 & \text{for}\quad r-\tau\in(2^{-n}k,2^{-n}(k{+}1)]\quad\text{and}\quad 2^{-n}k+\tau<0\end{cases},$$

where it is necessary that $r\le 2^{-n}$ for the lower option to take place. This gives that $X_n^{(N,\tau)}\in S_T$. In order to prove (9.3), it is therefore enough to show that

$$\liminf_{n\to\infty} \mathbf{E}\Big\{\int_0^T (X_n^{(N,\tau)}(r) - X^{(N)}(r))^2\,dr\Big\} = 0 \qquad \text{for some} \quad \tau\in[0,1], \tag{9.4}$$

because then there exists an $n\in\mathbb{N}$ such that we have the desired

$$\mathbf{E}\Big\{\int_0^T (X_n^{(N,\tau)}(r) - X^{(N)}(r))^2\,dr\Big\} \le \frac{\varepsilon}{4}.$$

To prove (9.4), it is enough to show that

$$\lim_{n\to\infty} \mathbf{E}\Big\{\int_{\tau=0}^{\tau=1}\int_{r=0}^{r=T} (X_n^{(N,\tau)}(r) - X^{(N)}(r))^2\,dr d\tau\Big\} = 0. \tag{9.5}$$

This is so, because if (9.4) does not hold, then *Fatou's Lemma* together with *Fubibi's Theorem* show that (9.5) does not hold, in the following way

$$\begin{aligned} 0 &< \int_{\tau=0}^{\tau=1} \liminf_{n\to\infty} \mathbf{E}\Big\{\int_{r=0}^{r=T} (X_n^{(N,\tau)}(r) - X^{(N)}(r))^2\,dr\Big\}\,d\tau \\ &\le \liminf_{n\to\infty} \mathbf{E}\Big\{\int_{\tau=0}^{\tau=1}\int_{r=0}^{r=T} (X_n^{(N,\tau)}(r) - X^{(N)}(r))^2\,dr d\tau\Big\}. \end{aligned}$$

To prove (9.5), it is enough to show that

$$\lim_{h\downarrow 0} \mathbf{E}\Big\{\int_h^T (X^{(N)}(r) - X^{(N)}(r-h))^2\,dr\Big\} = 0. \tag{9.6}$$

This is so, because if (9.6) holds, then the expectation in (9.5) is at most

$$\begin{aligned} &\mathbf{E}\Big\{\int_{r=2^{-n}}^{r=T} \sum_{\{k\,:\,2^{-n}\le r-2^{-n}k\le 1\}} \int_{\tau\in[r-2^{-n}(k+1),\,r-2^{-n}k)} (X^{(N)}(2^{-n}k+\tau) - X^{(N)}(r))^2\,dr d\tau\Big\} \\ &\qquad\qquad + \mathbf{E}\Big\{\int_{r=2^{-n}}^{r=T}\int_{\tau\in[0,2^{-n})\cup(1-2^{-n},1]} (X_n^{(N,\tau)}(r) - X^{(N)}(r))^2\,dr d\tau\Big\} \\ &\qquad\qquad + \mathbf{E}\Big\{\int_{r=0}^{r=2^{-n}}\int_{\tau=0}^{\tau=1} X^{(N)}(r)^2\,dr d\tau\Big\} \\ &\le 2^n\,\mathbf{E}\Big\{\int_{r=2^{-n}}^{r=T}\int_{\tilde\tau=0}^{\tilde\tau=2^{-n}} (X^{(N)}(r-\tilde\tau) - X^{(N)}(r))^2\,dr d\tau\Big\} \\ &\qquad\qquad + (T-2^{-n})\,(2^{-n}+2^{-n})\,(2N)^2 \;+\; 2^{-n}N^2 \\ &\le \sup_{h\in[0,2^{-n}]} \mathbf{E}\Big\{\int_h^T (X^{(N)}(r-h) - X^{(N)}(r))^2\,dr\Big\} + (T2^{3-n}+2^{-n})\,N^2 \to 0 \qquad \text{as} \quad n\to\infty. \end{aligned}$$

Let $C_T^{(N)}$ be the class of stochastic processes $\{Z(r)\}_{r\in[0,T]}$ that are continuous with $\sup_{t\in[0,T]}|Z(t)| \le N$. To prove (9.6), it is enough to show that, to each $\varepsilon>0$, we have

$$\mathbf{E}\Big\{\int_0^T (X^{(N)}(r) - Z(r))^2\,dr\Big\} \le \varepsilon \qquad \text{for some} \quad Z\in C_T^{(N)}. \tag{9.7}$$

This is so, because if (9.7) holds, then we have [using that $(x+y+z)^2 \le 3x^2+3y^2+3z^2$]

$$\begin{aligned} &\mathbf{E}\Big\{\int_h^T (X^{(N)}(r) - X^{(N)}(r-h))^2\,dr\Big\} \\ &\qquad \le 3\,\mathbf{E}\Big\{\int_h^T (X^{(N)}(r) - Z(r))^2\,dr\Big\} + 3\,\mathbf{E}\Big\{\int_h^T (Z(r) - Z(r-h))^2\,dr\Big\} \\ &\qquad\qquad + 3\,\mathbf{E}\Big\{\int_h^T (Z(r-h) - X^{(N)}(r-h))^2\,dr\Big\} \end{aligned}$$

$$\leq 3\,\varepsilon + 3T\,\mathbf{E}\Big\{\sup_{r\in[h,T]} (Z(r)-Z(r-h))^2\Big\} + 3\,\varepsilon \to 6\,\varepsilon \qquad \text{as} \quad h\downarrow 0,$$

by (uniform) continuity together with the *Dominated Convergence Theorem*. Since this holds for each $\varepsilon>0$, the limit in (9.6) must be zero.

To prove (9.7), define

$$Y(t) \equiv \int_0^t X^{(N)}(r)\,dr \qquad \text{for} \quad t\in[0,T], \tag{9.8}$$

and (with the notation $t^+ = \max\{t,0\}$)

$$Z_n(t) \equiv n\int_{(t-1/n)^+\wedge T}^{t^+\wedge T} X^{(N)}(r)\,dr = n\Big(Y(t^+\wedge T) - Y((t-1/n)^+\wedge T)\Big) \qquad \text{for} \quad t\in[0,T].$$

Notice that $|Z_n(t)|\leq N$, since the integral is over an interval of length at most $1/n$, and $|nX^{(N)}(r)|\leq nN$. Further, Z_n is continuous, so that $Z_n\in C_T^{(N)}$, since

$$Z_n(t+\varepsilon) - Z_n(t) = n\int_{t^+\wedge T}^{(t+\varepsilon)^+\wedge T} X^{(N)}(r)\,dr - n\int_{(t-1/n)^+\wedge T}^{(t+\varepsilon-1/n)^+\wedge T} X^{(N)}(r)\,dr \to 0$$

as $\varepsilon\to 0$, because the integrals are over intervals of length at most ε [and $nX^{(N)}$ is bounded]. Obviously, Y is absolutely continuous (cf. Example 3.1), with derivative

$$Y'(t) = \lim_{n\to\infty} Z_n(t) = X^{(N)}(t) \qquad \text{for almost all} \quad t\in[0,T].$$

Since $(X^{(N)}(t)-Z_n(t))^2 \leq 4N^2$, the *Dominated Convergence Theorem* thus gives

$$\lim_{n\to\infty} \mathbf{E}\Big\{\int_0^T (X^{(N)}(t)-Z_n(t))^2\,dt\Big\} = 0.$$

Picking a sufficiently large $n_0\in\mathbb{N}$, it follows that $Z=Z_{n_0}$ satisfies (9.7). □

Exercise 64 Explain why $Y(t) = \int_0^t X(r)\,dr$ is adapted if X is progressively measurable (and the integral well-defined). Explain why this possibly could fail if X is not progressively measurable. (This exercise is solved in Appendix C.)

Exercise 65 Give a direct proof of (9.3), in the case when X is continuous and adapted, so that $X^{(N)}$ is continuous, bounded and adapted.

Exercise 66 Show that if the processes $\{Y(t)\}_{t\in[0,T]}$ and $\{\tilde{Y}(t)\}_{t\in[0,T]}$ are versions of each other, then adaptedness for one of them, to an augmented filtration $\{\mathcal{F}_t\}_{t\geq 0}$, implies adaptedness for the other. (**Hint:** See the proof of Theorem 11.6.)

Lemma 9.2 *Let $\{X(t)\}_{t\in[0,T]}$ and $\{\tilde{X}(t)\}_{t\in[0,T]}$ be measurable stochastic processes that are versions of each other. We have*

$$\mathbf{P}\Big\{\int_0^t X(r)\,dr \neq \int_0^t \tilde{X}(r)\,dr\Big\} = 0 \qquad \textit{for} \quad t\in[0,T],$$

provided that the integrals are well-defined (but not necessarily finite).

Proof. Clearly, if $X(r)=\tilde{X}(r)$ for almost all $r\in[0,t)$, then we have $\int_0^t X(r)\,dr = \int_0^t \tilde{X}(r)\,dr$. It follows that

$$\mathbf{P}\Big\{\int_0^t X(r)\,dr \neq \int_0^t \tilde{X}(r)\,dr\Big\} \leq \mathbf{P}\Big\{\int_0^t I_{\{X(r)\neq\tilde{X}(r)\}}\,dr>0\Big\}. \tag{9.9}$$

If the probability on the right-hand side of (9.9) is non-zero, then the expected value of $\int_0^t I_{\{X(r)\neq\tilde{X}(r)\}}\,dr$ is non-zero, so that, by *Fubini's Theorem*,

$$0<\mathbf{E}\Big\{\int_0^t I_{\{X(r)\neq\tilde{X}(r)\}}\,dr>0\Big\}=\int_0^t \mathbf{E}\{I_{\{X(r)\neq\tilde{X}(r)\}}\,dr=\int_0^t \mathbf{P}\{X(r)\neq\tilde{X}(r)\}\,dr=0,$$

since X and $\tilde{X}$ are versions of each other. This is a contradiction, so that the probability on the right-hand side of (9.9) must be zero. This establishes the lemma. □

We now show how the proof of Theorem 9.1 can be somewhat shortened, if one assumes that $\{\mathcal{F}_t\}_{t\geq 0}$ is augmented, and uses Theorem 8.7.

**Second Proof of Theorem 9.1* (after [22, pp. 133-134], see also [27, p. 95]). By the first paragraph of the proof of Theorem 9.1, it is enough to prove (9.3). By the last paragraph, we have the approximation (9.7) of $X^{(N)}$, for some $Z\in C_T^{(N)}$. Since Z is continuous and bounded, Z can be approximated by a simple $\hat{X}\in S_T$, in the sense (9.3), by Exercise 65, thereby proving (9.3) also for $X^{(N)}$ [by (9.7)], if Z is adapted.

To prove that Z is adapted, it is enough to prove that Y in (9.8) is. Let $\tilde{X}^{(N)}$ be a progressively measurable version of $X^{(N)}$ (recall Theorem 8.7), and put $\tilde{Y}(t)=\int_0^t \tilde{X}^{(N)}(r)\,dr$. Since $\tilde{Y}$ is adapted, by Exercise 64, it is enough to prove that Y and $\tilde{Y}$ are versions of each other, by Exercise 66. This follows from Lemma 9.2. □

10.1 Properties of Stochastic Integrals in E_T

Definition 10.1 *For $X \in E_T$ we define the Itô integral process*

$$\int_s^t X(r)\,dB(r) \equiv \int_0^t X(r)\,dB(r) - \int_0^s X(r)\,dB(r) \qquad \textit{for} \quad s,t \in [0,T]$$

(where the integrals on the right-hand side are defined in Definition 8.1).

Theorem 10.2 *For the Itô integral process of $X, Y \in E_T$, we have*

(1) $\int_0^t (aX+bY)\,dB = a\int_0^t X\,dB + b\int_0^t Y\,dB$ *for* $a, b \in \mathbb{R}$*;*

(2) $\int_0^t X\,dB = \int_0^s X\,dB + \int_s^t X\,dB$*;*

(3) $\mathbf{E}\{\int_0^t X\,dB\} = 0$*;*

(4) $\mathbf{E}\Big\{(\int_0^t X\,dB)\,(\int_0^t Y\,dB)\Big\} = \mathbf{E}\Big\{\int_0^t X(r)Y(r)\,dr\Big\}$*;*

(5) $\{\int_0^t X\,dB, \mathcal{F}_t\}_{t\in[0,T]}$ *is a square-integrable martingale;*

(6) $\{\int_0^t X\,dB\}_{t\in[0,T]}$ *is continuous and progressively measurable, with probability one.*

Proof. (1) Taking $\{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty \subseteq S_T$ such that

$$\mathbf{E}\Big\{\int_0^T \big(X(r)-X_n(r)\big)^2 dr\Big\} \to 0 \qquad \text{and} \qquad \mathbf{E}\Big\{\int_0^T \big(Y(r)-Y_n(r)\big)^2 dr\Big\} \to 0 \tag{10.1}$$

as $n \to \infty$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{E}\Big\{\int_0^T \Big((aX(r)+bY(r)) - (aX_n(r)+bY_n(r))\Big)^2 dr\Big\} \\
&\leq 2a^2 \limsup_{n\to\infty} \mathbf{E}\Big\{\int_0^T \big(X(r)-X_n(r)\big)^2 dr\Big\} + 2b^2 \limsup_{n\to\infty} \mathbf{E}\Big\{\int_0^T \big(Y(r)-Y_n(r)\big)^2 ds\Big\} = 0.
\end{aligned}$$

Since $\{aX_n+bY_n\}_{n=1}^\infty \subseteq S_T$, we have (recall Definition 8.1 and Exercise 57)

$$\begin{aligned}
\int_0^t \Big((aX(r)+bY(r)\Big)\,dB(r) &= \operatorname*{l.i.m.}_{n\to\infty} \int_0^t \Big((aX_n(r)+bY_n(r)\Big)\,dB(r) \\
&= \operatorname*{l.i.m.}_{n\to\infty} \Big(a\int_0^t X_n(r)\,dB(r) + b\int_0^t Y_n(r)\,dB(r)\Big) \\
&= a \operatorname*{l.i.m.}_{n\to\infty} \int_0^t X_n(r)\,dB(r) + b \operatorname*{l.i.m.}_{n\to\infty} \int_0^t Y_n(r)\,dB(r) \\
&= a\int_0^t X(r)\,dB(r) + b\int_0^t Y(r)\,dB(r).
\end{aligned}$$

(2) This is only a rearrangement of Definition 10.1.

(3) Taking $\{X_n\}_{n=1}^\infty \subseteq S_T$ as in (10.1), we have, by Exercise 62,

$$\mathbf{E}\Big\{\int_0^t X(r)\,dB(r)\Big\} = \mathbf{E}\Big\{\operatorname*{l.i.m.}_{n\to\infty}\int_0^t X_n(r)\,dB(r)\Big\} = \lim_{n\to\infty} \mathbf{E}\Big\{\int_0^t X_n(r)\,dB(r)\Big\} = 0$$

(since mean-square limits commute with expectations).

(4) Taking $\{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty \subseteq S_T$ as in (10.1), we have, by Theorem 7.8,

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\int_0^t X(r)\,dB(r)\Big)\Big(\int_0^t Y(s)\,dB(s)\Big)\Big\} - \mathbf{E}\Big\{\int_0^t X(r)Y(r)\,dr\Big\} \\
&= \mathbf{E}\Big\{\Big(\underset{n\to\infty}{\text{l.i.m.}}\int_0^t X_n(r)\,dB(r)\Big)\Big(\underset{n\to\infty}{\text{l.i.m.}}\int_0^t Y_n(s)\,dB(s)\Big)\Big\} - \mathbf{E}\Big\{\int_0^t X(r)Y(r)\,dr\Big\} \\
&= \lim_{n\to\infty}\mathbf{E}\Big\{\Big(\int_0^t X_n(r)\,dB(r)\Big)\Big(\int_0^t Y_n(s)\,dB(s)\Big)\Big\} - \mathbf{E}\Big\{\int_0^t X(r)Y(r)\,dr\Big\} \\
&= \lim_{n\to\infty}\mathbf{E}\Big\{\int_0^t X_n(r)Y_n(r)\,dr\Big\} - \mathbf{E}\Big\{\int_0^t X(r)Y(r)\,dr\Big\} \\
&= \lim_{n\to\infty}\Big(\mathbf{E}\Big\{\int_0^t X(r)\,(Y_n(r)-Y(r))\,dr\Big\} + \mathbf{E}\Big\{\int_0^t (X_n(r)-X(r))\,Y(r)\,dr\Big\} \\
&\qquad\qquad + \mathbf{E}\Big\{\int_0^t (X_n(r)-X(r))\,(Y_n(r)-Y(r))\,dr\Big\}\Big)
\end{aligned}$$

(since mean-square limits commute with second moments). For the first term on the right-hand side, Cauchy-Schwarz inequality gives

$$\begin{aligned}
\mathbf{E}\Big\{\Big|\int_0^t X(r)\,(Y_n(r)-Y(r))\,dr\Big|\Big\} &\le \mathbf{E}\Big\{\sqrt{\int_0^t X(r)^2\,dr}\,\sqrt{\int_0^t (Y_n(s)-Y(s))^2\,ds}\Big\} \\
&\le \sqrt{\mathbf{E}\Big\{\int_0^t X(r)^2\,dr\Big\}\,\mathbf{E}\Big\{\int_0^t (Y_n(r)-Y(r))^2\,dr\Big\}} \to 0
\end{aligned}$$

as $n\to\infty$, and similarly the other two terms on the right-hand side go to zero. □

(5) Since the Itô integral $\int_0^t X\,dB$ is a mean-square limit of adapted random variables, it is adapted by Exercise 67 below. Taking $\{X_n\}_{n=1}^\infty \subseteq S_T$ as in (10.1), Theorem 7.7 together with Exercise 68 below further show that

$$\begin{aligned}
\mathbf{E}\Big\{\int_0^t X(r)\,dB(r)\,\Big|\,\mathcal{F}_s\Big\} &= \mathbf{E}\Big\{\underset{n\to\infty}{\text{l.i.m.}}\int_0^t X_n(r)\,dB(r)\,\Big|\,\mathcal{F}_s\Big\} \\
&= \underset{n\to\infty}{\text{l.i.m.}}\,\mathbf{E}\Big\{\int_0^t X_n(r)\,dB(r)\,\Big|\,\mathcal{F}_s\Big\} \\
&= \underset{n\to\infty}{\text{l.i.m.}}\int_0^s X_n(r)\,dB(r) = \int_0^s X(r)\,dB(r).
\end{aligned}$$

(6) By (5) together with Theorem 8.8, it is enough to prove continuity: We have

$$\mathbf{E}\Big\{\int_0^T \big(X(r)-X_n(r)\big)^2\,dr\Big\} \le \frac{1}{3}\,2^{-n} \qquad \text{for} \quad n\in\mathbb{N}, \quad \text{for some} \quad \{X_n\}_{n=1}^\infty \subseteq S_T$$

(by Theorem 9.1). It follows that

$$\begin{aligned}
&\mathbf{E}\Big\{\int_0^T \big(X_{i+1}(r)-X_i(r)\big)^2\,dr\Big\} \\
&\le 2\,\mathbf{E}\Big\{\int_0^T \big(X_{i+1}(r)-X(r)\big)^2\,dr\Big\} + 2\,\mathbf{E}\Big\{\int_0^T \big(X(r)-X_i(r)\big)^2\,dr\Big\} \le \frac{2^{-i}+2^{1-i}}{3} = 2^{-i}
\end{aligned}$$

for $i\in\mathbb{N}$. Notice that (by Definition 8.1)

$$\int_0^t X\,dB = \underset{n\to\infty}{\text{l.i.m.}}\int_0^t X_n\,dB = \int_0^t X_1\,dB + \underset{n\to\infty}{\text{l.i.m.}}\sum_{i=1}^{n-1}\int_0^t (X_{i+1}-X_i)\,dB.$$

Since the terms of the sum are continuous, by Exercise 59, we get the continuity desired if the sum converges uniformly with probability one, that is, if

$$\sum_{i=1}^{\infty} \sup_{t\in[0,T]} \left|\int_0^t (X_{i+1}-X_i)\, dB\right| \quad \text{converges with probability one.}$$

To that end, it is enough to show that

$$\mathbf{P}\Big\{\sup_{t\in[0,T]} \left|\int_0^t (X_{i+1}-X_i)\, dB\right| > \frac{1}{i^2} \ \text{ for infinitely many } \ i\in\mathbb{N}\Big\} = 0,$$

which in turn, by the *Borel-Cantelli Lemma*, will follow if

$$\sum_{i=1}^{\infty} \mathbf{P}\Big\{\sup_{t\in[0,T]} \left|\int_0^t (X_{i+1}-X_i)\, dB\right| > \frac{1}{i^2}\Big\} < \infty.$$

Since $\int_0^t X_{i+1}\, dB$ and $\int_0^t X_i\, dB$ are martingales (by Theorem 7.7), $\int_0^t (X_{i+1}-X_i)\, dB$ is a martingale. Hence the *Doob-Kolmogorov Inequality* and Theorem 7.8 give

$$\begin{aligned}\mathbf{P}\Big\{\sup_{t\in[0,T]} \left|\int_0^t (X_{i+1}-X_i)\, dB\right| > \frac{1}{i^2}\Big\} &\le \mathbf{E}\Big\{\Big(\int_0^T (X_{i+1}-X_i)\, dB\Big)^2\Big\}\Big/\Big(\frac{1}{i^2}\Big)^2 \\ &= \mathbf{E}\Big\{\int_0^T (X_{i+1}(r)-X_i(r))^2\, dr\Big\}\Big/\frac{1}{i^4} \le \frac{i^4}{2^i}. \quad \Box\end{aligned}$$

EXERCISE 67 Explain why, and in what sense, a mean-square limit, or more generally a limit in probability, of a sequence of random variables adapted to a σ-algebra $\mathcal{G}$, will be adapted to that σ-algebra.

EXERCISE 68 Let $\mathcal{G}$ be a σ-algebra, and $Z, Z_1, Z_2, \ldots$ random variables with finite means. Show that

$$\begin{cases} \lim_{n\to\infty} \mathbf{E}\{|Z_n - Z|\} = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbf{E}\Big\{\Big|\mathbf{E}\{Z_n|\mathcal{G}\} - \mathbf{E}\{Z|\mathcal{G}\}\Big|\Big\} = 0 \\ \text{l.i.m.}_{n\to\infty} Z_n = Z \quad \Rightarrow \quad \text{l.i.m.}_{n\to\infty} \mathbf{E}\{Z_n|\mathcal{G}\} = \mathbf{E}\{Z|\mathcal{G}\} \end{cases}.$$

10.2 Stochastic Integration of Processes in P_T

Theorem 11.1 states that, for $X \in P_T$, we have, in the sense of convergence in probability (P-lim) [convergence in measure ($\mathbb{L}^0(\mathbf{P})$-convergence in mathematics)]

$$\operatorname*{P-lim}_{n\to\infty} \int_0^T (X_n(r)-X(r))^2\, dr = 0 \qquad \text{for some sequence} \quad \{X_n\}_{n=1}^{\infty} \subseteq E_T, \tag{10.2}$$

provided that $\{\mathcal{F}_t\}_{t\ge 0}$ is augmented. Recall that (10.2) means that

$$\lim_{n\to\infty} \mathbf{P}\Big\{\Big|\int_0^T (X_n(r)-X(r))^2\, dr\Big| > \delta\Big\} = 0 \qquad \text{for each} \quad \delta > 0. \tag{10.3}$$

Theorem 11.2 states that, for $X \in E_T$ and a constant $C > 0$, we have

$$\mathbf{P}\Big\{\sup_{t\in[0,T]} \Big|\int_0^t X(r)\, dB(r)\Big| > \lambda\Big\} \le \frac{C}{\lambda^2} + \mathbf{P}\Big\{\int_0^T X(r)^2\, dr > C\Big\} \qquad \text{for} \quad \lambda > 0. \tag{10.4}$$

Definition 10.3 *Let $\{\mathcal{F}_t\}_{t\geq 0}$ be augmented* (Definition 2.20). *For $X\in P_T$, the Itô integral process $\{\int_0^t X\,dB\}_{t\in[0,T]}$ is defined, in the sense of convergence in probability,*

$$\int_0^t X(r)\,dB(r) = \operatorname*{P-lim}_{n\to\infty}\int_0^t X_n(r)\,dB(r) \qquad \textit{where} \quad \{X_n\}_{n=1}^{\infty}\subseteq E_T \quad \textit{satisfies } (10.2).$$

Theorem 10.4 *The Itô integral process $\int_0^t X\,dB$ of $X\in P_T$ is well-defined.*

Proof. To prove that the limit in probability in Definition 10.3 exists, it is enough to check the Cauchy criterion for convergence in probability

$$\lim_{m,n\to\infty}\mathbf{P}\Big\{\Big|\int_0^t X_m(r)\,dB(r)-\int_0^t X_n(r)\,dB(r)\Big|>\delta\Big\}=0 \qquad \text{for each} \quad \delta>0. \qquad (10.5)$$

To that end, pick an $\varepsilon>0$, and notice that (10.3) and (10.4) give

$$\begin{aligned}
&\limsup_{m,n\to\infty}\mathbf{P}\Big\{\Big|\int_0^t X_m\,dB-\int_0^t X_n\,dB\Big|>\delta\Big\}\\
&\leq \frac{\varepsilon}{\delta^2}+\limsup_{m,n\to\infty}\mathbf{P}\Big\{\int_0^T (X_m(r)-X_n(r))^2\,dr>\varepsilon\Big\}\\
&\leq \frac{\varepsilon}{\delta^2}+\limsup_{m,n\to\infty}\mathbf{P}\Big\{2\int_0^T (X_m(r)-X(r))^2\,dr+2\int_0^T (X(r)-X_n(r))^2\,dr>\varepsilon\Big\}\\
&\leq \frac{\varepsilon}{\delta^2}+\limsup_{m,n\to\infty}\mathbf{P}\Big\{\Big\{\int_0^T (X_m(r)-X(r))^2\,dr>\frac{\varepsilon}{4}\Big\}\cup\Big\{\int_0^T (X(r)-X_n(r))^2\,dr>\frac{\varepsilon}{4}\Big\}\Big\}\\
&\leq \frac{\varepsilon}{\delta^2}+\limsup_{m,n\to\infty}\mathbf{P}\Big\{\int_0^T (X_m(r)-X(r))^2\,dr>\frac{\varepsilon}{4}\Big\}+\limsup_{m,n\to\infty}\mathbf{P}\Big\{\int_0^T (X(r)-X_n(r))^2\,dr>\frac{\varepsilon}{4}\Big\}\\
&=\varepsilon/\delta^2+0+0 \qquad \text{for} \quad t\in[0,T].
\end{aligned}$$

Since the left-hand side does not depend on ε, and the established inequality holds for every $\varepsilon>0$, the left-hand side must be zero, which is precisely (10.5).

The limit in Definition 8.1 does not depend on what particular sequence $\{X_n\}_{n=1}^{\infty}\subseteq E_T$ we choose that satisfies (10.2). This is so, because if $\{X_n'\}_{n=1}^{\infty}\subseteq E_T$ is another sequence that satisfies (10.2), and $\delta,\varepsilon>0$ are constants, then (10.3) and (10.4) give

$$\begin{aligned}
&\mathbf{P}\Big\{\Big|\operatorname*{P-lim}_{n\to\infty}\int_0^t X_n\,dB-\operatorname*{P-lim}_{n\to\infty}\int_0^t X_n'\,dB\Big|>\delta\Big\}\\
&\leq \mathbf{P}\Big\{\Big|\operatorname*{P-lim}_{n\to\infty}\int_0^t X_n\,dB-\int_0^t X_k\,dB\Big|+\Big|\int_0^t (X_k-X_k')\,dB\Big|+\Big|\int_0^t X_k'\,dB-\operatorname*{P-lim}_{n\to\infty}\int_0^t X_n'\,dB\Big|>\delta\Big\}\\
&\leq \mathbf{P}\Big\{\Big|\operatorname*{P-lim}_{n\to\infty}\int_0^t X_n\,dB-\int_0^t X_k\,dB\Big|>\frac{\delta}{3}\Big\}+\mathbf{P}\Big\{\Big|\int_0^t X_k'\,dB-\operatorname*{P-lim}_{n\to\infty}\int_0^t X_n'\,dB\Big|>\frac{\delta}{3}\Big\}\\
&\qquad\qquad+\mathbf{P}\Big\{\Big|\int_0^t (X_k-X_k')\,dB\Big|>\frac{\delta}{3}\Big\}\\
&\leq o(1)+o(1)+\frac{\varepsilon}{(\delta/3)^2}+\mathbf{P}\Big\{\int_0^T (X_k(r)-X_k'(r))^2\,dr>\varepsilon\Big\}\\
&\leq o(1)+o(1)+\frac{9\varepsilon}{\delta^2}+\mathbf{P}\Big\{\int_0^T (X_k(r)-X(r))^2\,dr>\frac{\varepsilon}{4}\Big\}+\mathbf{P}\Big\{\int_0^T (X(r)-X_k'(r))^2\,dr>\frac{\varepsilon}{4}\Big\}\\
&\to 9\varepsilon/\delta^2 \qquad \text{as} \quad k\to\infty.
\end{aligned}$$

Here $o(1)$ denotes probabilities that goes to zero as $k\to\infty$, by the convergence in probability shown in the first part of the proof. Since $\varepsilon>0$ is arbitrary, the left-hand side is zero for each $\delta>0$, so that the two limits in probability therein coincide. □

EXERCISE 69 A possible problem when extending the Itô integral from E_T to P_T [S_T to E_T], would be that for $X\in E_T$ [$X\in S_T$], with the integral given by Definition 8.1 [Definition 7.5], the integral obtained from Definition 10.3 [Definition 8.1] would differ from that of Definition 8.1 [Definition 7.5]. Why does this not happen?

10.3 Approximating Sums for Itô Integrals of Processes in E_T

Given $X\in E_T$ and $\varepsilon>0$, an inspection of the proofs of Theorem 9.1 reveals that

$$\mathbf{E}\Big\{\int_0^T (X(r)-Z(r))^2\,dr\Big\}\leq\varepsilon \qquad \text{for some continuous and bounded} \quad Z\in E_{\tilde{T}}.$$

Here $\tilde{T}=T$ if $T<\infty$, but $\tilde{T}<T$ if $T=\infty$. If X is continuous, we may take $Z(t)=I_{[0,\tilde{T}]}(t)\,X^{(N)}(t)$, where $X^{(N)}=\max\{\min\{X,N\},-N\}$, for some sufficiently large $N\in\mathbb{N}$. For the corresponding Itô integral process, Theorem 10.2 gives

$$\mathbf{E}\Big\{\Big(\int_0^t X\,dB-\int_0^t Z\,dB\Big)^2 dr\Big\}\leq\mathbf{E}\Big\{\int_0^T (X(r)-Z(r))^2\,dr\Big\}\leq\varepsilon \qquad \text{for} \quad t\in[0,T].$$

For a continuous and bounded $Z\in E_{\tilde{T}}$, *Dominated Convergence* gives

$$\lim\left\{\mathbf{E}\Big\{\int_0^T\Big(Z(r)-\sum_{i=1}^n Z(t_{i-1})I_{(t_{i-1},t_i]}(r)\Big)^2 dr\Big\}: \begin{array}{c} 0=t_0<t_1<\ldots<t_n=T \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0\end{array}\right\}=0.$$

Since $\sum_{i=1}^n Z(t_{i-1})I_{(t_{i-1},t_i]}\in S_T$, Definition 7.5 shows that

$$\int_0^t \sum_{i=1}^n Z(t_{i-1})I_{(t_{i-1},t_i]}\,dB=\sum_{t_i<t} Z(t_{i-1})\,(B(t_i)-B(t_{i-1}))+I_{(t_j,t_{j+1}]}(t)\,Z(t_j)\,(B(t)-B(t_j)).$$

Hence we have, using Theorem 10.2 together with the fact that Z is bounded,

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\int_0^t Z\,dB-\sum_{t_i<t} Z(t_{i-1})\,(B(t_i)-B(t_{i-1}))\Big)^2\Big\}\\
&\leq 2\mathbf{E}\Big\{\Big(\int_0^t Z\,dB-\int_0^t\sum_{i=1}^n Z(t_{i-1})I_{(t_{i-1},t_i]}\,dB\Big)^2\Big\}+2\mathbf{E}\big\{I_{(t_j,t_{j+1}]}(t)\,Z(t_j)^2(B(t)-B(t_j))^2\big\}\\
&\leq 2\mathbf{E}\Big\{\int_0^T\Big(Z(r)-\sum_{i=1}^n Z(t_{i-1})I_{(t_{i-1},t_i]}(r)\Big)^2 dr\Big\}+2\,I_{(t_j,t_{j+1}]}(t)\,\mathbf{E}\{Z(t_j)^2\}\,(t_{j+1}-t_j)\ \to\ 0
\end{aligned}$$

uniformly for $t\in[0,T]$, as $\max_{1\leq i\leq n} t_i-t_{i-1}\to 0$: In particular, we have

Theorem 10.5 *Let $X\in E_T$ be continuous. Given an $\varepsilon>0$, there exist $N\in\mathbb{N}$ and $0=t_0<t_1<\ldots<t_n<\infty$, such that*

$$\mathbf{E}\Big\{\Big(\int_0^t X\,dB-\sum_{t_i<t} X^{(N)}(t_{i-1})\,(B(t_i)-B(t_{i-1}))\Big)^2\Big\}\leq\varepsilon \qquad \textit{for} \quad t\in[0,T].$$

11 Eleventh Lecture

11.1 Stochastic Integration of Processes in P_T (continued)

The two results of this section, Theorems 11.1 and 11.2 below, as well as their proofs, are identical (more or less) to [27, Lemmas 4.5 and 4.6]:

EXERCISE 70 Let τ be a stopping time wrt. the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Show that the events $\{\tau<t\}$ and $\{\tau=t\}$ are $\mathcal{F}_t$-measurable for $t\geq 0$.

Theorem 11.1 *Let $\{\mathcal{F}_t\}_{t\geq 0}$ be augmented. For a process $X\in P_T$, we have, in the sense of convergence in probability (*P-lim*)*

$$\operatorname*{P\text{-}lim}_{n\to\infty}\int_0^T (X_n(r)-X(r))^2\,dr = 0 \qquad \textit{for some sequence} \quad \{X_n\}_{n=1}^{\infty}\subseteq E_T.$$

Proof. Given an $n>0$, define

$$\tau_n \equiv T\wedge \inf\Big\{t\in[0,T]:\int_0^t X(r)^2\,dr\geq n\Big\} \quad \text{and} \quad X_n(t)\equiv X(t)I_{\{t\leq\tau_n\}} \tag{11.1}$$

(recall that $\inf\{\emptyset\}=\infty$). Provided that $\int_0^t X(r)^2\,dr$ is adapted, we have

$$\{\tau_n\leq t\} = \Big\{\int_0^t X(r)^2\,dr\geq n\Big\}\in\mathcal{F}_t \qquad \text{for} \quad t\in[0,T), \tag{11.2}$$

so that τ_n is a stopping time. This in turn makes X_n adapted, by Exercise 70, since

$$\begin{aligned}\{X_n(t)\in B\} &= \big(\{t\leq\tau_n\}\cap\{X(t)\in B\}\big)\cup\big(\{t>\tau_n\}\,I_B(0)\big)\\ &= \big(\{\tau_n<t\}^c\cap\{X(t)\in B\}\big)\cup\begin{cases}\{\tau_n\geq t\}^c & \text{if } 0\in B\\ \emptyset & \text{if } 0\notin B\end{cases}\in\mathcal{F}_t \quad \text{for } B\in\mathcal{B}(\mathbb{R}).\end{aligned}$$

Since X^2 is measurable and adapted, there exists a progressively measurable version $\tilde{X}^2$ of X^2, by Theorem 8.7. The process $\{\int_0^t \tilde{X}(r)^2\,dr\}_{t\in[0,T]}$ is adapted, by Exercise 64, and a version of $\{\int_0^t X(r)^2\,dr\}_{t\in[0,T]}$, by Lemma 9.2. Hence $\int_0^t X(r)^2\,dr$ is adapted, by Exercise 66, since $\{\mathcal{F}_t\}_{t\geq 0}$ is augmented.

By Exercise 71 below, X_n is measurable. Further we have

$$\int_0^T X_n(r)^2\,dr = \int_0^{\tau_n} X(r)^2\,dr\leq n. \tag{11.3}$$

This is so as an immediate consequence of the fact that $\int_0^t X(r)^2\,dr$ is a continuous function of t, by *Absolute Continuity of the Integral* (see Lemma 12.1 below).

Now pick an $\delta>0$. It is enough to prove that

$$\mathbf{P}\Big\{\int_0^T (X_n(r)-X(r))^2\,dr>\delta\Big\}\leq\varepsilon \qquad \text{for some} \quad n=n(\delta,\varepsilon)\in\mathbb{N}, \quad \text{for each} \quad \varepsilon>0.$$

Notice that $X_n\in E_T$, because $\int_0^T X_n(r)^2\,dr\leq n$. Since $X\in P_T$, we have

$$\mathbf{P}\Big\{\int_0^T X(r)^2\,dr \geq n\Big\} \leq \varepsilon \qquad \text{for some} \quad n = n(\varepsilon) \in \mathbb{N}.$$

It follows that

$$\begin{aligned}\mathbf{P}\Big\{\int_0^T (X_n(r) - X(r))^2\,dr > \delta\Big\} &\leq \mathbf{P}\big\{X(r) \neq X_n(r) \ \text{ for some } \ r \in [0,T]\big\} \\ &\leq \mathbf{P}\Big\{\int_0^T X(r)^2\,dr \geq n\Big\} \leq \varepsilon. \quad \square\end{aligned}$$

EXERCISE 71 Prove that the process X_n given by (11.1) is measurable.

Theorem 11.2 *For a process $X \in E_T$, we have, for each constant $C > 0$,*

$$\mathbf{P}\Big\{\sup_{t\in[0,T]}\Big|\int_0^t X(r)\,dB(r)\Big| > \lambda\Big\} \leq \frac{C}{\lambda^2} + \mathbf{P}\Big\{\int_0^T X(r)^2\,dr > C\Big\} \qquad \textit{for} \quad \lambda > 0.$$

Proof. By Exercise 72 below, it is enough to prove that, for each $n > 0$,

$$\mathbf{P}\Big\{\sup_{t\in[0,T]}\Big|\int_0^t X(r)\,dB(r)\Big| > \lambda\Big\} \leq \frac{n}{\lambda^2} + \mathbf{P}\Big\{\int_0^T X(r)^2\,dr \geq n\Big\} \qquad \text{for} \quad \lambda > 0. \tag{11.4}$$

When $\int_0^T X(r)^2\,dr < n$, we have $\tau_n = T$ and $X_n = X$, with the notation (11.1). By the *Doob-Komogorov Inequality*, together with Theorem 10.2 and (11.3), this gives

$$\begin{aligned}\mathbf{P}\Big\{\sup_{t\in[0,T]}\Big|\int_0^t X(r)\,dB(r)\Big| > \lambda\Big\} &\leq \mathbf{P}\Big\{\sup_{t\in[0,T]}\Big|\int_0^t X_n(r)\,dB(r)\Big| > \lambda\Big\} + \mathbf{P}\{X_n \neq X\} \\ &\leq \mathbf{E}\Big\{\Big(\int_0^T X_n(r)\,dB(r)\Big)^2\Big\}\Big/\lambda^2 + \mathbf{P}\Big\{\int_0^T X(r)^2\,dr \geq n\Big\} \\ &= \mathbf{E}\Big\{\int_0^T X_n(r)^2\,dr\Big\}\Big/\lambda^2 + \mathbf{P}\Big\{\int_0^T X(r)^2\,dr \geq n\Big\} \\ &\leq \frac{n}{\lambda^2} + \mathbf{P}\Big\{\int_0^T X(r)^2\,dr \geq n\Big\}. \quad \square\end{aligned}$$

EXERCISE 72 Show how Theorem 11.2 follows from the seemingly weaker (11.4).

EXERCISE 73 Construct the integral $\{\{\int_0^t f\,dB\}_{t\in[0,T]}\}_{f\in\mathbb{L}^2([0,T])}$, in mean-square sense, by first making $\{\{\int_0^t f\,dB\}_{t\in[0,T]}\}_{f\in S_T^{\emptyset}}$, and then $\{\{\int_0^t f\,dB\}_{t\in[0,T]}\}_{f\in C([0,T])}$, where $S_T^{\emptyset} = \{f \in S_T : \sigma(f(s) : s \in [0,T]) = \{\emptyset, \Omega\}\}$ (that is, $f \in S_T$ non-random).

EXERCISE 74 Try to explain why a stochastic integral $\{\int_0^t f\,dB\}_{t\in[0,T]}$, of a non-random $f : [0,T] \to \mathbb{R}$, to have deacent properties, requires that $f \in \mathbb{L}^2([0,T])$.

Remark 11.3 The integral $\{\int_0^t f\,dB\}_{t\in[0,T]}$, of a non-random function $f : [0,T] \to \mathbb{R}$, is well-defined when (and only when, if properly understood) $f \in \mathbb{L}^2([0,T])$ (cf. Exercises 73 and 74). It is notable that the Itô integral $\{\int_0^t X\,dB\}_{t\in[0,T]}$ is well-defined when (X is adapted and measurable with) $\mathbf{P}\{X \in \mathbb{L}^2([0,T])\} = 1$, which in the particular case of a non-random X becomes $X \in \mathbb{L}^2([0,T])$. #

*Remark 11.4 The construction in Exercise 73 is of Daniell type (cf. [18, Section 9]). It is not a Riemann integral, although starting out as that. (The "usual" Lebesgue integral can be got from the Riemann integral in a similar way.) Many authors get the integral from a Hilbertspace isomorfism between $\mathbb{L}^2([0,T])$ and the values of the integral. This approach is non-constructive, and does not extend beoynd the Gaussian setting [unless (random) Orlicz (Musielak-Orlicz) spaces are employed (e.g., [26])]. See [35, Section 3] on more methods to integrate, albeit only worked out for α-stable stochastic processes. For the general case, try the more demanding [26]. #

11.2 Properties of Stochastic Integrals in P_T

Definition 11.5 *Let $\{\mathcal{F}_t\}_{t\geq 0}$ be augmented. For $X\in P_T$ we define the Itô inegral*

$$\underline{\int_s^t X(r)\,dB(r)} \equiv \int_0^t X(r)\,dB(r) - \int_0^s X(r)\,dB(r) \qquad \textit{for} \quad s,t\in[0,T]$$

(where the integrals on the right-hand side are defined in Defintion 10.3).

Exercise 75 Prove that, for $X\in P_T$, the definition

$$\int_s^t X(r)\,dB(r) \equiv \int_0^t I_{(s,t]}(r)X(r)\,dB(r) \qquad \text{for} \quad s,t\in[0,T],$$

gives the same result as Definition 11.5.

Theorem 11.6 *For the Itô integral process of $X,Y\in P_T$, we have*

(1) $\int_0^t (aX+bY)\,dB = a\int_0^t X\,dB + b\int_0^t Y\,dB$ *for* $a,b\in\mathbb{R}$;

(2) $\int_0^t X\,dB = \int_0^s X\,dB + \int_s^t X\,dB$;

(3) $\{\int_0^t X\,dB\}_{t\in[0,T]}$ *is continuous and progressively measurable, with probability one.*

Proof. (1) This is Exercise 76 below.

(2) This is only a rearrangement of Definition 11.5.

(3) When $\int_0^T X(r)^2\,dr < n$, we have $\tau_n = T$, with the notation (11.1), so that $X_n = X$. By Theorem 10.2, recalling that $X_n\in E_T$, this shows that $\int_0^t X\,dB = \int_0^t X_n\,dB$ is continuous on $\{n-1\leq\int_0^T X(r)^2\,dr < n\}$. This gives continuity for $\int_0^t X\,dB$, with probability one, because

$$\mathbf{P}\bigg\{\bigcup_{n=1}^{\infty}\bigg\{n-1\leq\int_0^T X(r)^2\,dr < n\bigg\}\bigg\} = 1 \qquad \text{for} \quad X\in P_T.$$

Since $\int_0^t X\,dB$ is a limit in probability of random variables adapted to $\mathcal{F}_t$, by Theorem 10.2 and Defintion 10.3, it is an a.s. limit of a subsequence of such random variables. Hence $\int_0^t X\,dB$ is the limit of $\mathcal{F}_t$-measurable functions, except on a null-set. Therefore the inverse image under $\int_0^t X\,dB$, of a Borel-set in $\mathbb{R}$, deviates from that under the limit, which is $\mathcal{F}_t$-measurable, on a subset of this null-event. Since the filtration is augmented (cf. Definition 10.3), it follows that the inverse image under $\int_0^t X\,dB$ is also $\mathcal{F}_t$-measurable, so that $\int_0^t X\,dB$ is adapted. By continuity, we now get progressive measurability from Theorem 8.8. □

EXERCISE 76 Prove property (1) in Theorem 11.6.

11.3 Uniform Integrability and Regular Martingales

Uniform integrability and regularity characterize martingales that converge in $\mathbb{L}^1$.

Definition 11.7 *A family $\{Y_\alpha\}_{\alpha\in\mathfrak{A}}$ of random variables is uniformly integrable if*

$$\lim_{y\to\infty}\sup_{\alpha\in\mathfrak{A}}\int_\Omega I_{\{|Y_\alpha|>y\}}\,|Y_\alpha|\,d\mathbf{P}=0.$$

EXERCISE 77 Show that a family $\{Y_\alpha\}_{\alpha\in\mathfrak{A}}$ of random variables is uniformly integrable if $\sup_{\alpha\in\mathfrak{A}}\mathbf{E}\{|Y_\alpha|^p\}<\infty$ for some constant $p>1$. (**Hint:** *Hölder's inequality*, followed by *Tjebysjev's inequality*.)

Example 11.8 Let $\{Y_\alpha\}_{\alpha\in\mathfrak{A}}$ be a family of random variables, such that

$$|Y_\alpha|\le Z \quad \text{a.s.} \qquad \text{for each} \quad \alpha\in\mathfrak{A}, \qquad \text{for some positive random variable} \quad Z.$$

If in addition $\mathbf{E}\{Z\}<\infty$, then $\{Y_\alpha\}_{\alpha\in\mathfrak{A}}$ is uniformly integrable, by *Absolute Continuity of the Integral* (Lemma 12.1 below), since

$$\sup_{\alpha\in\mathfrak{A}}\int_\Omega I_{\{|Y_\alpha|>y\}}\,|Y_\alpha|\,d\mathbf{P}\le\int_{\{\omega\in\Omega\,:\,Z(\omega)>y\}}Z\,d\mathbf{P}\to 0 \qquad \text{as} \quad y\to\infty. \quad \#$$

In basic probability theory, uniform integrability is used in the following two ways:

(1) For random variables $Y,Y_1,Y_2,\ldots$, such that $\mathbf{E}\{|Y_n|\}<\infty$ for $n\in\mathbb{N}$ and P-$\lim_{n\to\infty}Y_n=Y$, the following properties are equivalent (e.g., [8, Theorem 4.5.4])*

- $\mathbf{E}\{|Y|\}<\infty$ and $\lim_{n\to\infty}\mathbf{E}\{|Y_n-Y|\}=0$;
- $\mathbf{E}\{|Y|\}<\infty$ and $\lim_{n\to\infty}\mathbf{E}\{|Y_n|\}=\mathbf{E}\{|Y|\}$; (11.5)
- $\{Y_n\}_{n=1}^\infty$ is uniformly integrable.

(2) For random variables $Y,Y_1,Y_2,\ldots$, such that $Y_n\to_{\text{distribution}}Y$, we have (e.g., [5, Theorem 3.5])*

- $\{Y_n\}_{n=1}^\infty$ is uniformly integrable $\Rightarrow$ $\lim_{n\to\infty}\mathbf{E}\{Y_n\}=\mathbf{E}\{Y\}$. (11.6)

Definition 11.9 *A martingale* $\{Y(t),\mathcal{F}_t\}_{t\in T}$ *is* <u>*regular*</u> *(or* <u>*closed*</u>*), if there exists a random variable* Y_∞*, with* $\mathbf{E}\{|Y_\infty|\}<\infty$*, such that*

$$\mathbf{E}\{Y_\infty|\mathcal{F}_t\} = Y(t) \qquad \text{for} \quad t\in T.$$

EXERCISE 78 Show that the martingales $\{Y(t)\}_{t\in\{0,\ldots,n\}}$ and $\{Y(t)\}_{t\in[0,T]}$ are regular.

Lemma 11.10 (e.g., [8, Theorem 9.4.6])* *For a martingale* $\{Y(n)\}_{n\geq 0}$ *the following properties are equivalent*

- Y *is a regular martingale;*
- $\lim_{n\to\infty}\mathbf{E}\{|Y(n)-Y_\infty|\} = 0$ *for some random variable* Y_∞ *[with* $\mathbf{E}\{|Y_\infty|\}<\infty$*];*
- $\{Y(n)\}_{n=1}^{\infty}$ *is uniformly integrable.*

Theorem 11.11 *For a right-continuous martingale* $\{Y(t)\}_{t\geq 0}$*, the following properties are equivalent*

- Y *is a regular martingale;*
- $\lim_{t\to\infty}\mathbf{E}\{|Y(t)-Y_\infty|\} = 0$ *for some random variable* Y_∞ *[with* $\mathbf{E}\{|Y_\infty|\}<\infty$*];*
- $\{Y(t)\}_{t\geq 0}$ *is uniformly integrable.*

*EXERCISE 79 Work on a derivation of Theorem 11.11 from Lemma 11.10.

11.4 Approximating Sums for Itô Integrals of Processes in P_T

The following important result substantially improves on our peace-meal findings in Section 10.3. However, a proof requires a more abstract approach to Itô integration, than the one we have taken (but still very similar to it in its essential parts), that uses local martingales. These, by the way, we introduce in the next lecture.

Theorem 11.12 (e.g., [33, Chapter IV, Proposition 2.13])* *For* $X\in P_T$*, we have*

$$\int_0^t X\,dB = \text{P-lim}\left\{\sum_{i=1}^n X(t_{i-1})\,(B(t_i)-B(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}.$$

12 Twelfth Lecture

12.1 Optional Stopping

Lemma 12.1 (Absolute Continuity of the Integral) (e.g., [18, p. 176])* *Let Y be a random variable with $\mathbf{E}\{|Y|\}<\infty$. Given a constant $\varepsilon>0$, we have*

$$\int_A |Y|\,d\mathbf{P}<\varepsilon \quad \text{for} \quad A\in\mathcal{F} \quad \text{with} \quad \mathbf{P}\{A\}<\delta, \quad \text{for some constant} \quad \delta>0.$$

*Remark 12.2 From the proof of Theorem 4.20, we have that, for a right-continuous and adapted process Y, and a stopping time τ,

$$\Big\{\Lambda\in\sigma(Y(\tau)):\Lambda\cap\{\tau<t\}\in\mathcal{F}_t\Big\}=\sigma(Y(\tau)).$$

This together with Exercise 70 shows that $Y(t\wedge\tau_n)$ is adapted, since for $B\in\mathcal{B}(\mathbb{R})$

$$\{Y(t\wedge\tau_n)\in B\}=\Big(\{\tau_n<t\}\cap\{Y(\tau_n)\in B\}\Big)\cup\Big(\{\tau_n\geq t\}\cap\{Y(t)\in B\}\Big)\in\mathcal{F}_t. \quad \#$$

Theorem 12.3 (Optional Stopping Theorem) *If $\{Y(t),\mathcal{F}_t\}_{t\geq 0}$ is a right-continuous martingale and τ a stopping time, then $\{Y(t\wedge\tau),\mathcal{F}_t\}_{t\geq 0}$ is a martingale.*

Proof. We have $Y(t\wedge\tau)$ adapted, by Remark 12.2. Writing $\tau_n=\lfloor 2^n\tau+1\rfloor/2^n$, we have $\tau_n=2^{-n}k$ for $\tau\in(2^{-n}(k-1),2^{-n}k]$, so that $\{\tau_n=2^{-n}k\}\in\mathcal{F}_{2^{-n}k}$. Further, $\tau_n\downarrow\tau$ as $n\to\infty$. In addition, $\{Y(t\wedge\tau_n)\}_{n=1}^{\infty}$ is uniformly integrable: To see this, recall that $|Y|$ is a submartingale, by Exercise 55. It follows that

$$\begin{aligned}\sup_{n\in\mathbb{N}}\int_{\{|Y(t\wedge\tau_n)|>y\}}|Y(t\wedge\tau_n)|\,d\mathbf{P}&=\sup_{n\in\mathbb{N}}\sum_{k=0}^{\infty}\int_{\{|Y(t\wedge 2^{-n}k)|>y,\,\tau_n=2^{-n}k\}}|Y(t\wedge 2^{-n}k)|\,d\mathbf{P}\\&\leq\sup_{n\in\mathbb{N}}\sum_{k=0}^{\infty}\int_{\{|Y(t\wedge 2^{-n}k)|>y,\,\tau_n=2^{-n}k\}}\mathbf{E}\Big\{|Y(t)|\,\Big|\,\mathcal{F}_{2^{-n}k}\Big\}\,d\mathbf{P}\\&=\sup_{n\in\mathbb{N}}\sum_{k=0}^{\infty}\int_{\{|Y(t\wedge 2^{-n}k)|>y,\,\tau_n=2^{-n}k\}}|Y(t)|\,d\mathbf{P}\\&=\sup_{n\in\mathbb{N}}\int_{\{|Y(t\wedge\tau_n)|>y\}}|Y(t)|\,d\mathbf{P}.\end{aligned}\tag{12.1}$$

In particular, this holds for $y=0$, so that, by *Markov's Inequality*,

$$\sup_{n\in\mathbb{N}}\mathbf{P}\{|Y(t\wedge\tau_n)|>y\}\leq\sup_{n\in\mathbb{N}}\frac{\mathbf{E}\{|Y(t\wedge\tau_n)|\}}{y}\leq\frac{\mathbf{E}\{|Y(t)|\}}{y}\to 0 \qquad \text{as} \quad y\to\infty.$$

By *Absolute Continuity of the Integral*, we therefore have that the right-hand side of (12.1) goes to zero as $y\to\infty$, which establishes uniform integrability.

By uniform integrability and right-continuity, we have $\lim_{n\to\infty}\mathbf{E}\{|Y(t\wedge\tau_n)-Y(t\wedge\tau)|\}=0$ [recall (11.5)]. Notice that $\{\tau_n=2^{-n}k\}\in\mathcal{F}_s$ for $2^{-n}k\leq s$, and that

$$\{\tau_n > s\} = \{\tau_n \le s\}^c = \Bigg(\bigcup_{\{k\,:\,2^{-n}k\le s\}} \Big\{\tau \in (2^{-n}(k-1), 2^{-n}k]\Big\}\Bigg)^c \in \mathcal{F}_s.$$

Using Exercise 68, we therefore obtain, with convergence in mean $(\mathbb{L}^1)$,

$$\begin{aligned}
&\mathbf{E}\{Y(t\wedge\tau)|\mathcal{F}_s\}\\
&= \lim_{n\to\infty} \mathbf{E}\{Y(t\wedge\tau_n)|\mathcal{F}_s\}\\
&= \lim_{n\to\infty}\Bigg(\mathbf{E}\Big\{I_{\{\tau_n>s\}}Y(t\wedge\tau_n)\Big|\mathcal{F}_s\Big\} + \sum_{\{k\,:\,2^{-n}k\le s\}} \mathbf{E}\Big\{I_{\{\tau_n=2^{-n}k\}}Y(t\wedge 2^{-n}k)\Big|\mathcal{F}_s\Big\}\Bigg)\\
&= \lim_{n\to\infty}\Bigg(I_{\{\tau_n>s\}}\mathbf{E}\{Y(t\wedge\tau_n)|\mathcal{F}_s\} + \sum_{\{k\,:\,2^{-n}k\le s\}} I_{\{\tau_n=2^{-n}k\}}\mathbf{E}\Big\{Y(t\wedge 2^{-n}k)\Big|\mathcal{F}_s\Big\}\Bigg)\\
&= \lim_{n\to\infty}\Bigg(\sum_{\{k\,:\,2^{-n}k>s\}} I_{\{\tau_n=2^{-n}k\}}\mathbf{E}\{Y(t\wedge 2^{-n}k)|\mathcal{F}_s\} + \sum_{\{k\,:\,2^{-n}k\le s\}} I_{\{\tau_n=2^{-n}k\}}Y(2^{-n}k)\Bigg)\\
&= \lim_{n\to\infty}\Bigg(\sum_{\{k\,:\,2^{-n}k>s\}} I_{\{\tau_n=2^{-n}k\}}Y(s) + \sum_{\{k\,:\,2^{-n}k\le s\}} I_{\{\tau_n=2^{-n}k\}}Y(2^{-n}k)\Bigg)\\
&= \lim_{n\to\infty} Y(s\wedge\tau_n) = Y(s\wedge\tau) \qquad \text{for} \quad 0\le s<t. \quad \Box
\end{aligned}$$

12.2 Local Martingales

We now define local smartingales (of which only local martingales are used). These are fundamental in stohastic calculus, and do not require integrability as smartingales.

Definition 12.4 *An adapted stochastic process $\{Y(t)\}_{t\ge 0}$ is a local (s)martingale, if there exists a localizing sequence of stopping times $0 \le \tau_1 \le \tau_2 \le \dots,$ with $\lim_{n\to\infty}\tau_n = \infty$ a.s., such that*

$$\{Y(t\wedge\tau_n)\}_{t\ge 0} \quad \textit{is a uniformly integrable (s)martingale for every} \quad n\in\mathbb{N}.$$

Theorem 12.5 *A right-continuous martingale $\{Y(t),\mathcal{F}_t\}_{t\ge 0}$ is a local martingale.*

Proof. The non-random variables $\tau_n = n$, $n\in\mathbb{N}$, are adapted to $\{\emptyset,\Omega\}$, and thus to each $\mathcal{F}_t$. They satisfy $0\le\tau_1\le\tau_2\le\dots$ and $\mathbf{P}\{\lim_{n\to\infty}\tau_n=\infty\} = 1$. The process $Y(t\wedge\tau_n) = Y(t\wedge n)$ is uniformly integrable, by Theorem 11.11, since $\lim_{t\to\infty}\mathbf{E}\{|Y(t\wedge\tau_n) - Y(n)|\} = 0$. Moreover, it is a martingale, by the *Optional Stopping theorem.* $\Box$

We have the following more or less trivial, but important observation:

Theorem 12.6 *A right-continuous adapted process $\{Y(t)\}_{t\ge 0}$ is a local martingale, if there exist stopping times $0\le\tau_1\le\tau_2\le\dots,$ with $\lim_{n\to\infty}\tau_n=\infty$ a.s., such that*

$$\{Y(t\wedge\tau_n)\}_{t\ge 0} \quad \textit{is a martingale for every} \quad n\in\mathbb{N}.$$

Proof. For the stopping times $\tilde{\tau}_n \equiv n\wedge\tau_n$, we have $\lim_{n\to\infty}\tilde{\tau}_n=\infty$ a.s. Therefore it is enough to prove that (cf. Definition 12.4)

$$\{Y(t\wedge\tilde{\tau}_n)\}_{t\geq 0} \quad \text{is a uniformly integrable martingale for every } n\in\mathbb{N}.$$

Notice that $Y(t\wedge\tilde{\tau}_n) = M(t\wedge n)$, where $M(t)=Y(t\wedge\tau_n)$ is a martingale, by assumption. Hence $M(t\wedge n)$ is a uniformly integrable martingale, by Theorem 12.5. □

*Remark 12.7 In general, it is not true that integrable local martingales are martingales (e.g., [33, Exercise V.2.12]), even if practice suggests this.

Because of Theorem 12.6, some authors (e.g., [22]) only require that $Y(t\wedge\tau_n)$ is a martingale (not necessarily uniformly integrable), in the definition of local martingale.

By Section 16.1, the Itô integral $\int_0^t X\,dY$ is well-defined, for Y a continuous local martingale and X a continuous adapted process. In particular $Y(t)-Y(0) = \int_0^t dY$, so that continuous local martingales Y with $Y(0)=0$ are Itô integrals.

Some authors require (right-) continuity (e.g., [33]) in the definition of a local martingale Y, and/or $Y(0)=0$ (e.g., [22]). This is due to a wish to identify local martingales with Itô integrals (see above). We will have more to say on that later. #

Definition 12.8 *A stochastic process $\{Y(t)\}_{t\in T}$ is bounded if $|Y(t)|\leq K$ for all $t\in T$, with probability one, for some constant $K>0$.*

Theorem 12.9 *A continuous local martingale $\{Y(t)\}_{t\geq 0}$ such that $|Y(t)|\leq Z_T$ a.s. for each $t\in[0,T]$, for a random variable Z_T with $\mathbf{E}\{Z_T\}<\infty$, for each $T\geq 0$, is a martingale. In particular, a continuous local martingale $\{Y(t)\}_{t\geq 0}$, such that $\{Y(t)\}_{t\in[0,T]}$ is bounded for each $T\geq 0$, is a martingale.*

Proof. Pick a localizing sequence $\{\tau_k\}_{k=1}^{\infty}$. Since $\{Y(t\wedge\tau_k)\}_{k=1}^{\infty}$ is uniformly integrable, by Example 11.8, and $Y(t\wedge\tau_k)\to Y(t)$ a.s. as $k\to\infty$, (11.5) shows that $\mathbf{E}\{|Y(t\wedge\tau_k)-Y(t)|\}\to 0$ for $t\geq 0$. Hence Exercise 68 gives

$$\mathbf{E}\{Y(t)|\mathcal{F}_s\} \leftarrow \mathbf{E}\{Y(t\wedge\tau_k)|\mathcal{F}_s\} = \mathbf{E}\{Y(s\wedge\tau_k)\} \to \mathbf{E}\{Y(s)\} \qquad \text{as} \quad k\to\infty$$

for $0\leq s<t$ (with convergence in $\mathbb{L}^1$), since $Y(t\wedge\tau_k)$ is a martingale. □

12.3 Properties of Stochastic Integrals in P_T (continued)

Exercise 80 Take $0<T_1<T_2$ and $X\in P_{T_2}$. Explain why the Itô integral processes $\{\int_0^t X\,dB\}_{t\in[0,T_1]}$ and $\{\int_0^t X\,dB\}_{t\in[0,T_2]}$, obtained by using Definition 10.3 on P_{T_1} and P_{T_2}, respectively, are versions of each other, when restricted to $[0,T_1]$.

Theorem 12.10 *For* $X \in \bigcup_{T>0} P_T$, *there exist stopping times* $0 \leq \tau_1 \leq \tau_2 \leq \ldots$, *with* $\mathbf{P}\{\lim_{n\to\infty} \tau_n = \infty\} = 1$, *such that*

$$\Big\{\int_0^{t\wedge\tau_n} X\,dB\Big\}_{t\geq 0} \quad \textit{is a bounded martingale for every } n\in\mathbb{N}.$$

In particular, the Itô integral process $\{\int_0^t X\,dB\}_{t\geq 0}$ *is a local martingale.*

Theorem 12.10 can be got from Theorem 12.3 (which is a very important result). See Exercise 82 for some gymnastics on this. However, there is a simple direct argument:

Proof of Theorem 12.10. Define [see (11.1)]

$$\tau_n \equiv n \wedge \inf\Big\{t\geq 0 : \int_0^t X(r)^2\,dr \geq n\Big\} \qquad \text{and} \qquad X_n(t) \equiv X(t) I_{\{t\leq\tau_n\}}.$$

Recall from the proof of Theorem 11.1, that τ_n is a stopping time and $X_n \in E_n$ (E_T with $T=n$). Since $\int_0^{\tau_n+1} X(r)^2\,dr \geq n$ when $\tau_n \leq m \leq n-1$, we have

$$\begin{aligned}\mathbf{P}\Big\{\lim_{n\to\infty}\tau_n<\infty\Big\} = \mathbf{P}\Big\{\bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\{\tau_n\leq m\}\Big\} &\leq \mathbf{P}\Big\{\bigcup_{m=1}^{\infty}\bigcap_{n=m+1}^{\infty}\{\tau_n\leq m\}\Big\}\\ &= \mathbf{P}\Big\{\bigcup_{m=1}^{\infty}\bigcap_{n=m+1}^{\infty}\Big\{\tau_n\leq m, \int_0^{\tau_n+1} X(r)^2\,dr\geq n\Big\}\Big\}\\ &\leq \mathbf{P}\Big\{\bigcup_{m=1}^{\infty}\Big\{\int_0^{m+1} X(r)^2\,dr=\infty\Big\}\Big\} = 0\end{aligned}$$

(because $X\in P_{m+1}$ for $m\in\mathbb{N}$), so that $\mathbf{P}\{\lim_{n\to\infty}\tau_n=\infty\} = 1$, as required.

Now, notice that (see Exercise 81 below)

$$\int_0^{t\wedge\tau_n} X(r)\,dB(r) = \int_0^t I_{\{r\leq\tau_n\}} X(r)\,dB(r) = \int_0^t X_n(r)\,dB(r) = \int_0^{t\wedge n} X_n(r)\,dB(r).$$

Hence $\{\int_0^{t\wedge\tau_n} X\,dB\}_{t\geq 0}$ is a martingale, since $\{\int_0^t X_n\,dB\}_{t\geq 0}$ is, by Theorem 10.2 (since $X_n\in E_n$). Moreover, we have (check $n\leq t$ and $n>t$ separately)

$$\int_0^{t\wedge\tau_n} X(r)\,dB(r) = \int_0^{t\wedge n} X_n(r)\,dB(r) = \mathbf{E}\Big\{\int_0^n X_n(r)\,dB(r)\,\Big|\,\mathcal{F}_t\Big\} \qquad \text{for} \quad t\geq 0$$

(since $\int_0^t X_n\,dB$ is a martingale), so that $\{\int_0^{t\wedge\tau_n} X\,dB\}_{t\geq 0}$ is a regular martingale. Hence it is also uniformly integrable, by Theorem 11.11. □

Exercise 81 Prove that, for $X\in P_T$, we have

$$\int_0^{s\wedge t} X(r)\,dB(r) = \int_0^t I_{[0,s]}(r) X(r)\,dB(r) \qquad \textit{for} \quad s,t\in[0,T].$$

*Exercise 82 Work on a proof of Theorem 12.10 that builds on Theorem 12.3.

13 Thirteenth Lecture

13.1 Examples of Stochastic Integration

Example 13.1 The Itô integral $\int_s^t r^{-\alpha}\,dB(r) = \int_0^\infty I_{(s,t]}(r)\,r^{-\alpha}\,dB(r)$ is well-defined for $\alpha\in(1/2,\infty)$ $[\alpha\in(-\infty,1/2)]$ when $0<s\leq t\leq\infty$ $[0\leq s\leq t<\infty]$, because these conditions give $\int_0^\infty(I_{(s,t]}(r)\,r^{-\alpha})^2\,dr<\infty$ (a.s.), so that $I_{(s,t]}(r)\,r^{-\alpha}\in\underline{P}_\infty$. #

Example 13.2 Given a constant $t_0\in[0,\infty)$, we have $X(\cdot)\equiv I_{(0,t_0]}(\cdot)\in S_{t_0}\subseteq S_\infty$, because X is measurable [since left-continuous (Exercise 43)], and adapted [since $X(r)$ is a constant (0 or 1) at each $r\geq 0$], with $\int_0^{t_0}X(r)^2\,dr=\int_0^\infty X(r)^2\,dr$ $= t_0<\infty$. By Definition 7.5 of the Itô integral on S_{t_0} $(\underline{S_\infty})$, we have

$$\int_0^\infty X(r)\,dB(r)=\int_0^{t_0}X(r)\,dB(r)=\int_0^{t_0}dB(r)=B(t_0)-B(0)=B(t_0).\quad \#$$

Example 13.3 (Wald's Identity) (Weak version) For a stopping time τ, the process $X\equiv I_{(0,\tau]}$ is measurable, since left-continuous, and adapted, since

$$X(t)^{-1}(B)=I_{\{0\in B\}}\{\tau<t\}+I_{\{1\in B\}}\{\tau\geq t\}\in\mathcal{F}_t\qquad\text{for}\quad B\in\mathcal{B}(\mathbb{R})$$

(with obvious notation), by Exercise 70. Taking $\mathbf{E}\{\tau\}<\infty$, we have $X\in\underline{E_\infty}$, since

$$\mathbf{E}\Big\{\int_0^\infty X(r)^2\,dr\Big\}=\int_0^\infty\mathbf{E}\{X(r)^2\}\,dr=\int_0^\infty\mathbf{P}\{\tau\geq r\}\,dr=\mathbf{E}\{\tau\}<\infty,$$

by *Fubini's theorem*. Using Example 13.2 together with Theorem 10.2, we thus obtain

$$\begin{array}{lllll}\boxed{\mathbf{E}\{B(\tau)\}} & = & \mathbf{E}\Big\{\int_0^\tau dB(r)\Big\} & =\mathbf{E}\Big\{\int_0^\infty X(r)\,dB(r)\Big\} & \boxed{=0}\\ \boxed{\mathbf{E}\{B(\tau)^2\}} & =\mathbf{E}\Big\{\Big(\int_0^\infty X(r)\,dB(r)\Big)^2\Big\} & = & \mathbf{E}\Big\{\int_0^\infty X(r)^2\,dr\Big\} & \boxed{=\mathbf{E}\{\tau\}}\end{array}.\quad\#$$

13.2 Introduction to Quadratic Variation of the Itô Integral

Theorem 13.4 *For $X\in S_\infty$, the Itô integral process has quadratic variation*

$$\Big[\int_0^\cdot X\,dB\Big](t)=\int_0^t X(r)^2\,dr\qquad\text{for}\quad t\geq 0,\quad\text{with convergence in mean }(\mathbb{L}^1).$$

Proof. By Definition 7.4, there exist constants $0=\tau_0<\ldots<\tau_m<\infty$, such that

$$X(t)=X(0)I_{\{0\}}(t)+\sum_{i=1}^m X_{\tau_{i-1}}I_{(\tau_{i-1},\tau_i]}(t)\qquad\text{for}\quad t\in(0,\infty).$$

Pick a $t>0$ together with partition $0=t_0<\ldots<t_n=t$ of $[0,t]$. Let $0=\tilde{t}_0<\ldots<\tilde{t}_N=t$ $(N\geq n)$ be the partition $\{t_i\}_{i=0}^n$ "enriched" with the points $\tau_1,\ldots,\tau_{\tilde{m}}$ that belong to $(0,t)$. Using Example 13.2 and Theorem 2.13, we get

$$
\begin{aligned}
\sum_{i=1}^{n}\Big(\int_0^{\tilde t_i} X\,dB-\int_0^{\tilde t_{i-1}} X\,dB\Big)^2 &= \sum_{i=1}^{n}\Big(\int_{\tilde t_{i-1}}^{\tilde t_i} X\,dB\Big)^2\\
&= \sum_{j=1}^{m}\sum_{\tilde t_i\in(\tau_{j-1},\tau_j]}\Big(\int_{\tilde t_{i-1}}^{\tilde t_i} X\,dB\Big)^2\\
&= \sum_{j=1}^{m} X(\tau_{j-1})^2\sum_{\tilde t_i\in(\tau_{j-1},\tau_j]}\Big(\int_{\tilde t_{i-1}}^{\tilde t_i} dB\Big)^2\\
&= \sum_{j=1}^{m} X(\tau_{j-1})^2\sum_{\tilde t_i\in(\tau_{j-1},\tau_j]}(B(\tilde t_i)-B(\tilde t_{i-1}))^2\\
&\to \sum_{j=1}^{m} X(\tau_{j-1})^2\,(\tau_j-\tau_{j-1}) = \int_0^t X(r)^2\,dr
\end{aligned}
$$

in mean square (and thus in mean) as $\max_{1\le i\le n} t_i-t_{i-1}\to 0$. Moreover, we have

$$
\begin{aligned}
&\mathbf{E}\bigg\{\bigg|\sum_{i=1}^{n}\Big(\int_0^{\tilde t_i} X\,dB-\int_0^{\tilde t_{i-1}} X\,dB\Big)^2-\sum_{i=1}^{n}\Big(\int_0^{t_i} X\,dB-\int_0^{t_{i-1}} X\,dB\Big)^2\bigg|\bigg\}\\
&\quad= \mathbf{E}\bigg\{\bigg|\sum_{j=1}^{\tilde m}\sum_{\{i\,:\,t_{i-1}<\tau_j<t_i\}}\bigg(\Big(\int_{\tau_j}^{t_i} X\,dB\Big)^2+\Big(\int_{t_{i-1}}^{\tau_j} X\,dB\Big)^2-\Big(\int_{t_{i-1}}^{t_i} X\,dB\Big)^2\bigg)\bigg|\bigg\}\\
&\quad= \mathbf{E}\bigg\{\bigg|-2\sum_{j=1}^{\tilde m}\sum_{\{i\,:\,t_{i-1}<\tau_j<t_i\}}\Big(\int_{\tau_j}^{t_i} X\,dB\Big)\Big(\int_{t_{i-1}}^{\tau_j} X\,dB\Big)\bigg|\bigg\}\\
&\quad\le \sum_{j=1}^{\tilde m}\sum_{\{i\,:\,t_{i-1}<\tau_j<t_i\}}\bigg(\mathbf{E}\Big\{\Big(\int_{\tau_j}^{t_i} X\,dB\Big)^2\Big\}+\mathbf{E}\Big\{\Big(\int_{t_{i-1}}^{\tau_j} X\,dB\Big)^2\Big\}\bigg)\\
&\quad= \sum_{j=1}^{\tilde m}\sum_{\{i\,:\,t_{i-1}<\tau_j<t_i\}}\mathbf{E}\Big\{\int_{t_{i-1}}^{t_i} X(r)^2\,dr\Big\}\to 0
\end{aligned}
$$

as $\max_{1\le i\le n} t_i-t_{i-1}\to 0$. Here we used that $\int_{t_{i-1}}^{t_i} X(r)^2\,dr\to 0$ [because of *Absolute Continuity of the Integral* (Lemma 12.1)], together with dominated convergence [$\int_{t_{i-1}}^{t_i} X(r)^2\,dr\le\int_0^t X(r)^2\,dr$, which is integrable]. □

13.3 Quadratic Variation of the Itô Integral

Theorem 13.5 *For $X\in E_T$, the Itô integral process has quadratic variation*

$$\Big[\int_0^{\cdot} X\,dB\Big](t)=\int_0^t X(r)^2\,dr \qquad \textit{for}\quad t\in[0,T],\quad \textit{with convergence in probability.}$$

Exercise 83 Can the Itô integral process be continuously differentiable?

Example 13.6 Let $\{a(t)\}_{t\in[0,T]}$ and $\{b(t)\}_{t\in[0,T]}$ be stochastic processes, such that $\sqrt{|a|}\in P_T$ and $b\in E_T$. The Lebesgue integral $\int_0^t a(r)\,dr$ is well-defined, since $\sqrt{|a|}\in P_T$ means that $a\in\mathbb{L}^1([0,T])$. Hence we may define

$$Y(t)=\int_0^t a(r)\,dr+\int_0^t b(r)\,dB(r)\qquad\text{for}\quad t\in[0,T].$$

Here $\int_0^\cdot a(r)\,dr = \int_0^\cdot a(r)^+\,dr - \int_0^\cdot a(r)^-\,dr$ is continuous [because of *Absolute Continuity of the Integral* (Lemma 12.1)], and has finite variation [because a difference between two increasing functions (Theorem 1.4)]. Now Theorems 1.12 and 13.5 give

$$[Y](t) = \Big[\int_0^\cdot a(r)\,dr\Big](t) + 2\Big[\int_0^\cdot a(r)\,dr, \int_0^\cdot b\,dB\Big](t) + \Big[\int_0^\cdot b\,dB\Big](t) = 0+0+\int_0^t b(r)^2\,dr,$$

where the zeros on the right-hand side follow from Theorem 1.12, together with the established properties of $\int_0^\cdot a(r)\,dr$ and continuity of $\int_0^\cdot b\,dB$ (Theorem 10.2). #

Rather than building on Theorem 13.4, one proves Theorem 13.5 by *Doob-Meyer decomposition* of $(\int_0^t X\,dB)^2$, using an abstract notion of quadratic variation. In Section 15.1 we get at why this quadratic variation coincides with that in Section 1.2.

Theorem 13.7 (Doob-Meyer Decomposition) (Simple form.) (e.g., [22, Section 1.4-1.5])* *For a continuous submartingale $\{Y(t)\}_{t\in[0,T]}$ $[\{Y(t)\}_{t\geq 0}]$, there exist unique continuous stochastic processes $\{M(t)\}_{t\in[0,T]}$ and $\{A(t)\}_{t\in[0,T]}$ $[\{M(t)\}_{t\geq 0}$ and $\{A(t)\}_{t\geq 0}]$, such that M is a martingale with $M(0)=Y(0)$, A* (sometimes called the compensator) *is increasing and adapted, and*

$$Y(t) = M(t) + A(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0].$$

Corollary 13.8 *For a continuous martingale $\{Y(t)\}_{t\in[0,T]}$ $[\{Y(t)\}_{t\geq 0}]$ that is square-integrable, there exist unique continuous stochastic processes $\{M(t)\}_{t\in[0,T]}$ and $\{A(t)\}_{t\in[0,T]}$ $[\{M(t)\}_{t\geq 0}$ and $\{A(t)\}_{t\geq 0}]$, such that M is a martingale with $M(0)=Y(0)^2$, A is increasing and adapted, and*

$$Y(t)^2 = M(t) + A(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0].$$

Proof. This follows from Theorem 13.7 together with Exercise 55. □

Definition 13.9 *The quadratic variation $[Y]$ of a continuous and square-integrable martingale $\{Y(t)\}_{t\in[0,T]}$ $[\{Y(t)\}_{t\geq 0}]$, is defined by $[Y](t) = A(t)$, where A is the continuous, increasing and adapted process that satisfies*

$$Y(t)^2 = M(t) + A(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0],$$

where M is a continuous martingale with $M(0)=Y(0)^2$ (cf. Corollary 13.8).

Lemma 13.10 *For $X,Y\in E_T$ and $0\leq s<t\leq T$, we have* (cf. Theorem 10.2)

$$\mathbf{E}\Big\{\Big(\int_s^t X(r)\,dB(r)\Big)\Big(\int_s^t Y(\hat{r})\,dB(\hat{r})\Big)\Big|\mathcal{F}_s\Big\} = \mathbf{E}\Big\{\int_s^t X(r)Y(r)\,dr\Big|\mathcal{F}_s\Big\}.$$

Proof. By Definition 8.1 together with Exercise 68 and a limit argument (see Exercise 84 below), it is enough to prove the lemma for $X, Y \in S_T$. To that end, we can assume that X and Y have a common grid $0 = t_0 < t_1 < \ldots < t_n = T$, such that $s = t_k$ and $t = t_{\ell+1}$, $k \le \ell$, belong to that grid (see Lecture 7). We have

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\int_s^t X\,dB\Big)\Big(\int_s^t Y\,dB\Big)\Big|\mathcal{F}_s\Big\} \\
&\quad = \mathbf{E}\Big\{\Big(\sum_{i=k}^{\ell} X_{t_i}(B(t_{i+1})-B(t_i))\Big)\Big(\sum_{j=k}^{\ell} Y_{t_j}(B(t_{j+1})-B(t_j))\Big)\Big|\mathcal{F}_s\Big\} \\
&\quad = \mathbf{E}\Big\{\sum_{i=k}^{\ell} X_{t_i}Y_{t_i}(B(t_{i+1})-B(t_i))^2\Big|\mathcal{F}_s\Big\} \\
&\qquad\qquad + \mathbf{E}\Big\{\sum_{k\le i,j\le \ell,\, i\ne j} X_{t_i}Y_{t_j}(B(t_{i+1})-B(t_i))(B(t_{j+1})-B(t_j))\Big|\mathcal{F}_s\Big\} \\
&\quad = \mathbf{E}\Big\{\int_s^t X(r)Y(r)\,dr\Big|\mathcal{F}_s\Big\} \\
&\qquad\qquad + \sum_{k\le i<j\le \ell} \mathbf{E}\Big\{X_{t_i}Y_{t_j}(B(t_{i+1})-B(t_i))\,\mathbf{E}\Big\{B(t_{j+1})-B(t_j)\Big|\mathcal{F}_{t_j}\Big\}\Big|\mathcal{F}_s\Big\} \\
&\qquad\qquad + \sum_{k\le j<i\le \ell} \mathbf{E}\Big\{X_{t_i}Y_{t_j}(B(t_{j+1})-B(t_j))\,\mathbf{E}\Big\{B(t_{i+1})-B(t_i)\Big|\mathcal{F}_{t_i}\Big\}\Big|\mathcal{F}_s\Big\}.
\end{aligned}$$

Here the last two terms on the right-hand side vanish, since $B(t_{i+1}) - B(t_i)$ has zero mean and is independent of $\mathcal{F}_{t_i}$. (All conditional expectations involved are well-defined, by the argument at the end of the proof of Theorem 7.8.) □

EXERCISE 84 Explain how Definition 8.1 together with Exercise 68 show that it is enough to prove Lemma 13.10 for $X, Y \in S_T$.

Lemma 13.11 *For* $X \in E_T$, $T \in (0,\infty)$, *the process*

$$M(t) \equiv \Big(\int_0^t X(r)\,dB(r)\Big)^2 - \int_0^t X(r)^2\,dr, \quad t\in[0,T], \qquad \textit{is a martingale.}$$

Proof. Leaving the details concerning requirements on integrability, adaptedness and measurability to the reader (recall Exercise 64), Lemma 13.10 gives

$$\begin{aligned}
\mathbf{E}\{M(t)|\mathcal{F}_s\} &= \mathbf{E}\Big\{\Big(\int_s^t X\,dB\Big)^2 - \int_s^t X(r)^2\,dr\Big|\mathcal{F}_s\Big\} + \mathbf{E}\Big\{2\Big(\int_0^s X\,dB\Big)\Big(\int_s^t X\,dB\Big)\Big|\mathcal{F}_s\Big\} \\
&\qquad + \mathbf{E}\Big\{\Big(\int_0^s X\,dB\Big)^2 - \int_0^s X(r)^2\,dr\Big|\mathcal{F}_s\Big\} \\
&= 0 + 0 + M(s) \qquad \text{for} \quad 0 \le s < t. \quad \square
\end{aligned}$$

EXERCISE 85 Explain the two zeros at the last line of the proof of Lemma 13.11.

EXERCISE 86 Piece together the above findings to a proof of Theorem 13.5.

14 Fourteenth Lecture

14.1 Introduction to Itô's Formula

Theorem 14.1 (Itô's Formula) (Simple form.) *For a function* $f \in \underline{\mathbb{C}^2(\mathbb{R})} \equiv \{(g:\mathbb{R}\to\mathbb{R}) : g$ is two times continuously differentiable$\}$, *we have*

$$f(B(t)) = f(0) + \frac{1}{2}\int_0^t f''(B(r))\,dr + \int_0^t f'(B(r))\,dB(r) \qquad \textit{for} \quad t\geq 0.$$

Expressed in terms of stochastic differentials, this means that

$$df(B(t)) = (1/2)\, f''(B(t))\,dt + f'(B(t))\,dB(t).$$

Proof. For partitions $0=t_0<\ldots<t_n=t$, Taylor expansion and Theorem 11.12 give

$$\begin{aligned} f(B(t))-f(0) &= \sum_{i=1}^n (f(B(t_i))-f(B(t_{i-1}))) \\ &\sim \sum_{i=1}^n f'(B(t_{i-1}))\,(B(t_i)-B(t_{i-1})) + \frac{1}{2}\sum_{i=1}^n f''(B(t_{i-1}))\,(B(t_i)-B(t_{i-1}))^2 \\ &\to \int_0^t f'(B(r))\,dB(r) + \frac{1}{2}\int_0^t f''(B(r))\,dr \qquad \text{as} \quad \max\nolimits_{1\leq i\leq n} t_i - t_{i-1} \to 0 \end{aligned}$$

with convergence in probability, where $\sim$ means asymptotic equality. Here we got the second term on the right-hand side from Theorem 2.13 and uniform continuity of $f''(B(\cdot))$, employing a second less coarse grid as in Example 6.3. □

Example 14.2 Taking $f(x)=x^2$ in *Itô's formula*, we get (cf. Example 6.4)

$$d(B(t))^2 = df(B(t)) = (1/2)\, f''(B(t))\,dt + f'(B(t))\,dB(t) = dt + 2\,B(t)\,dB(t). \quad \#$$

Example 14.3 Let $X(t) = X_0\,\mathrm{e}^{B(t)}$, where $X_0\in\mathbb{R}$ is a constant. We have

$$dX(t) = df(B(t)) = (1/2)\,X_0\,e^{B(t)}\,dt + X_0\,e^{B(t)}\,dB(t) = (1/2)\,X(t)\,dt + X(t)\,dB(t). \quad \#$$

14.2 Itô's Formula

Theorem 14.4 (Itô's Formula) *Let* $\{a(t)\}_{t\in[0,T]}$ *and* $\{b(t)\}_{t\in[0,T]}$ *be stochastic processes, such that* $\sqrt{|a|}, b \in P_T$, *and define*

$$\xi(t) = \xi_0 + \int_0^t a(r)\,dr + \int_0^t b(r)\,dB(r) \qquad \textit{for} \quad t\in[0,T],$$

where $\xi_0\in\mathbb{R}$ *is a constant. For* $f\in\mathbb{C}^2(\mathbb{R})$ *we have*

$$df(\xi(t)) = \Big(a(t)\,f'(\xi(t)) + \frac{b(t)^2}{2}\,f''(\xi(t))\Big)\,dt + b(t)\,f'(\xi(t))\,dB(t).$$

Example 14.5 Taking $a=0$ and $b=1$, Theorem 14.4 gives Theorem 14.1. #

Proof of Theorem 14.4. The claim is that

$$f(\xi(t)) = f(0) + \int_0^t \Big(a(r)\, f'(\xi(r)) + \frac{b(r)^2}{2}\, f''(\xi(r))\Big)\, dr + \int_0^t b(r)\, f'(\xi(r))\, dB(r).$$

Since $\sqrt{|a|},\, b \in P_T$, we have $\sqrt{|af'(\xi)|},\, b\sqrt{|f''(\xi)|},\, bf'(\xi) \in P_T$, by the assumptions on f together with continuity of ξ. Hence Theorem 11.12 shows that

$$\begin{aligned}
&\int_0^t \Big(a(r)\, f'(\xi(r)) + \frac{b(r)^2}{2}\, f''(\xi(r))\Big)\, dr + \int_0^t b(r)\, f'(\xi(r))\, dB(r) \\
&\quad \leftarrow \sum_{i=1}^n \Big(a(t_{i-1})\, f'(\xi(t_{i-1})) + \frac{b(t_{i-1})^2}{2}\, f''(\xi(t_{i-1}))\Big)\, (t_i - t_{i-1}) \\
&\qquad\qquad + \sum_{i=1}^n b(t_{i-1})\, f'(\xi(t_{i-1}))\, (B(t_i) - B(t_{i-1})) \\
&\quad = \sum_{i=1}^n \Big(a(t_{i-1})\, (t_i - t_{i-1}) + b(t_{i-1})\, (B(t_i) - B(t_{i-1}))\Big) f'(\xi(t_{i-1})) \\
&\qquad\qquad + \sum_{i=1}^n \frac{b(t_{i-1})^2}{2}\, (t_i - t_{i-1})\, f''(\xi(t_{i-1})) \qquad \text{as} \quad \max_{1\le i\le n} t_i - t_{i-1} \to 0
\end{aligned}$$

(where $0 = t_0 < \ldots < t_n = t$), in the sense of convergence in probability. Of course, Theorem 11.12 also gives

$$\xi(t) - \xi_0 \leftarrow \sum_{i=1}^n \Big(a(t_{i-1})\, (t_i - t_{i-1}) + b(t_{i-1})\, (B(t_i) - B(t_{i-1}))\Big). \tag{14.1}$$

By introducing a second coarser grid, as in Example 6.3 (see also the proof of Theorem 14.1), and using (14.1) together with the continuity of $f'(\xi)$, it follows that

$$\begin{aligned}
&\sum_{i=1}^n \Big(a(t_{i-1})\, (t_i - t_{i-1}) + b(t_{i-1})\, (B(t_i) - B(t_{i-1}))\Big) f'(\xi(t_{i-1})) \\
&\qquad\qquad \sim \sum_{i=1}^n (\xi(t_i) - \xi(t_{i-1}))\, f'(\xi(t_{i-1})) \qquad \text{as} \quad \max_{1\le i\le n} t_i - t_{i-1} \to 0.
\end{aligned}$$

In the same way, again by means of using a second coarser grid, the quadratic variation result in Example 13.6, together with continuity of $f''(\xi)$, gives

$$\sum_{i=1}^n \frac{b(t_{i-1})^2}{2}\, (t_i - t_{i-1})\, f''(\xi(t_{i-1})) \sim \sum_{i=1}^n \frac{(\xi(t_i) - \xi(t_{i-1}))^2}{2}\, f''(\xi(t_{i-1})).$$

Adding things up, we therefore obtain

$$\begin{aligned}
&\int_0^t \Big(a(r)\, f'(\xi(r)) + \frac{b(r)^2}{2}\, f''(\xi(r))\Big)\, dr + \int_0^t b(r)\, f'(\xi(r))\, dB(r) \\
&\quad \leftarrow \sum_{i=1}^n \Big((\xi(t_i) - \xi(t_{i-1}))\, f'(\xi(t_{i-1})) + \frac{(\xi(t_i) - \xi(t_{i-1}))^2}{2}\, f''(\xi(t_{i-1}))\Big).
\end{aligned}$$

Here, by Taylor expansion, the right-hand side is asymptotically the same as

$$\sum_{i=1}^n \Big(f(\xi(t_i)) - f(\xi(t_{i-1}))\Big) = f(\xi(t)) - f(\xi(0)) = f(\xi(t)) - f(\xi_0),$$

which establishes *Itô's formula*. (It is a consequence of elementary Taylor expansion arguments, together with continuity of ξ, that the convergence of the sum with the second order terms, ensures that higher order terms are not needed.) □

*Remark 14.6 Instead of relying on the unproved Theorem 11.12, we can get *Itô's formula*, for $\sqrt{|a|}, b \in E_T$, using Section 10.3. For this, one approximates the Itô integral processes $\int_0^t a(r)\, f'(\xi(r))\, dr$, $\int_0^t \frac{1}{2} b(r)^2 f''(\xi(r))\, dr$ and $\int_0^t b(r)\, f'(\xi(r))\, dB(r)$, with bounded and continuous processes in E_T, as in the proof of Theorem 11.1. Then a and b are approximated by the bounded and continuous processes $a^{(N)}$ and $b^{(N)}$, defined as in the proof of Theorem 9.1, and employ Section 10.3. #

Theorem 14.7 (Itô's Formula) (e.g., [27, Theorem 4.4])* *Let* $\{a(t)\}_{t\in[0,T]}$ *and* $\{b(t)\}_{t\in[0,T]}$ *be stochastic processes, such that* $\sqrt{|a|}, b \in P_T$, *and define*

$$\xi(t) = \xi_0 + \int_0^t a(r)\, dr + \int_0^t b(r)\, dB(r) \qquad \textit{for} \quad t\in[0,T],$$

with $\xi(0)=\xi_0$ *a constant. For* $f \in \underline{\mathbb{C}^{1,2}([0,T]\times\mathbb{R})} \equiv \{(g:[0,T]\times\mathbb{R}\to\mathbb{R}) : g$ has continuous partial derivatives $\partial_1 g = \frac{\partial g(t,x)}{\partial t}$, $\partial_2 g = \frac{\partial g(t,x)}{\partial x}$ *and* $\partial_2^2 g = \frac{\partial^2 g(t,x)}{\partial x^2}\}$ *we have*

$$\begin{aligned} f(t,\xi(t)) = f(0,\xi(0)) &+ \int_0^t \Big(\partial_1 f(r,\xi(r)) + a(r)\,\partial_2 f(r,\xi(r)) + \frac{b(r)^2}{2}\,\partial_2^2 f(r,\xi(r))\Big)\, dr \\ &+ \int_0^t b(r)\,\partial_2 f(r,\xi(r))\, dB(r) \qquad \textit{for} \quad t\in[0,T]. \end{aligned}$$

Example 14.8 Taking $f(t,x)=f(x)$, Theorem 14.7 gives Theorem 14.4. #

14.3 Examples of Stochastic Calculus

Example 14.9 We want to solve the (Itô sense) SDE

$$dX(t) = (1/2)\, X(t)\, dt + X(t)\, dB(t) \qquad \text{for} \quad t\in[0,T].$$

This means that we are looking for a process X that satsifies

$$X(t) = X_0 + \int_0^t (1/2)\, X(r)\, dr + \int_0^t X(r)\, dB(r) \qquad \text{for} \quad t\in[0,T], \tag{14.2}$$

where $\sqrt{|X|}, X \in P_T$ is required for the integrals to be well-defined. We try *Itô's formula* Theorem 14.4 for $f(t) = \ln(t)$. Since $a=X/2$ and $b=X$ [cf. (14.2)], we get

$$\begin{aligned} d(\ln(X(t))) &= \Big(a(t)\, f'(X(t)) + \frac{b(t)^2}{2}\, f''(X(t))\Big)\, dt + b(t)\, f'(X(t))\, dB(t) \\ &= \Big(\frac{X(t)}{2}\,\frac{1}{X(t)} - \frac{X(t)^2}{2}\,\frac{1}{X(t)^2}\Big)\, dt + X(t)\,\frac{1}{X(t)}\, dB(t) = dB(t). \end{aligned}$$

This gives $\ln(X(t))-\ln(X_0) = B(t)$, so that $X(t) = X_0\,\mathrm{e}^{B(t)}$. These calculations are only formal, because we do not even know that the process X that solves the equation exists, and much less that, for example, X is strictly positive (ln ...).

Given the candidate $X(t) = X_0\,\mathrm{e}^{B(t)}$ to a solution (not necessarily positive), we now have to check that the SDE is satisfied. However, this is Example 14.3. #

Example 14.10 In later lectures, we will consider stochastic integrals wrt. other processes that BM, in a systematic way. However, already now, such integrals can be introduced by means of stochastic differentials.

Let $\{a(t)\}_{t\in[0,T]}$ and $\{b(t)\}_{t\in[0,T]}$ be processes, such that $\sqrt{|a|}, b\in P_T$, and let

$$dY(t) = a(t)\,dt + b(t)\,dB(t) \qquad \text{for} \quad t\in[0,T].$$

Provided that $\{Z(t)\}_{t\in[0,T]}$ is a process, such that $\sqrt{|aZ|}, bZ\in P_T$, we may define

$$\int_0^t Z(r)\,dY(r) \equiv \int_0^t Z(r)\,a(r)\,dr + \int_0^t Z(r)\,b(r)\,dB(r) \qquad \text{for} \quad t\in[0,T].$$

Moreover, by inspection of the proof of Theorem 14.4, for $\sqrt{|a|}, b\in E_T$, we may, in the sense explained in Remark 14.6 (see also Section 10.3), evaluate the integrals as limits of approximating (Itô sense) sums. #

Example 14.11 (Integration by Parts) Consider two stochastic processes $\{X(t)\}_{t\in[0,T]}$ and $\{Y(t)\}_{t\in[0,T]}$, that are both stochastic differentials (see Example 14.10). In real analysis, we have the integration by parts formula

$$\int_0^t X(r)\,dY(r) = X(t)Y(t) - X(0)Y(0) - \int_0^t Y(r)\,dX(r).$$

In case of stochastic Itô integration, the corresponding formula is

$$\int_0^t X(r)\,dY(r) = X(t)Y(t) - X(0)Y(0) - \int_0^t Y(r)\,dX(r) - [X,Y](t).$$

Invoking Theroem 11.12, together with Theorem 15.3 and Definition 15.4 below, this follows by means of checking the corresponding approximating sums

$$\begin{aligned}
&\int_0^t X(r)\,dY(r) + \int_0^t Y(r)\,dX(r)\\
&= \text{P-lim}\sum_{i=1}^n \Big(X(t_{i-1})\,(Y(t_i)-Y(t_{i-1})) + Y(t_{i-1})\,(X(t_i)-X(t_{i-1}))\Big)\\
&= \text{P-lim}\sum_{i=1}^n \Big(X(t_i)Y(t_i)-X(t_{i-1})Y(t_{i-1}) - (X(t_i)-X(t_{i-1}))\,(Y(t_i)-Y(t_{i-1}))\Big)\\
&= X(t)Y(t)-X(0)Y(0) - [X,Y](t). \quad \#
\end{aligned}$$

15 Fifteenth Lecture

15.1 Quadratic Variation for Local Martingales

In the next section, we establish a quite general Itô formula. This requires a more abstract approach to quadratic variation, via continuous local martingales.

Theorem 15.1 (e.g., [33, Chapter IV, Theorem 1.8])* *For a continuous local martingale $\{Y(t)\}_{t\in[0,T]}$ [$\{Y(t)\}_{t\geq 0}$], there exist unique continuous stochastic processes $\{M(t)\}_{t\in[0,T]}$ and $\{A(t)\}_{t\in[0,T]}$ [$\{M(t)\}_{t\geq 0}$ and $\{A(t)\}_{t\geq 0}$], such that M is a local martingale with $M(0)=Y(0)^2$, A is increasing and adapted, and*

$$Y(t)^2 = M(t) + A(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0].$$

Moreover, we have

$$A(t) = \text{P-lim}\left\{\sum_{i=1}^{n}(Y(t_i)-Y(t_{i-1}))^2 : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}.$$

The first part of the theorem can be proved by *Doob-Meyer decomposition* of submartingales (Theorem 13.7), and a clever use of the stopping times from the definition of local martingales. Similarly, the second part of the theorem follows from a corresponding (non-local) martingale theorem, which we discuss in later lectures.

Definition 15.2 *The quadratic variation $[Y]$ of a continuous local martingale $\{Y(t)\}_{t\in[0,T]}$ [$\{Y(t)\}_{t\geq 0}$], is defined by $[Y](t) = A(t)$, where A is the continuous, increasing and adapted process that satisfies*

$$Y(t)^2 = M(t) + A(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0],$$

where M is a continuous local martingale with $M(0)=Y(0)^2$ (cf. Theorem 15.1).

Theorem 15.3 (e.g., [33, Chapter IV, Theorem 1.9])* *For two continuous local martingales $\{Y(t)\}_{t\in[0,T]}$ and $\{Z(t)\}_{t\in[0,T]}$ [$\{Y(t)\}_{t\geq 0}$ and $\{Z(t)\}_{t\geq 0}$] (wrt. a common filtration), there exist unique continuous stochastic processes $\{M(t)\}_{t\in[0,T]}$ and $\{C(t)\}_{t\in[0,T]}$ [$\{M(t)\}_{t\geq 0}$ and $\{C(t)\}_{t\geq 0}$], such that M is a local martingale with $M(0)=Y(0)Z(0)$, C has finite variation and is adapted, and*

$$Y(t)Z(t) = M(t) + C(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0].$$

Moreover, we have

$$C(t) = \text{P-lim}\left\{\sum_{i=1}^{n}(Y(t_i)-Y(t_{i-1}))\,(Z(t_i)-Z(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}.$$

Definition 15.4 *The covariation $[Y, Z]$ between two continuous local martingales $\{Y(t)\}_{t\in[0,T]}$ and $\{Z(t)\}_{t\in[0,T]}$ [$\{Y(t)\}_{t\geq 0}$ and $\{Z(t)\}_{t\geq 0}$] (wrt. a common filtration), is defined by $[Y, Z](t) = C(t)$, where C is the continuous and adapted process, with finite variation, that satisfies*

$$Y(t)Z(t) = M(t) + C(t) \qquad \text{for} \quad t\in[0,T] \quad [t\geq 0],$$

where M is a continuous local martingale with $M(0)=Y(0)Z(0)$ (cf. Theorem 15.3).

Exercise 87 Show that the sum of two local martingales (wrt. a common filtration), is again a local martingale. What about the difference?

Exercise 88 Derive the existence part of Theorem 15.3 from Theorem 15.1. (Uniqueness is more difficult.)

Since the Itô integrals $\int_0^t X\,dB$ and $\int_0^t Y\,dB$, of $X, Y \in P_T$, are local martingales (Theorem 12.10), their quadratic variations $[X]$ and $[Y]$ are well-defined, by Definition 15.2, as is their covariation $[X, Y]$, by Definition 15.4.

Theorem 15.5 *For $X\in P_T$, the Itô integral process has quadratic variation*

$$\Big[\int_0^\cdot X\,dB\Big](t) = \int_0^t X(r)^2\,dr \qquad \text{for} \quad t\in[0,T].$$

The proof is a typical demonstration of localization at work:

Proof of Theorem 15.5. Writing

$$\tau_n \equiv n\wedge \inf\Big\{t\geq 0 : \int_0^t X(r)^2\,dr \geq n\Big\} \qquad \text{and} \qquad X_n(t)\equiv X(t)I_{\{t\leq\tau_n\}},$$

we have $X_n\in E_\infty$ (cf. the proof of Theorem 12.10), so that, by Theorem 13.5,

$$\Big[\int_0^\cdot X\,dB\Big](\tau_n\wedge t) = \Big[\int_0^\cdot X_n\,dB\Big](t) = \int_0^t X_n(r)^2\,dr = \int_0^{\tau_n\wedge t} X(r)^2\,dr.$$

Sending $n\to\infty$, the theorem follows (since $\tau_n\to\infty$, by continuity of X). □

Theorem 15.6 *For $X, Y\in P_T$, the Itô integral has covariation*

$$\Big[\int_0^\cdot X\,dB, \int_0^\cdot Y\,dB\Big](t) = \int_0^t X(r)Y(r)\,dr \qquad \text{for} \quad t\in[0,T].$$

Theorem 15.7 *For $\sqrt{|a_1|}, \sqrt{|a_2|}, b_1, b_2 \in P_T$, we have*

$$\Big[\int_0^\cdot a_1(r)\,dr + \int_0^\cdot b_1\,dB, \int_0^\cdot a_2(r)\,dr + \int_0^\cdot b_2\,dB\Big](t) = \int_0^t b_1(r)b_2(r)\,dr \qquad \text{for} \quad t\in[0,T].$$

Corollary 15.8 *Let* $\sqrt{|a_1|}, \sqrt{|a_2|}, b_1, b_2 \in P_T$. *For the stochastic differentials*

$$dX(t) = a_1(t)\,dt + b_1(t)\,dB(t) \qquad \textit{and} \qquad dY(t) = a_2(t)\,dt + b_2(t)\,dB(t)$$

we have

$$\underline{dX(t)dY(t)} \equiv d[X,Y](t) = b_1(t)b_2(t)\,dt.$$

EXERCISE 89 Prove Theorems 15.6 and 15.7, by Theorem 15.5 and polarization.

15.2 Itô's Formula (continued)

Theorem 15.9 (ITÔ'S FORMULA) (e.g., [33, Chapter IV, Theorem 3.3])* *Let* $\sqrt{|a_1|}, \sqrt{|a_2|}, b_1, b_2 \in P_T$. *For the stochastic differentials*

$$dX(t) = a_1(t)\,dt + b_1(t)\,dB(t) \qquad \textit{and} \qquad dY(t) = a_2(t)\,dt + b_2(t)\,dB(t),$$

and a function $f \in \mathbb{C}^2(\mathbb{R}^2) \equiv \{(g : \mathbb{R} \to \mathbb{R}) : g$ has continuous second order partial derivatives$\}$, *we have*

$$\begin{aligned}
&df(X(t), Y(t)) \\
&= \partial_1 f(X,Y)\,dX(t) + \partial_2 f(X,Y)\,dY(t) \\
&\qquad +\frac{1}{2}\partial_1^2 f(X,Y)\,dX(t)^2 + \partial_1\partial_2 f(X,Y)\,dX(t)dY(t) + \frac{1}{2}\partial_2^2 f(X,Y)\,dY(t)^2 \\
&= \partial_1 f(X,Y)\,dX(t) + \partial_2 f(X,Y)\,dY(t) \\
&\qquad +\frac{1}{2}\partial_1^2 f(X,Y)\,d[X](t) + \partial_1\partial_2 f(X,Y)\,d[X,Y](t) + \frac{1}{2}\partial_2^2 f(X,Y)\,d[Y](t) \\
&= \partial_1 f(X,Y)\,dX(t) + \partial_2 f(X,Y)\,dY(t) \\
&\qquad +\frac{1}{2}\partial_1^2 f(X,Y)\,b_1(t)^2\,dt + \partial_1\partial_2 f(X,Y)\,b_1(t)b_2(t)\,dt + \frac{1}{2}\partial_2^2 f(X,Y)\,b_2(t)^2\,dt.
\end{aligned}$$

Using Corollary 15.8, this general Itô formula is proved in the same way as Theorem 14.4. It has an obvious multidimensional extension to functions $f : \mathbb{R}^n \to \mathbb{R}$. Some authors (e.g., the influential [22]) call Itô's formula Itô's rule.

EXERCISE 90 Show how Theorem 14.7 follows from Theorem 15.9.

Example 15.10 Taking $f(x,y) = xy$ in Theorem 15.9, we get

$$d(X(t)Y(t)) = Y(t)\,dX(t) + X(t)\,dY(t) + d[X,Y](t).$$

Expressed in terms of integrals, this recovers Example 14.11 (see also Example 14.10)

$$X(t)Y(t) - X(0)Y(0) = \int_0^t X(r)\,dY(r) + \int_0^t Y(r)\,dX(r) + [X,Y](t). \quad \#$$

15.3 Stochastic Exponentials

Definition 15.11 *Let* $\sqrt{|a|}, b \in P_T$, *and put*

$$dX(t) = a(t)\,dt + b(t)\,dB(t) \qquad \text{for} \quad t \geq 0.$$

The *stochastic exponential* $\mathcal{E}X$ *of* X *is given by*

$$(\mathcal{E}X)(t) \equiv \mathrm{e}^{X(t)-X(0)-[X](t)/2} = \mathrm{e}^{X(t)-X(0)-\frac{1}{2}\int_0^t b(r)^2\,dr} \qquad \text{for} \quad t \geq 0.$$

Theorem 15.12 *Let* $\sqrt{|a|}, b \in P_T$, *and put*

$$dX(t) = a(t)\,dt + b(t)\,dB(t) \qquad \text{for} \quad t \geq 0.$$

The stochastic exponential $\mathcal{E}X$ *of* X *satisfies the SDE*

$$d(\mathcal{E}X)(t) = (\mathcal{E}X)(t)\,dX(t) \qquad \text{for} \quad t \geq 0, \qquad (\mathcal{E}X)(0) = 1.$$

Proof. Using Itô's formula on $f(X,Y)$, where $f(x,y) = \mathrm{e}^{x-y}$ and $Y(t) = X(0) + \frac{1}{2}[X,X](t)$, so that $dY = \frac{1}{2}d[X,X]$ and $d[Y,Y] = d[X,Y] = 0$, we obtain

$$d(\mathcal{E}X) = df(X,Y) = \mathcal{E}X\,\Big(dX - dY + d[X]/2 - d[X,Y]/2 + d[Y]/2\Big) = \mathcal{E}X\,dX. \quad \square$$

Exercise 91 Why is $d[X,Y] = d[Y,Y] = 0$ in the proof of Theorem 15.12?

Example 15.13 The stochastic exponential of BM is $(\mathcal{E}B)(t) = \mathrm{e}^{B(t)-B(0)-t/2}$. #

Example 15.14 (Black-Scholes Formula) In *Mathematical Finance*, the *stock price* S is the stochastical exponential of the *return* R, which is given by

$$dR(t) = \mu\,dt + \sigma\,dB(t) \qquad \text{where} \quad \mu \in \mathbb{R} \quad \text{and} \quad \sigma \geq 0 \quad \text{are constants.}$$

Inserting in Definition 15.11, the stock price becomes

$$S(t) = (\mathcal{E}R)(t) = \mathrm{e}^{R(t)-R(0)-[R](t)/2} = \mathrm{e}^{\mu t + \sigma B(t) - \sigma^2 t/2}. \quad \#$$

As we will see in Section 18.3, stochastic exponentials are fundamental for *Girsanov's Theorem*, which is one of the more important results in stochastic calculus.

15.4 Introduction to Diffusion Type SDE wrt. BM

Definition 15.15 *A* *diffusion type SDE* *(or* *diffusion*) *with* *drift* $\mu : \mathbb{R}^2 \to \mathbb{R}$ *and* *dispersion* $\sigma : \mathbb{R}^2 \to \mathbb{R}$, *which are measurable functions, is given by*

$$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dB(t).$$

A more general SDE (than one of diffusion type), is given by

$$dX(t) = \mu(t, \mathcal{F}_t)\, dt + \sigma(t, \mathcal{F}_t)\, dB(t) \qquad \text{(with obvious notation)}.$$

However, diffusions are still very versatile. Moreover, as we shall see, they have an array of remarkable and useful properties. Argubly, diffusions are the most important objects of probability theory. Certainly, they are the most studied ones.

We will almost exclusively deal with SDE of diffusion type, which we will simply refer to as SDE (rather than SDE of diffusion type).

Definition 15.16 *Let $X_0 \in \mathbb{R}$ be a constant. A strong solution to the SDE*

$$dX(t) = \mu(t, X(t))\, dt + \sigma(t, X(t))\, dB(t) \qquad \text{for} \quad t \in [0,T], \qquad X(0) = X_0,$$

is a stochastic process $\{X(t)\}_{t\in[0,T]}$, such that $\sqrt{|\mu(\cdot, X(\cdot))|}, \sigma(\cdot, X(\cdot)) \in P_T$ and

$$X(t) = X_0 + \int_0^t \mu(r, X(r))\, dr + \int_0^t \sigma(r, X(r))\, dB(r) \qquad \text{for} \quad t \in [0,T].$$

Definition 15.17 (Simplified definition.) *Let $X_0 \in \mathbb{R}$ be a constant. A stochastic process $\{X(t)\}_{t\in[0,T]}$ is a weak solution to the SDE*

$$dX(t) = \mu(t, X(t))\, dt + \sigma(t, X(t))\, dB(t) \qquad \text{for} \quad t \in [0,T], \qquad X(0) = X_0,$$

if $\{X(t)\}_{t\in[0,T]} =_{\texttt{same fidi's}} \{\hat{X}(t)\}_{t\in[0,T]}$, where $\sqrt{|\mu(\cdot, \hat{X}(\cdot))|}, \sigma(\cdot, \hat{X}(\cdot)) \in P_T$, with

$$\hat{X}(t) = X_0 + \int_0^t \mu(r, \hat{X}(r))\, dr + \int_0^t \sigma(r, \hat{X}(r))\, d\hat{B}(r) \qquad \text{for} \quad t \in [0,T],$$

for some BM $\hat{B}$ (possibly defined on another probability space than B).

Each strong solution is a weak solution. However, a weak solution is a strong solution on some possibly different probability space. We will return to this later.

Definition 15.18 *Any solution to a diffusion type SDE is a diffusion process.*

Example 15.19 (Langevin Equation) Consider the SDE

$$dX(t) = (\alpha - \beta\, X(t))\, dt + \sigma\, dB(t) \qquad \text{for} \quad t \geq 0, \tag{15.1}$$

where $\alpha, \sigma \in \mathbb{R}$ and $\beta > 0$ are constants. Writing $Y(t) = e^{\beta t}(X(t) - \alpha/\beta)$, Theorem 14.7 with $f(t,x) = e^{\beta t}(x - \alpha/\beta)$, $a(t) = (\alpha - \beta\, X(t))$ and $b(t) = \sigma$, gives

$$dY(t) = \beta\, Y\, dt + e^{\beta t}\Big((\alpha - \beta\, X(t))\, dt + \sigma\, dB(t)\Big) = \sigma\, e^{\beta t}\, dB(t).$$

Hence we have

$$Y(t) = Y(0) + \int_0^t \sigma\, e^{\beta r}\, dB(r) = X(0) - \alpha/\beta + \int_0^t \sigma\, e^{\beta r}\, dB(r),$$

so that

$$X(t) = \alpha/\beta + e^{-\beta t} Y(t) = \alpha/\beta + e^{-\beta t}\Big(X(0) - \alpha/\beta + \int_0^t \sigma\, e^{\beta r}\, dB(r)\Big). \qquad (15.2)$$

The above calculations are only formal, since we do not know that a process X that solves the equation exists. Thus we have only derived necessary requirements for such a solution. However, it is a straightforward matter to use Itô's formula to check that the formal solution really is a (strong) solution to the SDE.

A solution X to the *Langevin equation* (15.1) (where usually $\alpha = 0$), is an Ornstein-Uhlenbeck process. However, in non-Markov literature, X is called an Ornstein-Uhlenbek process only when it is started according to the *stationary distribution* (see Definition 25.4 and Example 25.9 below), so that X is stationary. #

Example 15.20 (LINEAR SDE) The general linear SDE of diffusion type

$$dX(t) = \big(\alpha(t) + \beta(t)X(t)\big)\, dt + \big(\gamma(t) + \delta(t)X(t)\big)\, dB(t), \qquad (15.3)$$

where $\alpha, \beta, \gamma, \delta : \mathbb{R} \to \mathbb{R}$ are locally integrable functions, has an explicit strong solution. See for example [21, Proposition 18.2]* and [25, p. 123]*. Due to the resemblance with the *Langevin equation* in Example 15.19, which (15.3) generalizes, the solution X is sometimes called an inhomogeneous Ornstein-Uhlenbeck process.

In the particular case when $\alpha = \gamma = 0$, we have

$$dX(t) = \beta X\, dt + \delta X\, dB(t) = X\, dY(t) \qquad \text{where} \qquad dY(t) = \beta\, dt + \delta\, dB(t).$$

It follows that $X(t)/X(0)$ is a stochastic exponential of $Y(t)$, so that

$$X(t) = X(0)\, \exp\Big\{Y(t) - \frac{[Y,Y](t)}{2}\Big\} = X(0)\, \exp\Big\{\int_0^t \Big(\beta(r) - \frac{\delta(r)^2}{2}\Big)\, dr + \int_0^t \delta\, dB\Big\}.$$

The solution of (15.3) for general α and γ is only notationally more complicated. #

A measurable function $f : T \to \mathbb{R}$, $T \subseteq \mathcal{B}(\mathbb{R}^n)$, is locally integrable if $\int_B |f(x)|\, dx < \infty$ for each bounded set $B \in \mathcal{B}(\mathbb{R}^n)$.

16 Sixteenth Lecture

16.1 Itô's Formula for Local Martingales

Theory in this section has been gathered from [33, Chapter IV, Sections 1-3].

Pick a constant $T \in (0,\infty)$. Let $\{X(t)\}_{t\in[0,T]}$ be a continuous local martingale, and $\{a(t)\}_{t\in[0,T]}$ and $\{b(t)\}_{t\in[0,T]}$ progressively measurable processes such that

$$\int_0^T |a(r)|\,dr < \infty \qquad \text{and} \qquad \int_0^T b(r)^2\,d[X,X](r) < \infty \qquad \text{with probability one}$$

(pathwise Lebesgue integral). In particular, this holds if a and b also are continuous local martingales (recall Theorem 8.8). The integral processes

$$\Big\{\int_0^t a(r)\,dr\Big\}_{t\in[0,T]} \qquad \text{and} \qquad \Big\{\int_0^t b(r)\,dX(r)\Big\}_{t\in[0,T]} \qquad \text{are well-defined,}$$

in the sense of a Lebesgue integral and an *Itô integral*, respectively. The first of these processes has finite variation, while the second is a continuous local martingale. From this we get the first of the following two equalities (cf. Exercise 89)

$$\Big[\int_0^\cdot a(r)\,dr+\int_0^\cdot b\,dX,\int_0^\cdot a(r)\,dr+\int_0^\cdot b\,dX\Big](t) = \Big[\int_0^\cdot b\,dX,\int_0^\cdot b\,dX\Big](t) = \int_0^t b^2\,d[X,X]$$

for $t\in[0,T]$. As for the Itô integral wrt. BM, we have (cf. Theorem 11.12)

$$\int_0^t b\,dX \;=\; \text{P-lim}\Big\{\sum_{i=1}^n b(t_{i-1})\,(X(t_i)-X(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ n\in\mathbb{N},\ \max_{1\le i\le n} t_i-t_{i-1}\to 0\end{array}\Big\}.$$

In Section 24.1, we see that these properties follow from the corresponding properties of Itô integrals wrt. BM, since Itô integrals wrt. continuous local martingales can be transformed to integrals wrt. BM, by random change of time.

Now let $\{X_i(t)\}_{t\in[0,T]}$, $\{a_i(t)\}_{t\in[0,T]}$, $\{b_i(t)\}_{t\in[0,T]}$, $i=1,\ldots,n$, be continuous local martingales, wrt. a common filtration. We have

$$\Big[\int_0^\cdot a_i(r)\,dr+\int_0^\cdot b_i\,dX_i,\int_0^\cdot a_j(r)\,dr+\int_0^\cdot b_j\,dX_j\Big](t) = \int_0^t b_ib_j\,d[X_i,X_j].$$

For the stochastic differentials (defined by means of integrals)

$$dY(t) = \big(dY_1(t),\ldots,dY_n(t)\big) = \big(a_1(t)\,dt+b_1(t)\,dX_1(t),\ \ldots,\ a_n(t)\,dt+b_n(t)\,dX_n(t)\big),$$

we therefore have (again expressing integral statements with differentials)

$$d[Y_i,Y_j](t) = b_i(t)b_j(t)\,dt \qquad \text{for} \quad i,j=1,\ldots,n.$$

For $f\in \underline{\mathbb{C}^2(\mathbb{R}^n)} \equiv \{(g:\mathbb{R}^n\to\mathbb{R}) : g$ has continuous second order partial derivatives$\}$, we have <u>Itô's formula</u>

$$\begin{aligned} df(Y) &= \sum_{i=1}^n \partial_i f(Y)\,dY_i(t) && +\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \partial_i\partial_j f(Y)\,d[Y_i,Y_j](t) \\ &= \sum_{i=1}^n \partial_i f(Y)\,\big(a_i\,dt+b_i\,dX_i(t)\big) &&+\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \partial_i\partial_j f(Y)\,b_ib_j\,d[X_i,X_j](t). \end{aligned}$$

In particular, $f(Y)$ is again a continuous local martingale (compare with Example 14.10). [Of course, for $n=2$ and $X_1=X_2=B$ (BM), this reduces to Theorem 15.9.]

Example 16.1 (INTEGRATION BY PARTS) Let $dX(t)$ and $dY(t)$ be stochastic differentials, in the above sense. Taking $f(x,y)=xy$ in Itô's formula, we get

$$d(X(t)Y(t)) = Y(t)\,dX(t) + X(t)\,dY(t) + d[X,Y](t).$$

Of course, expressed in terms of integrals, this means that

$$X(t)Y(t) - X(0)Y(0) = \int_0^t X(r)\,dY(r) + \int_0^t Y(r)\,dX(r) + [X,Y](t). \quad \#$$

Example 16.2 (STOCHASTIC EXPONENTIAL) The stochastic exponential $\mathcal{E}X$ of a continuous local martingale X [or more generally, of a continuous semimartingale (see Section 24.2)], is given by $(\mathcal{E}X)(t) = \mathrm{e}^{X(t)-X(0)-[X](t)/2}$. Obviously, $(\mathcal{E}X)(0)=1$, and using *Itô's formula* from above [or more generally, from Section 24.2], as in the proof of Theorem 15.12, we get that $d(\mathcal{E}X)=\mathcal{E}X\,dX$. #

At [33, p. 139] of their very densly written and famous 533 pages long book, the authors state that *"To some extent, the whole sequel of this book is but an unending series of applications of Itô's formula."* One may thus safely say that Itô's formula is one of the few most important results in probability.

16.2 Beginning Values for SDE

In Section 15.4, we considered a SDE

$$X(t) = X(0) + \int_0^t \mu(r,X(r))\,dr + \int_0^t \sigma(r,X(r))\,dB(r) \qquad \text{for} \quad t\in[0,T],$$

with constant beginning value $X(0)=X_0\in\mathbb{R}$. In general, it is desirable to allow a random beginning value $X(0)=X_0$, that is independent of B.

But the defintion of the Itô integral requires that $\sigma(\cdot,X(\cdot))\in P_T$, which in particular means that $\sigma(\cdot,X(\cdot))$ is adapted, so that

$$\sigma(0,X(0)) \in \mathcal{F}_0 = \sigma(B(0))_{\text{augmented}} = \sigma(0)_{\text{augmented}} = \{\emptyset,\Omega\}_{\text{augmented}}.$$

Hence $\sigma(0,X(0))$ is constant a.s., so that $X(0)$ cannot be random (except on a null-event), for a general σ. One can handle this is the following (e.g., [33, Chapter IX])*:

(1) *Solve a SDE for all non-random beginning values x_0 in the non-random range $X_0(\Omega)$ of the random beginning value X_0. Substitute the non-random beginning value x_0 with the random X_0*, that is independent of the solutions started at $x_0\in X_0(\Omega)$, due to the adaptedness of the solution (see the proof of Theorem 16.8 below).

On might draw a parallell with the study of a Markov process X by means of its transition probabilities, which specify the development of X for every possible non-random beginning value of X [and impossible ones as well (recall Remark 3.8)].

We handle the problem in a (seemingly) different way, following [22, Chapter 5]*:

(2.a) *Associate with the given BM B an augmented filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$, such that B is adapted to $\mathbb{F}$,* but that is not necessarily $\{\sigma(B(r): 0\leq r\geq t)_{\text{augmented}}\}_{t\geq 0}$.
(2.b) *As long as $B(t)-B(s)$ is independent of $\mathcal{F}_s$ for $0\leq s<t$, results and proofs for the Itô integral $\int_0^t X\,dB$ of $X\in P_T$ remain valid.* Of course, now P_T is the class of measurable processes that are adapted to $\mathbb{F}$, with $\int_0^T X(r)^2\,dr<\infty$ a.s.

An important, as we shall see, particular case of this approach is the following:

(3.a) *Associate with B the augmented filtration generated by X_0 and B*

$$\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0} = \Big\{\sigma\Big(\sigma(B(r): 0\leq r\geq t), \sigma(X_0)\Big)_{\text{augmented}}\Big\}_{t\geq 0}. \tag{16.1}$$

(3.b) *Results and proofs for the Itô integral $\int_0^t X\,dB$ of $X\in P_T$ remain valid, since $B(t)-B(s)$ is independent of $\mathcal{F}_s$ for $0\leq s<t$,* because X_0 is independent of B.

We use **(2)** to define weak solution, and **(3)** for strong solutions, of a SDE.

Definition 16.3 *Let X_0 be a random variable that is independent of B. Associate with B the augmented filtration generated by X_0 and B [*cf. (16.1)*]. A strong solution to the SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \textit{for} \quad t\in[0,T], \qquad X(0)=X_0,$$

is a stochastic process $\{X(t)\}_{t\in[0,T]}$, such that $\sqrt{|\mu(\cdot,X(\cdot))|}, \sigma(\cdot,X(\cdot)) \in P_T$ and

$$X(t) = X_0 + \int_0^t \mu(r,X(r))\,dr + \int_0^t \sigma(r,X(r))\,dB(r) \qquad \textit{for} \quad t\in[0,T].$$

Definition 16.4 *Let X_0 be a random variable that is independent of B. Consider a random variable $\tilde{X}_0 =_{\text{distribution}} X_0$, together with a BM $\tilde{B}$ that is independent of $\tilde{X}_0$, on some probability space. Associate with $\tilde{B}$ an augmented filtration that contains the filtration generated by $\tilde{X}_0$ and $\tilde{B}$* [with the independence property specified in **(2.b)**]. *A stochastic process $\{\tilde{X}(t)\}_{t\in[0,T]}$ is a weak solution to the SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \textit{for} \quad t\in[0,T], \qquad X(0)=X_0,$$

if

$$\tilde{X}(t) = \tilde{X}_0 + \int_0^t \mu(r,\tilde{X}(r))\,dr + \int_0^t \sigma(r,\tilde{X}(r))\,d\tilde{B}(r) \qquad \textit{for} \quad t\in[0,T].$$

The important difference between a strong solution and a weak one, is that a strong solution $X(t)$ is a "function" (functional) of X_0 together with BM B until time t (by being adapted to the σ-algebra generated by them), while a weak solution, which is not strong, does not have this functional relation with X_0 and B.

If a strong solution exists on a particular probability space, large enough to "house" the filtration $\mathbb{F}$ given by (16.1), then it exists on every such probability space, that is, for every given BM with such a filtration.

For a SDE that lacks a strong solution (e.g., Example 17.5 below), but has a weak one, there will not exist a weak solution on every probability space that houses "copies" $\tilde{X}_0$ and $\tilde{B}$ of X_0 and B (see Definition 16.4), together with a filtration to which they are adapted [with the independence property specified in **(2.b)**]. For example, there will not exist a weak solution (on that probability space), if the filtration is the augmented one generated by $\tilde{X}_0$ and $\tilde{B}$ (since this would be a strong solution).

16.3 Uniqueness of Solutions to SDE

There is some variation in the definitions of uniqueness of solutions to a SDE in the literature. We follow [22, Chapter 5]:

Definition 16.5 *A SDE has unique strong solution, if for any pair of strong solutions* $\{X_1(t)\}_{t\in[0,T]}$ *and* $\{X_2(t)\}_{t\in[0,T]}$ *to it* (defined on a common probability space), *we have*

$$\mathbf{P}\Big\{X_1(t)=X_2(t) \ \ \textit{for all} \ \ t\in[0,T]\Big\} = 1.$$

Definition 16.6 *A SDE has unique weak solution, if for any pair of weak solutions* $\{X_1(t)\}_{t\in[0,T]}$ *and* $\{X_2(t)\}_{t\in[0,T]}$ *to it, we have*

$$\{X_1(t)\}_{t\in[0,T]} =_{\texttt{same fidi's}} \{X_2(t)\}_{t\in[0,T]}.$$

Exercise 92 Does strong uniqueness trivially imply weak uniqueness? Does weak uniqueness trivially imply strong uniqueness?

Exercise 93 Pick a constant $\alpha\in(0,1)$, and consider the SDE (ODE)

$$dX(t) = \mu(t,X(t))\,dt \qquad \text{for} \quad t\geq 0, \qquad X(0)=0,$$

where $\mu(t,x) = |x|^\alpha$. Show that, for any constant $t_0\in(0,\infty]$, it is solved by

$$X(t) = \begin{cases} 0 & \textit{for} \quad t\in[0,t_0) \\ ((1-\alpha)\,(t-t_0))^{1/(1-\alpha)} & \textit{for} \quad t\in[t_0,\infty) \end{cases} \in \bigcup_{T>0} P_T.$$

EXERCISE 94 Let sign$(x) \equiv 2\, I_{[0,\infty)}(x) - 1$, and pick a $T \in (0,\infty)$. Show that

$$dX(t) = \mathrm{sign}(X(t))\, dB(t) \qquad \text{for} \quad t \in [0,T], \qquad X(0) = 0,$$

does not have unique strong solution. (We will see later that it does not have a strong solution at all. But it has a weak solution, that is unique!) (**Hint:** Consider $-X$.)

Theorem 16.7 (GRÖNWALL'S LEMMA) *Let $u, v \in \mathbb{C}([0,T])$ satisfy*

$$v(t) \leq C + \int_0^t u(r)v(r)\, dr \equiv V(t) \qquad \textit{for} \quad t \in [0,T],$$

for some constant $C \geq 0$. We have

$$v(t) \leq V(t) \leq C\, \exp\Big\{\int_0^t u(r)\, dr\Big\} \qquad \textit{for} \quad t \in [0,T].$$

Proof. Picking an $\varepsilon > 0$, we have

$$v(t) \leq (C+\varepsilon) + \int_0^t u(r)v(r)\, dr \equiv V_\varepsilon(t) \qquad \text{for} \quad t \in [0,T].$$

Since V_ε is strictly positive, it follows that

$$(\ln(V_\varepsilon(t)))' = V_\varepsilon'(t)/V_\varepsilon(t) = u(t)v(t)/V_\varepsilon(t) \leq u(t) \qquad \text{for} \quad t \in [0,T].$$

This in turn gives

$$\ln(V_\varepsilon(t)) = \ln(V_\varepsilon(0)) + \int_0^t (\ln(V_\varepsilon(r)))'\, dr \leq \ln(C+\varepsilon) + \int_0^t u(r)\, dr \qquad \text{for} \quad t \in [0,T],$$

so that

$$v(t) \leq V_\varepsilon(t) \leq \exp\Big\{\ln(C+\varepsilon) + \int_0^t u(r)\, dr\Big\} = (C+\varepsilon)\, \exp\Big\{\int_0^t u(r)\, dr\Big\}$$

for $t \in [0,T]$. Sending $\varepsilon \downarrow 0$, we get the statement of the theorem. □

As we will now see exemplified, *Grönwall's lemma* is often used with $C = 0$.

Theorem 16.8 *Consider the SDE*

$$dX(t) = \mu(t, X(t))\, dt + \sigma(t, X(t))\, dB(t) \qquad \textit{for} \quad t \in [0,T], \qquad X(0) = X_0$$

(where X_0 is independent of B). Assume that, to each $n \in \mathbb{N}$, there exists a constant $K_n > 0$ such that the following Lipschitz condition holds

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K_n |x-y| \qquad \textit{for} \quad |x|, |y| \leq n \quad \textit{and} \quad t \in [0,T].$$

We have strong uniqueness for solutions to the SDE (but not necessarily existence).

Proof (after [22, pp. 287-288])*. Consider two strong solutions $\{X_1(t)\}_{t\in[0,T]}$ and $\{X_2(t)\}_{t\in[0,T]}$ to the SDE, on a common probability space. It is enough to prove that

$$\mathbf{P}\{X_1(t)=X_2(t)\} = 1 \qquad \text{for each} \quad t\in[0,T],$$

because this gives

$$\mathbf{P}\Big\{X_1(t)=X_2(t) \ \text{ for all } \ t\in\mathbb{Q}\cap[0,T]\Big\} = 1.$$

Since $\{X_1(t)\}_{t\in[0,T]}$ and $\{X_2(t)\}_{t\in[0,T]}$ are continuous, and thus determined by their values on $\mathbb{Q}\cap[0,T]$, it follows that they are equal on $[0,T]$, with probability one.

The solution X_i is adapted, because $\int_0^t \sigma(r,X_i(r))\,dB(r)$ is a martingale (and thus adapted), and because $\int_0^t \mu(r,X_i(r))\,dr$ is adapted, by the argument used in the proof of Theorem 11.1, since $\mu(\cdot,X_i(\cdot))$ is measurable and adapted. [Recall that we require that μ is measurable (Definition 15.15).] Since X_i is adapted,

$$\tau_{i,n} \equiv \inf\Big\{t\in[0,T] : |X_i(t)|\geq n\Big\}$$

is a stopping time, because (by continuity)

$$\{\tau_{i,n}\leq t\} = \Big\{\sup_{s\in[0,t]}|X_i(s)|\geq n\Big\} = \bigg(\bigcap_{n=0}^{\infty}\bigcap_{k=0}^{[nt]}\{|X_i(k/n)|<n\}\bigg)^c \in \mathcal{F}_t.$$

It follows that $\tau_n \equiv \tau_{1,n}\wedge\tau_{2,n}$ is a stopping time. Notice that $X_i^{(n)}(t) \equiv X_i(t\wedge\tau_n)$ satisfies $|X_i^{(n)}(t)|\leq n$ for $t\in[0,T]$ (by continuity), and that

$$X_i^{(n)}(t) = \int_0^t \mu(r,X_i(r))\,I_{[0,\tau_n]}(r)\,dr + \int_0^t \sigma(r,X_i(r))\,I_{[0,\tau_n]}(r)\,dB(r) \qquad \text{for} \quad t\in[0,T].$$

Since $\tau_n\uparrow\infty$ as $n\to\infty$, by continuity of X_1 and X_2, it is enough to prove that $X_1^{(n)}(t)=X_2^{(n)}(t)$ with probability one, for $t\in[0,T]$ and $n\in\mathbb{N}$. However, Theorem 10.2 and Cauchy-Schwarz inequality, together with the Lipschitz condition, give

$$\begin{aligned}
\mathbf{E}\Big\{(X_1^{(n)}(t)-X_2^{(n)}(t))^2\Big\} &\leq 2\,\mathbf{E}\bigg\{\bigg(\int_0^t\Big(\mu(r,X_1(r))-\mu(r,X_2(r))\Big)\,I_{[0,\tau_n]}(r)\,dr\bigg)^2\bigg\}\\
&\quad+2\,\mathbf{E}\bigg\{\bigg(\int_0^t\Big(\sigma(r,X_1(r))-\sigma(r,X_2(r))\Big)\,I_{[0,\tau_n]}(r)\,dB(r)\bigg)^2\bigg\}\\
&\leq 2\,\mathbf{E}\bigg\{\bigg(\int_0^t dr\bigg)\bigg(\int_0^t\Big(\mu(r,X_1(r))-\mu(r,X_2(r))\Big)^2\,I_{[0,\tau_n]}(r)\,dr\bigg)\bigg\}\\
&\quad+2\,\mathbf{E}\bigg\{\int_0^t\Big(\sigma(r,X_1(r))-\sigma(r,X_2(r))\Big)^2\,I_{[0,\tau_n]}(r)\,dr\bigg\}\\
&\leq 2\,(T+1)\,K_n^2\int_0^t\mathbf{E}\Big\{(X_1^{(n)}(t)-X_2^{(n)}(t))^2\Big\}\,dr \qquad \text{for} \quad t\in[0,T].
\end{aligned}$$

Hence *Grönwall's lemma*, with $C=0$, $u(s) = 2\,(T+1)\,K_n^2$ and $v(s) = \mathbf{E}\Big\{(X_1^{(n)}(s)-X_2^{(n)}(s))^2\Big\}$ (that is continuous, by continuity of X_i and *Dominated Convergence*), gives $v(t)=0$. This means that $X_1^{(n)}(t)=X_2^{(n)}(t)$ with probability one. $\square$

17 Seventeenth Lecture

17.1 Uniqueness of Solutions to SDE (continued)

Example 17.1 Pick a $T\in(0,\infty)$. The following SDE has unique strong solution

$$dX(t) = X(t)^2\,dt \qquad \text{for} \quad t\in[0,T], \qquad X(0)=1,$$

by Theorem 16.8, since the drift $\mu(x)=x^2$ satisfies a Lipschitz condition

$$|x^2-y^2|^2 = |x+y|^2|x-y|^2 \le 4\,n^2\,|x-y|^2 \qquad \text{for} \quad |x|,|y|\le n.$$

In more general settings, than ours with one-dimesional SDE of diffusion type, uniqueness result like Theorem 16.8 are the standard. However, for one-dimesional diffusions, there are other uniqueness results, as for example the following one:

Theorem 17.2 (Yamada-Watanabe) (e.g., [22, pp. 291-292]*) *Consider the SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \text{for} \quad t\in[0,T], \qquad X(0)=X_0$$

(where X_0 is independent of B). Assume that there exist a constant $K>0$, together with a strictly increasing function $h:[0,\infty)\to[0,\infty)$ that satisfies

$$h(0)=0 \qquad \text{and} \qquad \int_0^\varepsilon h(r)^2\,dr = \infty \qquad \text{for each} \quad \varepsilon>0,$$

such that the following conditions hold

$$\begin{cases} |\mu(t,x)-\mu(t,y)| \le K\,|x-y| \\ |\sigma(t,x)-\sigma(t,y)| \le h(|x-y|) \end{cases} \qquad \text{for} \quad x,y\in\mathbb{R} \quad \text{and} \quad t\in[0,T].$$

We have strong uniqueness for solutions to the SDE (but not necessarily existence).

Theorem 17.2 is typically used in the following way:

Corollary 17.3 *Consider the SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \text{for} \quad t\in[0,T], \qquad X(0)=X_0$$

(where X_0 is independent of B). Assume that there exist constants $K>0$ and $\alpha\in[1/2,\infty)$ such that the following Hölder conditions hold

$$\begin{cases} |\mu(t,x)-\mu(t,y)| \le K\,|x-y| \\ |\sigma(t,x)-\sigma(t,y)| \le |x-y|^\alpha \end{cases} \qquad \text{for} \quad x,y\in\mathbb{R} \quad \text{and} \quad t\in[0,T].$$

We have strong uniqueness for solutions to the SDE (but not necessarily existence).

Famous Example 17.4 Given an $\alpha\in[1/2,1]$, the SDE

$$dX(t) = |X(t)|^\alpha\, dB(t) \qquad \text{for} \quad t\in[0,T], \qquad X(0)=0,$$

has unique solution $X(t)=0$, by Corollary 17.3, since $||x|^\alpha-|y|^\alpha| \leq |x-y|^\alpha$. #

17.2 Existence of Strong Solutions to SDE

Famous Example 17.5 (e.g., [22, pp. 301-302])* We see in Example 19.10 that

$$dX(t) = \operatorname{sign}(X(t))\, dB(t) \qquad \text{for} \quad t\in[0,T], \qquad X(0)=0, \tag{17.1}$$

has a weak solution. However, it has no strong solution. Formally, assuming that X is a strong solution, we get this using *Itô's formula* on $|X|$, which gives

$$d|X|(t) = \operatorname{sign}(X(t))\, dX(t) + \ldots = (\operatorname{sign}(X(t)))^2\, dB(t) + \ldots = dB(t) + \ldots\ .$$

Albeit $|\cdot| \notin \mathbb{C}^2(\mathbb{R})$, $|\cdot|$ is convex, and there is an *Itô's formula* for convex functions [that need not be $\mathbb{C}^2(\mathbb{R})$ (see Remark 26.5 below)]. This formula gives $d|X|(t) = dB(t)$, plus a function of $\{|X(s)|\}_{s\leq t}$ coming from a second order "generalized derivative" of $|\cdot|$ at zero. Rearranging, we find that $B(t)$ is a function of $\{|X(s)|\}_{s\leq t}$. Since $X(t)$ in turn is a function of $\{B(s)\}_{s\leq t}$ [solving (17.1)], $X(t)$ is adapted to $\sigma(|X(s)|: s\leq t)$. This contradicts that X is BM, by Exercise 95 below. #

***Exercise 95** Show that a solution X to (17.1) is BM. (**Hint:** Show that $\mathbf{E}\{e^{\sum_{j=1}^N \theta_j X(\tau_j)}\} = \lim \mathbf{E}\{e^{\sum_{j=1}^N \theta_j \sum_{t_i\leq\tau_j} X(t_{i-1})(B(t_i)-B(t_{i-1}))}\} = \mathbf{E}\{e^{\sum_{j=1}^N \theta_j B^0(\tau_j)}\}$.)

Remark 17.6 By Theorem 15.6, X in (17.1) has $[X](t) = \int_0^t \operatorname{sign}(r)^2\, dr = t$, and so X is BM, by Theorem 18.1 below: This is an economic way to do Exercise 95! #

Exercise 96 Pick a $T\geq 1$ Show that the following SDE (ODE) does not have a strong solution (cf. Example 17.1);

$$dX(t) = X(t)^2\, dt \qquad \text{for} \quad t\in[0,T], \qquad X(0)=1,$$

Theorem 17.7 *Consider the SDE*

$$dX(t) = \mu(t,X(t))\, dt + \sigma(t,X(t))\, dB(t) \qquad \textit{for} \quad t\in[0,T], \qquad X(0)=X_0$$

(where X_0 is independent of B). Assume that

$$\begin{cases} |\mu(t,x)-\mu(t,y)| + |\sigma(t,x)-\sigma(t,y)| \leq K\,|x-y| \\ \mu(t,x)^2+\sigma(t,x)^2 \leq K^2\,(1+x^2) \end{cases} \quad \textit{for} \quad x,y\in\mathbb{R} \quad \textit{and} \quad t\in[0,T],$$

for some constant $K>0$. There exists a unique strong solution to the SDE.

Rather than Theorem 17.7, we prove the following result, using so called *Picard-Lindelöf iteration*. It gives Theorem 17.7 by truncation of X_0 ([22, Problem 5.2.12])*:

Theorem 17.8 *Consider the SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \textit{for} \quad t\in[0,T], \qquad X(0)=X_0,$$

where (X_0 is independent of B) with $\mathbf{E}\{X_0^2\}<\infty$. *Assume that*

$$\begin{cases} |\mu(t,x)-\mu(t,y)| + |\sigma(t,x)-\sigma(t,y)| \leq \ K\,|x-y| \\ \qquad \mu(t,x)^2+\sigma(t,x)^2 \qquad\qquad \leq K^2\,(1+x^2) \end{cases} \qquad \textit{for} \quad x,y\in\mathbb{R} \quad \textit{and} \quad t\in[0,T],$$

for some constant $K>0$. There exists a unique strong solution to the SDE. Moreover, the solution is square-integrable.

**Proof* (after [22, pp. 289-291]). We have uniqueness from Theorem 16.8, so that it is enough to prove existence and square-integrability. Define $X^{(0)}(t)=X_0$ and

$$X^{(k+1)}(t) = X_0 + \int_0^t \mu(r,X^{(k)}(r))\,dr + \int_0^t \sigma(r,X^{(k)}(r))\,dB(r) \qquad \text{for} \quad k\in\mathbb{N},$$

for $t\in[0,T]$. To establish that the process $X^{(k+1)}(t)$ on the left-hand side is well-defined, it is enough to show that the process on the right-hand side $X^{(k)}(t)$ satifies

$$\sup\nolimits_{t\in[0,T]} \mathbf{E}\{X^{(t)}(r)^2\} < \infty, \tag{17.2}$$

because (by *Fubini's theorem*) this shows that

$$\mathbf{E}\Big\{\int_0^T \sigma(r,X^{(k)}(r))^2\,dr\Big\} \leq \mathbf{E}\Big\{\int_0^T K^2\big(1+X^{(k)}(r)^2\big)\,dr\Big\} = \int_0^T K^2\big(1+\mathbf{E}\{X^{(k)}(r)^2\}\big)\,dr$$

is finite, so that $\sigma(\cdot,X^{(k)}(\cdot))\in E_T$, and similarly, by Cauchy-Schwarz inequality,

$$\begin{aligned}\Big(\mathbf{E}\Big\{\int_0^T |\mu(r,X^{(k)}(r))|\,dr\Big\}\Big)^2 &\leq \mathbf{E}\Big\{\Big(\int_0^T |\mu(r,X^{(k)}(r))|\,dr\Big)^2\Big\} \\ &\leq \mathbf{E}\Big\{\Big(\int_0^T dr\Big)\Big(\int_0^T \mu(r,X^{(k)}(r))^2\,dr\Big)\Big\} \\ &\leq T\,\mathbf{E}\Big\{\int_0^T K^2\big(1+X^{(k)}(r)^2\big)\,dr\Big\} < \infty,\end{aligned}$$

so that $\int_0^T |\mu(r,X^{(k)}(r)|\,dr < \infty$ a.s. Notice that (17.2) holds trivially for $k=0$ [since $\mathbf{E}\{X_0^2\}<\infty$]. If we have (17.2) for a certain k, we further get (cf. above)

$$\begin{aligned}&\mathbf{E}\{X^{(k+1)}(r)^2\} \\ &\leq 3\,\mathbf{E}\{X_0^2\} + 3\,\mathbf{E}\Big\{\Big(\int_0^t |\mu(r,X^{(k)}(r))|\,dr\Big)^2\Big\} + 3\,\mathbf{E}\Big\{\Big(\int_0^t \sigma(r,X^{(k)}(r))\,dB(r)\Big)^2\Big\} \\ &\leq 3\,\mathbf{E}\{X_0^2\} + 3T\,\mathbf{E}\Big\{\int_0^t K^2\big(1+X^{(k)}(r)^2\big)\,dr\Big\} + 3\,\mathbf{E}\Big\{\int_0^t K^2\big(1+X^{(k)}(r)^2\big)\,dr\Big\} \\ &\leq C_1 + C_1\int_0^t \mathbf{E}\{X^{(k)}(r)^2\}\,dr \qquad \text{for} \quad t\in[0,T], \quad \text{for some constant} \quad C_1>0,\end{aligned}$$

by Theorem 10.2. This gives (17.2) for $k+1$, and thus for all $k \in \mathbb{N}$, by induction.

Now consider the processes [well-defined, because of (17.2)]

$$\begin{cases} B^{(k)}(t) \equiv \int_0^t \Big(\mu(r, X^{(k+1)}(r)) - \mu(r, X^{(k)}(r))\Big)\, dr \\ M^{(k)}(t) \equiv \int_0^t \Big(\sigma(r, X^{(k+1)}(r)) - \sigma(r, X^{(k)}(r))\Big)\, dB(r) \end{cases} \quad \text{for} \quad k \in \mathbb{N} \quad \text{and} \quad t \in [0,T].$$

Since $\{M^{(k)}(t)\}_{t\in[0,T]}$ is a square-integrable martingale [by the fact that $\sigma(\cdot, X^{(k)}(\cdot))$, $\sigma(\cdot, X^{(k+1)}(\cdot)) \in E_T$, together with Theorem 10.2], *Doob Maximal inequality* yields

$$\begin{aligned} \mathbf{E}\Big\{\sup_{s\in[0,t]} M^{(k)}(s)^2\Big\} &\leq \sqrt{2}\,\mathbf{E}\{M^{(k)}(t)^2\} = \sqrt{2}\,\mathbf{E}\Big\{\int_0^t \Big(\sigma(r, X^{(k+1)}(r)) - \sigma(r, X^{(k)}(r))\Big)^2 dr\Big\} \\ &\leq \sqrt{2}\int_0^t K^2\, \mathbf{E}\big\{|X^{(k+1)}(r) - X^{(k)}(r)|^2\big\}\, dr \end{aligned}$$

for $t \in [0,T]$ (using Theorem 10.2 again). Further, Cauchy-Schwarz inequality gives

$$\begin{aligned} \mathbf{E}\Big\{\sup_{s\in[0,t]} B^{(k)}(s)^2\Big\} &\leq \mathbf{E}\Big\{\Big(\int_0^t dr\Big)\Big(\int_0^t \Big(\mu(r, X^{(k+1)}(r)) - \mu(r, X^{(k)}(r))\Big)^2 dr\Big)\Big\} \\ &\leq T\int_0^t K^2\, \mathbf{E}\big\{|X^{(k+1)}(r) - X^{(k)}(r)|^2\big\}\, dr \end{aligned}$$

for $t \in [0,T]$. Putting these finding together, we conclude that

$$\begin{aligned} \mathbf{E}\Big\{\sup_{s\in[0,t]} |X^{(k+1)}(s) - X^{(k)}(s)|^2\Big\} &= \mathbf{E}\Big\{\sup_{s\in[0,t]} |B^{(k-1)}(s) + M^{(k-1)}(s)|^2\Big\} \\ &\leq 2\,\mathbf{E}\Big\{\sup_{s\in[0,t]} |B^{(k-1)}(s)|^2\Big\} + 2\,\mathbf{E}\Big\{\sup_{s\in[0,t]} |M^{(k-1)}(s)|^2\Big\} \\ &\leq C_2\int_0^t \mathbf{E}\big\{|X^{(k)}(r) - X^{(k-1)}(r)|^2\big\}\, dr \\ &\Big(\leq C_2\int_0^t \mathbf{E}\Big\{\sup_{s\in[0,r]} |X^{(k)}(s) - X^{(k-1)}(s)|^2\Big\}\, dr\Big) \end{aligned} \qquad (17.3)$$

for some constant $C_2 > 0$. By means of induction (see Exercise 97 below), this gives

$$\mathbf{E}\Big\{\sup_{s\in[0,T]} |X^{(k+1)}(s) - X^{(k)}(s)|^2\Big\} \leq \sup_{s\in[0,T]} \mathbf{E}\big\{|X^{(1)}(s) - X^{(0)}(s)|^2\big\}\, \frac{C_2^k}{k!} = C_3\,\frac{C_2^k}{k!} \qquad (17.4)$$

[recall (17.2)]. By application of *Tjebysjev's inequality*, this implies that

$$\mathbf{P}\Big\{\sup_{t\in[0,T]} |X^{(k+1)}(s) - X^{(k)}(s)| > \frac{1}{2^{k+1}}\Big\} \leq C_3\,\frac{C_2^k}{k!}\Big/\frac{1}{(2^{k+1})^2} = \frac{4\,C_3\,(2\,C_2)^k}{k!}.$$

Hence we have

$$\sum_{k=0}^{\infty} \mathbf{P}\Big\{\sup_{t\in[0,T]} |X^{(k+1)}(s) - X^{(k)}(s)| > \frac{1}{2^{k+1}}\Big\} < \infty,$$

so that, by the *Borel-Cantelli lemma*,

$$\mathbf{P}\Big\{\sup_{t\in[0,T]} |X^{(k+1)}(s) - X^{(k)}(s)| > \frac{1}{2^{k+1}} \text{ for infinitely many } k \in \mathbb{N}\Big\} = 0.$$

In other words, we have

$$\sup_{t\in[0,T]} |X^{(k+1)}(s)-X^{(k)}(s)| \leq \frac{1}{2^{k+1}} \qquad \text{for} \quad k\geq k_0(\omega), \quad \text{for some} \quad k_0(\omega)\in\mathbb{N},$$

that is finite with probability one. By summation, this gives

$$\sup_{t\in[0,T]} |X^{(m)}(s)-X^{(n)}(s)| \leq \sum_{k=m\wedge n}^{m\vee n-1} \sup_{t\in[0,T]} |X^{(k+1)}(s)-X^{(k)}(s)| \leq \sum_{k=m\wedge n}^{\infty} \frac{1}{2^{k+1}} \leq \frac{1}{2^{m\wedge n}}$$

for $m\wedge n\geq k_0(\omega)$, where $\underline{m\vee n} = \max\{m,n\}$. Thus the processes $\{X^{(k)}\}_{k=1}^{\infty}$ constitute a Cauchy sequence wrt. uniform convergence of functions (processes) on the interval $[0,T]$. Since each of these processes is continuous [by Theorem 10.2, together with absolute continuity of $\int^t \mu(r,X^{(k-1)}(r))\,dr$], it follows (e.g., [12, pp. 1-2])* that there exists a stochastic process $\{X(t)\}_{t\in[0,T]}$, that is continuous with probability one, such that $\sup_{t\in[0,T]} |X^{(k)}(t)-X(t)| \to 0$ as $k\to\infty$, with probability one.

By the established convergence together with (17.4) and *Fatou's lemma*, we have

$$\begin{aligned} \mathbf{E}\Big\{\sup_{t\in[0,T]} |X(t)-X^{(k)}(t)|^2\Big\} &= \mathbf{E}\Big\{\liminf_{\ell\to\infty} \sup_{t\in[0,T]} |X^{(\ell)}(t)-X^{(k)}(t)|^2\Big\} \\ &\leq \liminf_{\ell\to\infty} \mathbf{E}\Big\{\sup_{t\in[0,T]} |X^{(\ell)}(t)-X^{(k)}(t)|^2\Big\} \\ &\leq \limsup_{\ell\to\infty} \sum_{n=k}^{\ell-1} 2^{n+1-k}\, \mathbf{E}\Big\{\sup_{t\in[0,T]} |X^{(n+1)}(t)-X^{(n)}(t)|^2\Big\} \\ &\leq 2^{1-k} \sum_{n=0}^{\infty} 2^n\, C_3 \frac{C_2^n}{n!} = C_3\, e^{2C_2}\, 2^{1-k} \end{aligned} \tag{17.5}$$

(see Exercise 98 below). In particular, we have [recall (17.2)]

$$\begin{aligned} \sup_{t\in[0,T]} \mathbf{E}\{X(t)^2\} &\leq \mathbf{E}\Big\{\sup_{t\in[0,T]} X(t)^2\Big\} \\ &\leq 2\,\mathbf{E}\Big\{\sup_{t\in[0,T]} |X(t)-X^{(1)}(t)|^2\Big\} + 2\,\mathbf{E}\Big\{\sup_{t\in[0,T]} |X^{(1)}(t)|^2\Big\} < \infty. \end{aligned}$$

From this we get that

$$Y(t) = X_0 + \int_0^t \mu(r,X(r))\,dr + \int_0^t \sigma(r,X(r))\,dB(r) \qquad \text{is well-defined for} \quad t\in[0,T],$$

in the same way as we got that $X^{(k+1)}$ is well-defined from (17.2). Hence it is enough to show that $X=Y$ a.s., to establish the existence of a square-integrable solution.

We have, by (17.5) together with the argument used to establish (17.3),

$$\mathbf{E}\Big\{\sup_{s\in[0,T]} |Y(t)-X^{(k+1)}(t)|^2\Big\} \leq C_2 \int_0^T \mathbf{E}\Big\{\sup_{s\in[0,r]} |X(s)-X^{(k)}(s)|^2\Big\}\,dr \leq C_2 T\, C_3\, e^{2C_2}\, 2^{1-k}.$$

Hence another application of (17.5) shows that

$$\begin{aligned} \mathbf{E}\Big\{|Y(t)-X(t)|^2\Big\} &\leq 2\,\mathbf{E}\Big\{\sup_{s\in[0,T]} |Y(t)-X^{(k+1)}(t)|^2\Big\} + 2\,\mathbf{E}\Big\{\sup_{s\in[0,T]} |X^{(k+1)}(t)-X(t)|^2\Big\} \\ &\leq C_2 T\, C_3 e^{2C_2}\, 2^{2-k} + C_3\, e^{2C_2}\, 2^{2-k} \to 0 \qquad \text{as} \quad k\to\infty. \end{aligned}$$

This gives $Y(t)=X(t)$ a.s. for $t\in[0,T]$, so that [by continuity of X and Y (used as in the proof of Theorem 16.8)], $\mathbf{P}\{Y(t)=X(t)$ for $t\in[0,T]\}=1$. □

EXERCISE 97 Show that (17.3) gives (17.4).

EXERCISE 98 Prove the inequality featuring in (17.5)

$$\mathbf{E}\Big\{\sup_{t\in[0,T]}|X^{(\ell)}(t)-X^{(k)}(t)|^2\Big\}\leq\sum_{n=k}^{\ell-1}2^{n+1-k}\,\mathbf{E}\Big\{\sup_{t\in[0,T]}|X^{(n+1)}(t)-X^{(n)}(t)|^2\Big\}\qquad\text{for}\quad \ell>k.$$

18 Eighteenth Lecture

18.1 Paul Lévy's Characterization of BM

The following celebrated result, by P. Lévy, is crucial in the use of the machinery of stochastic calculus. We have seen one indication of this in Exercise 95 (together with Remark 17.6). Other vital applications include *Girsanov's theorem* (Theorems 19.1 and 19.6 below), which is used to solve SDE below.

Theorem 18.1 (Lévy's Characterization of BM) *A continuous local martingale* $\{X(t),\mathcal{F}_t\}_{t\geq 0}$ *is BM wrt. the filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ (see Section 16.2), *iff. it is BM in the sense of Definition 2.1, iff.* $[X](t)=[X,X](t)=t$ *for* $t\geq 0$.

Proof (after [22, p. 157])*. The first implication to the right is trivial, and the second Theorem 2.13. It remains to prove that X is BM wrt. $\{\mathcal{F}_t\}_{t\geq 0}$ when $[X](t)=t$, i.e.,

$$\mathbf{P}\Big\{A\cap\{X(t)-X(s)\in\cdot\}\Big\}=\mathbf{P}\{A\}\,\mathbf{P}\{B(t)-B(s)\in\cdot\}\qquad\text{for}\quad A\in\mathcal{F}_s\quad\text{and}\quad 0\leq s<t.$$

[This gives independence of increments, as well as the desired distribution for them (taking $A=\Omega$).] By theory for characteristic functions, this follows if we can prove

$$\mathbf{E}\Big\{\mathrm{e}^{i\varphi I_A}\mathrm{e}^{i\theta(X(t)-X(s))}\Big\}=\mathbf{E}\{\mathrm{e}^{i\varphi I_A}\}\,\mathbf{E}\{\mathrm{e}^{i\theta(B(t)-B(s))}\}\qquad\text{for}\quad\theta,\varphi\in\mathbb{R}.$$

By conditioning, we have

$$\mathbf{E}\Big\{\mathrm{e}^{i\varphi I_A}\mathrm{e}^{i\theta(X(t)-X(s))}\Big\}=\mathbf{E}\Big\{\mathrm{e}^{i\varphi I_A}\,\mathbf{E}\{\mathrm{e}^{i\theta(X(t)-X(s))}|\mathcal{F}_s\}\Big\}\qquad\text{for}\quad\theta,\varphi\in\mathbb{R}.$$

It follows that it is enough to prove that

$$\mathbf{E}\{\mathrm{e}^{i\theta(X(t)-X(s))}|\mathcal{F}_s\}=\mathbf{E}\{\mathrm{e}^{i\theta(B(t)-B(s))}\}\qquad\text{for}\quad\theta\in\mathbb{R}.\tag{18.1}$$

By *Ito's formula* for continuous local martingales (Section 16.1), aplied to the function $f(x)=\mathrm{e}^{i\theta x}$ (considering the real and imaginary parts separately), we have

$$\mathrm{e}^{i\theta X(t)}=\mathrm{e}^{i\theta X(s)}+i\theta\int_s^t\mathrm{e}^{i\theta X(r)}\,dX(r)-(\theta^2/2)\int_s^t\mathrm{e}^{i\theta X(r)}\,dr.\tag{18.2}$$

Recall that $\int_0^t\mathrm{e}^{i\theta X(r)}\,dX(r)$ is a local martingale (more precisely, its real and imaginary parts are), so that there exist stopping times $0\leq\tau_1\ldots\leq\tau_n\uparrow\infty$ a.s. as $n\to\infty$, such that $\int_0^{t\wedge\tau_n}\mathrm{e}^{i\theta X(r)}\,dX(r)$ is a martingale. Hence we have (cf. Exercise 85)

$$\mathbf{E}\bigg\{\int_{s\wedge\tau_n}^{t\wedge\tau_n}\mathrm{e}^{i\theta(X(r)-X(s\wedge\tau_n))}\,dX(r)\,\bigg|\,\mathcal{F}_s\bigg\}=\mathrm{e}^{-i\theta X(s\wedge\tau_n)}\,\mathbf{E}\bigg\{\int_{s\wedge\tau_n}^{t\wedge\tau_n}\mathrm{e}^{i\theta X(r)}\,dX(r)\,\bigg|\,\mathcal{F}_s\bigg\}=0.$$

Since the properties of $\{\tau_n\}_{n=1}^{\infty}$ ensure that

$$\left\{\begin{array}{lcl}\mathrm{e}^{i\theta(X(t\wedge\tau_n)-X(s\wedge\tau_n))} & \to & \mathrm{e}^{-i\theta(X(t)-X(s))}\\ \displaystyle\int_{s\wedge\tau_n}^{t\wedge\tau_n}\mathrm{e}^{i\theta(X(r)-X(s\wedge\tau_n))}\,dr & \to & \displaystyle\int_s^t\mathrm{e}^{i\theta(X(r)-X(s))}\,dr\end{array}\right.\qquad\text{a.s.}\qquad\text{as}\quad n\to\infty,$$

where the random variables involved are bounded, *Dominated Convergence* give convergence in $\mathbb{L}^1$. By Exercise 68, this yields convergence for the corresponding conditional expectations in $\mathbb{L}^1$. Using (18.2), we thus get

$$\mathbf{E}\Big\{\mathrm{e}^{i\theta(X(t)-X(s))}-1+(\theta^2/2)\int_s^t \mathrm{e}^{i\theta(X(r)-X(s))}\,dr\,\Big|\,\mathcal{F}_s\Big\}$$

$$\leftarrow \mathbf{E}\Big\{\mathrm{e}^{i\theta(X(t\wedge\tau_n)-X(s\wedge\tau_n))}-1+(\theta^2/2)\int_{s\wedge\tau_n}^{t\wedge\tau_n} \mathrm{e}^{i\theta(X(r)-X(s\wedge\tau_n))}\,dr\,\Big|\,\mathcal{F}_s\Big\}$$

$$= \mathbf{E}\Big\{i\theta\int_{s\wedge\tau_n}^{t\wedge\tau_n} \mathrm{e}^{i\theta(X(r)-X(s\wedge\tau_n))}\,dX(r)\,\Big|\,\mathcal{F}_s\Big\}=0$$

(with convergence in $\mathbb{L}^1$), so that

$$\mathbf{E}\Big\{\mathrm{e}^{i\theta(X(t)-X(s))}\,\Big|\,\mathcal{F}_s\Big\}=1-(\theta^2/2)\,\mathbf{E}\Big\{\int_s^t \mathrm{e}^{i\theta(X(r)-X(s))}\,dr\,\Big|\,\mathcal{F}_s\Big\}.$$

Picking an event $\Lambda\in\mathcal{F}_s$, this gives

$$\begin{aligned}\mathbf{E}\Big\{I_\Lambda\,\mathrm{e}^{i\theta(X(t)-X(s))}\Big\}=\int_\Lambda \mathrm{e}^{i\theta(X(t)-X(s))}\,d\mathbf{P}&=\int_\Lambda \mathbf{E}\Big\{\mathrm{e}^{i\theta(X(t)-X(s))}\,\Big|\,\mathcal{F}_s\Big\}\,d\mathbf{P}\\ &=\int_\Lambda\Big(1-(\theta^2/2)\,\mathbf{E}\Big\{\int_s^t \mathrm{e}^{i\theta(X(r)-X(s))}\,dr\,\Big|\,\mathcal{F}_s\Big\}\Big)\,d\mathbf{P}\\ &=\mathbf{P}\{\Lambda\}-(\theta^2/2)\int_\Lambda\Big(\int_s^t \mathrm{e}^{i\theta(X(r)-X(s))}\,dr\Big)\,d\mathbf{P}\\ &=\mathbf{P}\{\Lambda\}-(\theta^2/2)\int_s^t \mathbf{E}\Big\{I_\Lambda\,\mathrm{e}^{i\theta(X(r)-X(s))}\Big\}\,dr.\end{aligned}$$

By Exercise 99 below, this equation has unique solution

$$\mathbf{E}\Big\{I_\Lambda\,\mathrm{e}^{i\theta(X(t)-X(s))}\Big\}=\mathbf{P}\{\Lambda\}\,\mathrm{e}^{-\theta^2(t-s)/2}=\mathbf{P}\{\Lambda\}\,\mathbf{E}\{\mathrm{e}^{i\varphi(B(t)-B(s))}\}.$$

Using this together with the above system of equations, we obtain

$$\int_\Lambda \mathbf{E}\Big\{\mathrm{e}^{i\theta(X(t)-X(s))}\,\Big|\,\mathcal{F}_s\Big\}\,d\mathbf{P}=\mathbf{P}\{\Lambda\}\,\mathbf{E}\{\mathrm{e}^{i\varphi(B(t)-B(s))}\}=\int_\Lambda \mathbf{E}\{\mathrm{e}^{i\varphi(B(t)-B(s))}\}\,d\mathbf{P}$$

for $\Lambda\in\mathcal{F}_s$, which is the sought after identity (18.1). □

EXERCISE 99 Show that the equation

$$\mathbf{E}\Big\{I_\Lambda\,\mathrm{e}^{i\theta(X(t)-X(s))}\Big\}=\mathbf{P}\{\Lambda\}-(\theta^2/2)\int_s^t \mathbf{E}\Big\{I_\Lambda\,\mathrm{e}^{i\theta(X(r)-X(s))}\Big\}\,dr \qquad \text{for}\quad 0\le s<t$$

(in the proof of Theorem 18.1), has unique solution

$$\mathbf{E}\Big\{I_\Lambda\,e^{i\theta(X(t)-X(s))}\Big\}=\mathbf{P}\{\Lambda\}\,e^{-\theta^2(t-s)/2} \qquad \text{for}\quad 0\le s<t.$$

18.2 Stratonovich Stochastic Integrals

Definition 18.2 *Let $\{X(t)\}_{t\in[0,T]}$ and $\{Y(t)\}_{t\in[0,T]}$ be continuous local martingales. The Stratonovich stochastic integral of Y wrt. X is given by*

$$\int_0^t Y(r)\circ dX(r)\equiv\int_0^t Y(r)\,dX(r)+\frac{1}{2}\,[X,Y](t) \qquad \textit{for}\quad t\in[0,T].$$

This is the rigorous definition of the *Stratonovich (Fisk-Stratonovich) integral* discussed in Lecture 6. Notice that, by continuity of Y, the condition for existence of $\int_0^t Y\,dX$ in Section 16.1 holds, that $\int_0^t Y^2\,d[X,X]<\infty$ with probability one.

The Stratonovich integral is defined for a much narrower class of integrands (continuous local martingales), than the Itô integral [defined for adapted and measurable processes with $\int_0^t Y^2\,d[X,X]<\infty$]. This is because $[X,Y]$ features in the definition.

By Theorem 15.3 together with Section 16.1, the Stratonovich integral satisfies

$$\int_0^t Y\circ dX = \text{P-lim}\left\{\sum_{i=1}^n \frac{Y(t_{i-1})+Y(t_{i-1})}{2}\,(X(t_i)-X(t_{i-1})) : \begin{matrix}0=t_0<t_1<\ldots<t_n=t\\ \max_{1\le i\le n} t_i-t_{i-1}\to 0\end{matrix}\right\}.$$

EXERCISE 100 (INTEGRATION BY PARTS) Use *Itô's formula* to show that

$$\boxed{\int_0^t Y(r)\circ dX(r) = Y(t)X(t)-Y(0)X(0)-\int_0^t X(r)\circ dY(r)}\,.$$

18.3 Introduction to Girsanov's Theorem

One very important result in stochastic calculus is *Girsanov's theorem*, with *Girsanov transformation*. It is a main tool to find weak solutions to SDE, and significantly generalizes the earlier *Cameron-Martin formula* (see Exercise 104 below).

The *stochastic exponential* $\mathcal{E}X=\mathrm{e}^{X-X(0)-[X]/2}$ of a continuous local martingale X satisfies $d(\mathcal{E}X)=(\mathcal{E}X)\,dX$ (by Example 16.2), and is thus a continuous local martingale (since an Itô integral). By Exercise 101 below, it is a supermartingale.

EXERCISE 101 Show that a positive local martingale $\{Y(t)\}_{t\in[0,T]}$, is a supermartingale, and that it is a martingale iff. $\mathbf{E}\{Y(0)\}=\mathbf{E}\{Y(T)\}$.

Theorem 18.3 (GIRSANOV'S THEOREM) (First version.) *For $X\in P_T$ such that*

$$Z(t)\equiv= (\mathcal{E}\textstyle\int_0^\cdot X\,dB)(t) = \exp\left\{\displaystyle\int_0^t X(r)\,dB(r)-\frac{1}{2}\int_0^t X(r)^2\,dr\right\},\quad t\in[0,T],$$

is a martingale, there exists a probability measure (on the measurable space under con- sideration)*, under which the following stochastic process is BM*

$$W(t)\equiv B(t)-B(0)-\int_0^t X(r)\,dr,\quad t\in[0,T].$$

EXERCISE 102 Explain why $X\in P_T$ ensures that $\int_0^t X(r)\,dr$ is well-defined.

Example 18.4 Let $X(t)=\mu(t,B(t))$ in *Girsanov's theorem*. Rearranging, we get

$$B(t)=B(0)+\int_0^t X(r)\,dr+W(t)=B(0)+\int_0^t \mu(r,B(r))\,dr+W(t)\qquad\text{for}\quad t\in[0,T].$$

This means that $Y=B$ is a weak solution to the general SDE with dispersion one

$$dY(t) = \mu(t, Y(t))\,dt + dW(t) \qquad \text{for} \quad t \in [0, T].$$

The probability measure has changed, so that B *is no longer BM!* By *Novikov's Criterion* below, *Girsanov's theorem* really applies here, if for example μ is bounded. #

> ⋆Theorem 18.5 (Kazamaki's Criterion) *For a continuous local martingale* $\{X(t)\}_{t\in[0,T]}$ *such that* $\{e^{(X(t)-X(0))/2}\}_{t\in[0,T]}$ *is a submartingale,* $\{(\mathcal{E}X)(t)\}_{t\in[0,T]}$ *is a martingale.*

⋆*Proof* (after [33, p. 307]). Since $\mathcal{E}X$ is a supermartingale, it is enough to show that $\mathbf{E}\{(\mathcal{E}X)(T)\} = \mathbf{E}\{(\mathcal{E}X)(0)\} = 1$, by Execise 101. Put $Z_a(t) = e^{a\,(X(t)-X(0))/(1+a)}$ for $a \in (0,1)$. Notice that

$$\mathcal{E}(aX) = e^{aX - aX(0) - a^2[X]/2} = (\mathcal{E}X)^{a^2} e^{(a-a^2)\,(X-X(0))} = (\mathcal{E}X)^{a^2} Z_a^{1-a^2}.$$

It is enough to prove that $\mathcal{E}(aX)$ is a martingale, because then *Hölder's Inequality* gives

$$1 = \mathbf{E}\{(\mathcal{E}(aX))(T)\} = \mathbf{E}\{(\mathcal{E}X)(T)^{a^2} Z_a(T)^{1-a^2}\} \le (\mathbf{E}\{(\mathcal{E}X)(T)\})^{a^2} (\mathbf{E}\{Z_a(T)\})^{1-a^2}.$$

Here $Z_a(T) \le e^{(X(T)-X(0))/2} + I_{\{X(T)-X(0)\le 0\}}$, where the right-hand side is integrable, since $Z(t) \equiv e^{(X(t)-X(0))/2}$ is a submartingale. Hence *Dominated Convergence* gives $\mathbf{E}\{Z_a(T)\} \to \mathbf{E}\{Z(t)\}$ as $a \uparrow 1$, so that $(\mathbf{E}\{Z_a(T)\})^{1-a^2} \to 1$. Inserting above, we get $\mathbf{E}\{(\mathcal{E}X)(T)\} \ge 1$. We already know that $\mathbf{E}\{(\mathcal{E}X)(T)\} \le 1$, since $\mathcal{E}X$ is a supermatingale (which have decreasing means), by Exercise 101, and so $\mathbf{E}\{(\mathcal{E}X)(T)\} = 1$.

By Theorem 23.14 below, the continuous local martingale $\{(\mathcal{E}(aX))(t)\}_{t\in[0,T]}$ is a martingale if $\{(\mathcal{E}(aX))(\tau) : \tau$ stopping time with $\tau \le T\}$ is uniformly integrable. This in turn follows if $\{Z_a(\tau) : \tau$ stopping time with $\tau \le T\}$ is uniformly integrable, since noticing that $(\mathcal{E}(aX))(t)^{1/(1+a)} \le Z_a(t)$, *Hölders Inequality* gives

$$\begin{aligned}\mathbf{E}\Big\{I_{\{(\mathcal{E}(aX))(\tau)>y\}}(\mathcal{E}(aX))(\tau)\Big\} &= \mathbf{E}\Big\{I_{\{(\mathcal{E}(aX))(\tau)^{1/(1+a)}>y^{1/(1+a)}\}}(\mathcal{E}X)(\tau)^{a^2} Z_a(\tau)^{1-a^2}\Big\} \\ &\le (\mathbf{E}\{(\mathcal{E}X)(\tau)\})^{a^2} \Big(\mathbf{E}\Big\{I_{\{Z_a(\tau)>y^{1/(1+a)}\}} Z_a(\tau)\Big\}\Big)^{1-a^2},\end{aligned}$$

where $\mathbf{E}\{(\mathcal{E}X)(\tau)\} = \mathbf{E}\Big\{\mathbf{E}\{(\mathcal{E}X)(\tau)|\mathcal{F}_0\}\Big\} \le \mathbf{E}\{(\mathcal{E}X)(0)\} = 1$, by *Optional Sampling* Corollary 23.7, since $\mathcal{E}X$ is a supermartingale.

Recalling that $Z_a(t) \le Z(t) + I_{\{X(t)-X(0)\le 0\}}$ (where the second term on the right-hand side is uniformly integrable, e.g., by Exercise 77), it is enough to show that $\{Z(\tau) : \tau$ stopping time with $\tau \le T\}$ is uniformly integrable. This is an application of *Optional Sampling* Corollary 23.7, to the submartingale Z: As $y \to \infty$, we have

$$\mathbf{E}\Big\{I_{\{Z(\tau)>y\}} Z(\tau)\Big\} \le \mathbf{E}\Big\{I_{\{Z(\tau)>y\}} \mathbf{E}\{Z(T)|\mathcal{F}_\tau\}\Big\} = \mathbf{E}\Big\{I_{\{Z(\tau)>y\}} Z(T)\Big\}$$

$$\leq \mathbf{E}\Big\{I_{\{\sup_{t\in[0,T]} Z(t)>y\}} Z(T)\Big\} \to 0$$

uniformly for $\tau \leq T$, by continuity of Z and *Absolute Continuity of the Integral*. □

> Corollary 18.6 *For a continuous martingale* $\{X(t)\}_{t\in[0,T]}$ *such that* $\mathbf{E}\{\mathrm{e}^{(X(T)-X(0))/2}\}<\infty$, $\{(\mathcal{E}X)(t)\}_{t\in[0,T]}$ *is a martingale.*

Proof. This follows from Theorem 18.5: Since $X(t)-X(0)$ is a martingale and $\mathrm{e}^{\cdot}$ convex, $Z(t)\equiv \mathrm{e}^{(X(t)-X(0))/2}$, $t\in[0,T]$, is a submartingale by Exercise 55, if $\mathbf{E}\{Z(t)\}<\infty$ for $t\in[0,T]$. This we get from *Jensen's Inequality*, using that $\mathbf{E}\{Z(T)\}<\infty$,

$$\mathbf{E}\{Z(T)\} = \mathbf{E}\Big\{\mathbf{E}\{\mathrm{e}^{(X(T)-X(0))/2}|\mathcal{F}_t\}\Big\} \geq \mathbf{E}\Big\{\mathrm{e}^{\mathbf{E}\{(X(T)-X(0))/2|\mathcal{F}_t\}}\Big\} = \mathbf{E}\{\mathrm{e}^{(X(t)-X(0))/2}\}. \quad □$$

> Corollary 18.7 (Novikov's Criterion) *For a continuous local martingale* $\{X(t)\}_{t\in[0,T]}$ *such that* $\mathbf{E}\{\mathrm{e}^{[X](T)/2}\}<\infty$, $\{(\mathcal{E}X)(t)\}_{t\in[0,T]}$ *is a martingale.*

Proof. This follows from Corollary 18.6: By *Burkholder-Davis-Gundy inequality* Theorem 23.18 below, and the fact that $\sqrt{x}\leq \mathrm{e}^{(x-1)/2}$ for $x\geq 0$, we have

$$\mathbf{E}\Big\{\sup_{s\in[0,T]}|X(s)-X(0)|\Big\} \leq K_{1/2}\,\mathbf{E}\{[X](T)^{1/2}\} \leq K_{1/2}\,\mathbf{E}\{\mathrm{e}^{([X](T)-1)/2}\} < \infty.$$

Hence $|X(t)-X(0)| \leq Z_T \equiv \sup_{s\in[0,T]}|X(s)-X(0)|$ for $t\in[0,T]$, where Z_T is integrable, and so $X(t)-X(0)$ is a martingale, by Theorem 12.9. Since $\mathcal{E}X$ has finite means (being a supermartingale), *Cauchy-Schwarz inequality* further gives

$$\mathbf{E}\{\mathrm{e}^{(X(T)-X(0))/2}\} = \mathbf{E}\Big\{\sqrt{(\mathcal{E}X)(T)}\sqrt{\mathrm{e}^{[X](T)/2}}\Big\} \leq \sqrt{\mathbf{E}\{(\mathcal{E}X)(T)\}\,\mathbf{E}\{\mathrm{e}^{[X](T)/2}\}} < \infty. \quad □$$

*Remark 18.8 *Novikov's Criterion* is known to be close to optimal (e.g., [27, Sections 6.2.4-6.2.5]). Since it comes as an application of *Kazamaki's Criterion*, that criterion is even sharper (as is also exemplified in [33, Section VIII.1]). #

The following result (much versatiler than it might first appear), is not used by us. We cite it anyway, since many authors (e.g., [21] and [22])* prefer it instead of *Kazamaki's Criterion* to derive *Novikov's Criterion* (see Exercise 122 below).

> Theorem 18.9 (Wald's Identity) *For BM* B *and a stopping time* τ, *we have*
>
> $$\mathbf{E}\{(\mathcal{E}B)(\tau)\} = \mathbf{E}\{\mathrm{e}^{B(\tau)-B(0)-\tau/2}\} = 1 \qquad \textit{when} \qquad \mathbf{E}\{\mathrm{e}^{\tau/2}\}<\infty.$$

Exercise 103 Show how this *Wald's Identity* gives the one in Example 13.3.

19 Nineteenth Lecture

19.1 Girsanov's Theorem

Theorem 19.1 (Girsanov's Theorem) *Let* $X \in P_T$, *and assume that the stochastic process*

$$Z(t) \equiv \exp\Big\{\int_0^t X(r)\,dB(r) - \frac{1}{2}\int_0^t X(r)^2\,dr\Big\}, \quad t \in [0,T],$$

is a martingale (wrt. the filtration associated with B). *Under the probability measure*

$$\tilde{\mathbf{P}}\{A\} \equiv \mathbf{E}\Big\{I_A \exp\Big\{\int_0^T X(r)\,dB(r) - \frac{1}{2}\int_0^T X(r)^2\,dr\Big\}\Big\} = \mathbf{E}\{I_A Z(T)\}, \quad A \in \mathcal{F}_T,$$

the following stochastic process W *is a BM* (wrt. the filtration associated with B)

$$W(t) \equiv B(t) - B(0) - \int_0^t X(r)\,dr, \quad t \in [0,T].$$

Lemma 19.2 (Bayes' Rule) *Let* $X \in P_T$, *and assume that*

$$Z(t) \equiv \exp\Big\{\int_0^t X(r)\,dB(r) - \frac{1}{2}\int_0^t X(r)^2\,dr\Big\}, \quad t \in [0,T],$$

is a martingale. For expectations $\tilde{\mathbf{E}}\{\cdot\}$ *wrt. the probability measure*

$$\tilde{\mathbf{P}}\{A\} = \mathbf{E}\{I_A Z(T)\}, \quad A \in \mathcal{F}_T,$$

and a $\mathcal{F}_t$*-measurable random variable* Y *with* $\tilde{\mathbf{E}}\{|Y|\} < \infty$ *and* $t \in [0,T]$, *we have*

$$\tilde{\mathbf{E}}\{Y|\mathcal{F}_s\} = \mathbf{E}\{YZ(t)|\mathcal{F}_s\}/Z(s) \qquad \text{for} \quad s \in [0,t].$$

Proof. Since Z is a martingale, we have

$$\tilde{\mathbf{E}}\{Y\} = \mathbf{E}\{YZ(T)\} = \mathbf{E}\Big\{Y\,\mathbf{E}\{Z(T)|\mathcal{F}_s\}\Big\} = \mathbf{E}\{YZ(s)\}$$

for $\mathcal{F}_s$-measurable random variables Y with $\tilde{\mathbf{E}}\{|Y|\} < \infty$. For $\Lambda \in \mathcal{F}_s$, this gives

$$\begin{aligned}
\tilde{\mathbf{E}}\{I_\Lambda Y\} = \mathbf{E}\{I_\Lambda YZ(T)\} = \mathbf{E}\Big\{\mathbf{E}\{I_\Lambda YZ(T)|\mathcal{F}_t\}\Big\} &= \mathbf{E}\Big\{I_\Lambda Y\,\mathbf{E}\{Z(T)|\mathcal{F}_t\}\Big\} \\
&= \mathbf{E}\{I_\Lambda YZ(t)\} \\
&= \mathbf{E}\Big\{\mathbf{E}\{I_\Lambda YZ(t)|\mathcal{F}_s\}\Big\} \\
&= \mathbf{E}\Big\{I_\Lambda\,\mathbf{E}\{YZ(t)|\mathcal{F}_s\}\Big\} \\
&= \tilde{\mathbf{E}}\Big\{I_\Lambda\,\mathbf{E}\{YZ(t)|\mathcal{F}_s\}/Z(s)\Big\}. \quad \square
\end{aligned}$$

Proof of Theorem 19.1. Since B is continuous, and $\int_0^\cdot X(r)\,dr$ has finite variation and is continuous, we have

$$[W,W](t) = [B,B](t) - 2\Big[\int_0^\cdot X(r)\,dr, B\Big](t) + \Big[\int_0^\cdot X(r)\,dr, \int_0^\cdot X(r)\,dr\Big](t) = t+0+0.$$

Hence, by *Lévy's characterization of BM*, it is enough to show that W is a local martingale under $\tilde{\mathbf{P}}$ (since W obviously is continuous, by definition).

Since Z is stochastic exponential of $dY(t) = X(t)\,dB(t)$ (by inspection of Definition 15.11), so that $dZ(t) = Z(t)\,dY(t) = Z(t)X(t)\,dB(t)$, integration by parts (Example 14.11), together with the definition of W, give

$$\begin{aligned}\int_0^t Z\,dB - \int_0^t Z(r)X(r)\,dr = \int_0^t Z\,dW \\ &= Z(t)W(t) - Z(0)W(0) - \int_0^t W\,dZ - [Z,W](t) \\ &= Z(t)W(t) - 0 - \int_0^t WZX\,dB - \int_0^t Z(r)X(r)\,dr,\end{aligned}$$

since $W(0)=0$ and

$$\begin{aligned}d[Z,W](t) = dZ(t)\,dW(t) &= (Z(t)X(t)\,dB(t))\,(dB(t)-X(t)dt) \\ &= Z(t)X(t)\,dB(t)^2 - o(dt) = Z(t)X(t)\,dt.\end{aligned}$$

Rearranging, we get that ZW is a local martingale, since

$$Z(t)W(t) = \int_0^t Z\,dB + \int_0^t WZX\,dB.$$

Pick stopping times $\{\tau_n\}_{n=1}^\infty$ such that $\tau_n \uparrow \infty$ a.s. and $ZW(\tau_n \wedge t)$ is a martingale. Let $\tilde{\tau}_n \equiv \inf\{t \geq 0 : |W(t)| \geq n\}$ and $\hat{\tau}_n \equiv \tau_n \wedge \tilde{\tau}_n$. Since W is continuous and adapted, by Exercise 64, $ZW(\hat{\tau}_n \wedge t) = ZW((\tau_n \wedge t)\wedge \tilde{\tau}_n)$ is a martingale, by *Optional Stopping*.

By Example 23.17, $Z(t)W(\hat{\tau}_n \wedge t)$ is martingale. Hence *Bayes' rule* gives

$$\tilde{\mathbf{E}}\{W(\hat{\tau}_n \wedge t)|\mathcal{F}_s\} = \mathbf{E}\{Z(t)W(\hat{\tau}_n \wedge t)|\mathcal{F}_s\}/Z(s) = Z(s)W(\hat{\tau}_n \wedge s)/Z(s) = W(\hat{\tau}_n \wedge s).$$

Consequently, $W(\hat{\tau}_n \wedge t)$ is a martingale, and W a local martingale, under $\tilde{\mathbf{P}}$. □

EXERCISE 104 Explain why $\tilde{\mathbf{P}}$ is a probability measure. Compute the Radon-Nikodym derivative $d\tilde{\mathbf{P}}/d\mathbf{P}$.

EXERCISE 105 One vesion of the *Cameron-Martin formula*, simply is Theorem 19.1 in the special case of non-random $X \in P_T$, that is, X is a (deterministic) function in $\mathbb{L}^2([0,T])$ (recall Exercise 73-74). (In this case, Z is always a martingale.) One indication of the depth of *Girsanov's theorem* is the difficulty of a direct proof of (the described version of) the *Cameron-Martin formula*: Try such a proof. (**Hint:** Use an adaption of the proof of *Lévy's characterization of BM*.)

19.2 Multidimensional SDE

We call an $\mathbb{R}^n$-valued stochastic process $\{B(t)\}_{t\geq 0}$ <u>n-dimensional BM</u>, if its components $B_1,\ldots,B_n$ are independent Brownian motions wrt. to a common filtration.

Given functions $\mu:[0,T]\times\mathbb{R}^n\to\mathbb{R}^n$ and $\sigma:[0,T]\times\mathbb{R}^n\to\mathbb{R}_{n|n}$ ($n|n$-matrices), with measurable components, we consider the *multidimensional diffusion type SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \text{for} \quad t\in[0,T], \qquad X(0)=X_0, \qquad (19.1)$$

where B is n-dimensional BM and X_0 is independent of B (see Section 16.2).

A strong solution to (19.1), is an $\mathbb{R}^n$-valued stochastic process $\{X(t)\}_{t\in[0,T]}$, with all components $\{\sqrt{|b_i(t,X(t))|}\}_{t\in[0,T]}$ and $\{\sigma_{i,j}(t,X(t))\}_{t\in[0,T]}$ in P_T, such that

$$X(t) = X_0 + \int_0^t \mu(t,X(t))\,dt + \int_0^t \sigma(t,X(t))\,dB(t) \qquad \text{for} \quad t\in[0,T].$$

Expressed componetwise, this means that (for each $i\in\{1,\ldots,n\}$)

$$X_i(t) = (X_0)_i + \int_0^t \mu_i(t,X(t))\,dt + \sum_{j=1}^n \int_0^t \sigma_{i,j}(t,X(t))\,dB_j(t) \qquad \text{for} \quad t\in[0,T].$$

[Here random beginning values X_0 are accomodated as described in Section 16.2.]

Any solution to a multidimensional SDE is a <u>multidimensional diffusion process</u>.

Methods for one-dimensional SDE developed so far, carry over to the multidimensional with only obvious changes. Except for *martingale problems* introduced below, we have in that way already covered basic theory for multidimensional SDE.

However, there exist a lot of results for the one-dimensional case, that do not carry over to the multidimensional setting. We will encounter such results then and then in the sequel. They constitute a classical subject matter, that was researched by "famous names" as Chung, Doob, Feller, Itô, Kolmogorov, Lévy, etc.

We now list multidimensional versions of important results, previously stated in one dimension. (Recall that we already have a multidimensional *Itô formula* from Section 16.1.) As mentioned, the proofs of these results are much the same as those in one dimension. See the one-dimensional results for references to the literature.

Introduce the vector norm $\underline{\|x\|} = \sqrt{x_1^2+\ldots+x_n^2}$ for $x\in\mathbb{R}^n$, and matrix norm $\underline{\|M\|} = \sqrt{M_{1,1}^2+\ldots+M_{n,n}^2}$ for $M\in\mathbb{R}_{n|n}$.

Theorem 19.3 (e.g., [22, Theorem 5.2.5])* *Consider the multidimensional SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \textit{for} \quad t\in[0,T], \qquad X(0)=X_0$$

(where X_0 is independent of B). Assume that, to each $N\in\mathbb{N}$, there exists a constant $K_N>0$ such that the following Lipschitz condition holds

$$\|\mu(t,x)-\mu(t,y)\|+\|\sigma(t,x)-\sigma(t,y)\| \le K_N\|x-y\| \qquad \textit{for} \quad \|x\|,\|y\|\le N \quad \textit{and} \quad t\in[0,T].$$

We have strong uniqueness for solutions to the multidimensional SDE.

(The *Yamada-Watanabe theorem* does not carry over to the multidimensional.)

Theorem 19.4 (e.g., [22, Problem 5.2.12])⋆ *Consider the multidimensional SDE*

$$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dB(t) \qquad \text{for} \quad t \in [0,T], \qquad X(0) = X_0$$

(where X_0 *is independent of* B*). Assume that, for some constant* $K > 0$*,*

$$\begin{cases} \|\mu(t,x) - \mu(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \le K\,\|x-y\| \\ \|\mu(t,x)\|^2 + \|\sigma(t,x)\|^2 \le K^2\,(1+\|x\|^2) \end{cases} \quad \text{for} \quad x,y \in \mathbb{R}^n \quad \text{and} \quad t \in [0,T].$$

There exists a unique strong solution to the multidimensional SDE.

Theorem 19.5 (Lévy's Characterization of BM) (e.g., [22, Theorem 3.2.16])⋆ *Let* $\{X(t)\}_{t\ge 0}$ *be an* $\mathbb{R}^n$*-valued stochastic process, the components of which are continuous local martingales wrt. a common filtration. The process* X *is* n*-dimensional BM iff. it has covariations*

$$[X_i, X_j](t) = \delta_{i,j} t \qquad \text{for} \quad t \ge 0 \quad \text{and} \quad i,j \in \{1,\ldots,n\}$$

(where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise).

Theorem 19.6 (Girsanov's Theorem) (e.g., [22, Theorem 3.5.1])⋆ *Let* B *be* n*-dimensional BM and let* $X_1,\ldots,X_n \in P_T$*. Assume that the stochastic process*

$$Z(t) \equiv \exp\left\{\sum_{i=1}^{n}\int_0^t X_i(r)\,dB_i(r) - \frac{1}{2}\int_0^t \|X(r)\|^2\,dr\right\}, \quad t \in [0,T],$$

is a martingale. Under the probability measure

$$\tilde{\mathbf{P}}\{A\} \equiv \mathbf{E}\{I_A Z(T)\}, \quad A \in \mathcal{F}_T,$$

the following stochastic process W *is an* n*-dimensional BM*

$$W(t) \equiv B(t) - B(0) - \int_0^t X(r)\,dr, \quad t \in [0,T].$$

Exercise 106 Explain how the multidimensional version of *Girsanov's theorem* can be used to find a weak solution to the multidimensional SDE

$$dX(t) = \mu(t, X(t))\,dt + dB(t) \qquad \text{for} \quad t \in [0,T].$$

19.3 One-Dimensional Time Homogeneous SDE

This section gives first examples on results that hold only for one-dimensional time homogeneous SDE, that is, SDE where the drift and dispersion do not depend on time, $\mu(t,x) = \mu(x)$ and $\sigma(t,x) = \sigma(x)$. Obviously, the method of proof does not carry over to the multidimensional, and is therefore not discussed.

Theorem 19.7 (ENGELBERT-SCHMIDT) (e.g., [22, pp. 332-334])* *The SDE*

$$dX(t) = \sigma(X(t))\,dB(t) \qquad \text{for} \quad t>0, \qquad X(0)=X_0,$$

has a weak solution for every choice of initial value X_0 *iff., for each* $x\in\mathbb{R}$, *we have*

$$\int_{-\varepsilon}^{\varepsilon} \frac{dy}{\sigma(x+y)^2} = \infty \quad \textit{for all} \quad \varepsilon>0 \quad \Rightarrow \quad \sigma(x)=0.$$

EXERCISE 107 Give an example of a non-continuous σ for which (by Theorem 19.7) there exists a weak solution to the SDE

$$dX(t) = \sigma(X(t))\,dB(t) \qquad \text{for} \quad t>0, \qquad X(0)=X_0.$$

Theorem 19.8 (ENGELBERT-SCHMIDT) (e.g., [22, p. 335])* *The SDE*

$$dX(t) = \sigma(X(t))\,dB(t) \qquad \text{for} \quad t>0, \qquad X(0)=X_0,$$

has a weak solution, that in addition is unique, for every choice of initial value X_0, *iff., for each* $x\in\mathbb{R}$, *we have*

$$\int_{-\varepsilon}^{\varepsilon} \frac{dy}{\sigma(x+y)^2} = \infty \quad \textit{for all} \quad \varepsilon>0 \quad \Leftrightarrow \quad \sigma(x)=0.$$

Corollary 19.9 *Let* $\sigma:\mathbb{R}\to\mathbb{R}$ *be continuous, or bounded away from zero (or both). The SDE*

$$dX(t) = \sigma(X(t))\,dB(t) \qquad \text{for} \quad t>0, \qquad X(0)=X_0,$$

has a weak solution, that in addition is unique, for every choice of initial value X_0.

A function $f(x)$ is bounded away from zero if $|f(x)|\geq\varepsilon$ for all x (for which the function is defined), for some constant $\varepsilon>0$.

The question of existence and uniqueness of solutions to a general one-dimensional time homogeneous SDE, with a general drift μ and dispersion σ, can be reduced to the case with zero drift attended to above, by means of *removal of drift* (when μ/σ^2 is locally integrable). See Corollary 26.6 below.

Example 19.10 Consider the SDE from Example 17.5

$$dX(t) = \operatorname{sign}(X(t))\,dB(t) \qquad \text{for} \quad t\in[0,T], \qquad X(0)=0.$$

By Corollary 19.9, it has a weak solution, that in addition is unique. #

Assume that we have a (weak or strong) solution X to the SDE

$$dX(t) = \sigma(X(t))\,dB(t) \qquad \text{for} \quad t>0, \qquad X(0)=X_0,$$

and that

$$dW(t) = dB(t) - \mu(X(t))/\sigma(X(t))\,dt.$$

It follows that

$$\sigma(X(t))\,dW(t) + \mu(X(t))\,dt = \sigma(X(t))\,dB(t) = dX(t), \qquad X(0)=X_0,$$

so that X is a weak solution to the general one-dimensional time-homogeneous diffusion type SDE $dX = \mu\,dt + \sigma\,dW$, if W is BM under some probability measure. By *Girsanov's theorem* (Theorem 18.3), this holds if the following process is a martingale

$$\exp\left\{\int_0^t \frac{\mu(X(t))}{\sigma(X(t))}\,dB(r) - \frac{1}{2}\int_0^t \frac{\mu(X(t))^2}{\sigma(X(t))^2}\,dr\right\}, \quad t\in[0,T].$$

19.4 Local Martingale Problems

Weak solutions to multidimensional SDE are studied by martingale methods, due to Stroock and Varadhan (see their famous and difficult monograph [36])*, where the SDE is related to an equivalent so called (local) *martingale problem*. It is uniqueness of solutions that are best dealt with in this way, while existence can be handled in other ways as well. Theory for the *Cauchy problem for parabolic PDE* come into play here. Unfortunately, these methods are a bit difficult and somewhat non-probabilistic. Nevertheless, one should know a little about what is going one here. This is so also because the approach connects to so called *diffusion theory*.

Exercise 108 Let $\hat{B}$ be $\mathbb{R}^n$-valued BM. Let the $\mathbb{R}^n$-valued process $\{Y(t)\}_{t\geq 0}$ be a weak solution to the multidimensional (not necessarily diffusion type) SDE

$$dY(t) = \hat{\mu}(t)\,dt + \hat{\sigma}(t)\,d\hat{B}(t) \qquad \text{for} \quad t\geq 0, \tag{19.2}$$

[where the processes $\{\hat{\mu}(t)\}_{t\geq 0}$ and $\{\hat{\sigma}(t)\}_{t\geq 0}$, with values in $\mathbb{R}^n$ and $\mathbb{R}_{n|n}$, respectively, satisfy $\sqrt{|\hat{\mu}_i|}, \hat{\sigma}_{i,j} \in P_T$ for $T>0$ and $i,j=1,\ldots,n$]. Define the operators

$$\begin{cases} (Ag)(t,x) \equiv \dfrac{1}{2}\displaystyle\sum_{i=1}^n\sum_{j=1}^n (\hat{\sigma}\hat{\sigma}^T)_{i,j}(t)\,\frac{\partial^2 g(t,x)}{\partial x_i \partial x_j} + \sum_{i=1}^n \hat{\mu}_i(t)\,\frac{\partial g(t,x)}{\partial x_i} \\ (Cg)(t,x) \equiv \quad \left(\displaystyle\sum_{i=1}^n \hat{\sigma}_{i,1}(t)\,\frac{\partial g(t,x)}{\partial x_i}, \ldots, \sum_{i=1}^n \hat{\sigma}_{i,n}(t)\,\frac{\partial g(t,x)}{\partial x_i}\right) \end{cases} \tag{19.3}$$

for functions $g \in \underline{\mathbb{C}^{1,2}([0,\infty)\times\mathbb{R}^n)} \equiv \{(\tilde{g}:[0,\infty)\times\mathbb{R}\to\mathbb{R}) : \tilde{g}$ has continuous partial derivatives $\frac{\partial \tilde{g}(t,x)}{\partial t}$ and $\frac{\partial^2 \tilde{g}(t,x)}{\partial x_i \partial x_j}$, $i,j=1,\ldots,n\}$. Show that

$$dg(t,Y(t)) - \Big(\partial_1 g(t,Y(t)) + (Ag)(t,Y(t))\Big)\,dt = \Big((Cg)(t,Y(t))\Big)^T dB(t).$$

Corollary 19.11 *Let the $\mathbb{R}^n$-valued process $\{Y(t)\}_{t\geq 0}$ be a weak solution to the multidimensional SDE* (19.2)*, and consider the operator A given by* (19.3)*. For $g\in\mathbb{C}^{1,2}([0,\infty)\times\mathbb{R}^n)$ the following process is a continuous local martingale*

$$g(t,Y(t))-g(0,Y(0))-\int_0^t\Big(\partial_1 g(r,Y(r))+(Ag)(r,Y(r))\Big)\,dr,\quad t\geq 0.$$

Proof. This follows from Exercise 108 together with Theorem 12.10. □

Given an $s\geq 0$, consider the multidimensional diffusion type SDE [cf. (19.1)]

$$X(t)=X(s)+\int_s^t\mu(r,X(r))\,dr+\int_s^t\sigma(r,X(r))\,dB(r)\qquad\text{for}\quad t\geq s. \tag{19.4}$$

The generator of this SDE is the second order partial differential operator (PDO)

$$(\mathcal{A}_t f)(x)\equiv\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(\sigma\sigma^T)_{i,j}(t,x)\,\frac{\partial^2 f(x)}{\partial x_i\partial x_j}+\sum_{i=1}^n\mu_i(t,x)\,\frac{\partial f(x)}{\partial x_i} \tag{19.5}$$

for $f\in\mathbb{C}^2(\mathbb{R}^n)$ [cf. (19.3)]. Here $\sigma\sigma^T:[0,\infty)\times\mathbb{R}^n\to\mathbb{R}_{n|n}$ is the diffusion matrix.

Corollary 19.12 *Let the $\mathbb{R}^n$-valued process $\{X(t)\}_{t\geq s}=\{X(t{+}s)\}_{t\geq 0}=\{\hat{X}(t)\}_{t\geq 0}$ be a weak solution to the multidimensional SDE* (19.4)*, with generator $\mathcal{A}_t$. For functions $f\in\mathbb{C}^2(\mathbb{R}^n)$, the following process is a continuous local martingale*

$$f(\hat{X}(t))-f(\hat{X}(0))-\int_0^t(\mathcal{A}_{r+s}f)(\hat{X}(r))\,dr,\quad t\geq 0.$$

Proof. Since $Y=\hat{X}$ solves (19.2), with $\hat{\sigma}(t)=\sigma(t{+}s,X(t{+}s))$, $\hat{\mu}(t)=\mu(t{+}s,X(t{+}s))$ and $\hat{B}(t)=B(t{+}s)-B(s)$ (which is also BM), Corollary 19.11 gives the result. □

Definition 19.13 *Consider the generator $\mathcal{A}_t$ in* (19.5)*, and pick an $s\geq 0$. A continuous adapted $\mathbb{R}^n$-valued process $\{X(t)\}_{t\geq 0}$ is a solution to the local martingale problem associated with $\mathcal{A}_{\cdot+s}$, if for each $f\in\mathbb{C}^2(\mathbb{R}^n)$, the following process is a continuous local martingale*

$$f(X(t))-f(X(0))-\int_0^t(\mathcal{A}_{r+s}f)(X(r))\,dr,\quad t\geq 0.$$

In the literature, local martingale problems are formulated analytically (without the process X), as measure theoretic statements (equivalent with Definition 19.13).

The next very important result helps us complete previous findings, on the relation between multidimensional diffusions and local martingale problems:

Theorem 19.14 (ABSOLUTE CONTINUITY WRT. BM) (e.g., [22, pp. 170-172 and 316-317])* *Let $\{M_i(t)\}_{t\geq 0}$ be continuous local martingales, and $\{\sigma_{i,j}(t)\}_{t\geq 0}$ measurable adapted processes, wrt. a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, for $i,j=1,\ldots,n$. Assume that*

$$[M_i,M_j](t) = \int_0^t (\sigma\sigma^T)_{i,j}(r)\,dr = \int_0^t \sum_{k=1}^n \sigma_{i,k}(r)\sigma_{j,k}(r)\,dr \quad \text{for } t\geq 0 \text{ and } i,j=1,\ldots,n.$$

There exists a BM $\{\tilde{B}(t)\}_{t\geq 0}$ in $\mathbb{R}^n$ wrt. a filtration $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}\supseteq\{\mathcal{F}_t\}_{t\geq 0}$, such that

$$M_i(t) = \int_0^t \sigma(r)\,d\tilde{B}(r) = \int_0^t \sum_{k=1}^n \sigma_{i,k}(r)\,d\tilde{B}_k(r)\,dr \qquad \text{for } t\geq 0 \text{ and } i=1,\ldots,n.$$

Proof for $n=1$. Let $\delta=1/\sigma$ for $\sigma\neq 0$, and $\delta=0$ for $\sigma=0$. Let $\{W(t)\}_{t\geq 0}$ be a BM that is independent of $\{\mathcal{F}_t\}_{t\geq 0}$, and put $\tilde{\mathcal{F}}_t = \sigma(\mathcal{F}_t, \sigma(W(s):s\leq t))$. The process

$$\tilde{B}(t) \equiv \int_0^t \delta(r)\,dM(r) + \int_0^t I_{\{\sigma(r)=0\}}\,dW(r) \qquad \text{for } t\geq 0$$

is well-defined, since $\int_0^t \delta^2\,d[M] = \int_0^t \delta(r)^2\sigma(r)^2\,dr \leq t$. Here $\int_0^t \delta\,dM$ is a continuous local martingale wrt. $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}$, since a local martingale wrt. $\{\mathcal{F}_t\}_{t\geq 0}$, and W is independent of $\{\mathcal{F}_t\}_{t\geq 0}$. In the same way, W is a BM wrt. $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}$, so that also $\int_0^t I_{\{\sigma=0\}}\,dW$ is a continuous local martingale wrt. $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}$. Further, we have

$$[\tilde{B}](t) = \int_0^t \delta^2\,d[M] + 2\int_0^t \delta I_{\{\sigma=0\}}\,d[M,W] + \int_0^t I_{\{\sigma=0\}}\,d[W] = \int_0^t (I_{\{\sigma\neq 0\}} + 0 + I_{\{\sigma=0\}})\,dr = t.$$

Hence *Lévy's Characterization of BM* shows that $\tilde{B}$ is BM. Moreover,

$$\int_0^t \sigma\,d\tilde{B} = \int_0^t I_{\{\sigma\neq 0\}}\sigma\,d\tilde{B} = \int_0^t I_{\{\sigma\neq 0\}}\sigma\,(\delta\,dM + I_{\{\sigma=0\}}dW) = \int_0^t I_{\{\sigma\neq 0\}}\,dM = M(t).$$

Here the last equality follows from Corollary 23.19 below, which shows that $\int_0^t I_{\{\sigma=0\}}\,dM = 0$, since its quadratic variation is $\int_0^t I_{\{\sigma=0\}}\,d[M] = \int_0^t I_{\{\sigma(r)=0\}}\sigma(r)^2\,dr = 0$. □

Theorem 19.15 *Consider the generator $\mathcal{A}_t$ given by (19.5). Pick a constant $s\geq 0$ (and an $\mathbb{R}^n$-valued random variable X_0). The multidimensional SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \text{for } t\geq s \tag{19.4}$$

has a weak solution $\{X(t)\}_{t\geq s}$ (with $X(s) =_{\text{distribution}} X_0$), iff. $\{\hat{X}(t)\}_{t\geq 0} = \{X(t+s)\}_{t\geq 0}$ solves the local martingale problem for $\mathcal{A}_{\cdot+s}$ (with $\hat{X}(0) =_{\text{distribution}} X_0$).

Proof for $s=0$. Since a weak solution to the SDE is a solution to the local martingale problem, by Corollary 19.12, it is enough to prove that a solution to the local martingale problem is a weak solution to the SDE: Let X solve the local martingale problem. Taking $f(x)=x_i$, we get that

$$M^{(i)}(t) \equiv X_i(t) - X_i(0) - \int_0^t \mu_i(r,X(r))\,dr \qquad \text{is a continuous local martingale}$$

for $i=1,\ldots,n$. By *Itô's formula* together with the hypothesis of Theorem 19.15,

$$
\begin{aligned}
df-\mathcal{A}_t f\,dt &= \sum_{i=1}^n \partial_i f\,dX_i+\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \partial_i\partial_j f\,d[X_i,X_j]-\mathcal{A}_t f\,dt\\
&= \sum_{i=1}^n \partial_i f\,dM^{(i)}+\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \partial_i\partial_j f\,d[M^{(i)},M^{(j)}]-\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(\sigma\sigma^T)_{i,j}\,\partial_i\partial_j f\,dt
\end{aligned}
$$

is the stochastic differential of a continuous local martingale, for every $f\in\mathbb{C}^2(\mathbb{R}^n)$. Since also $\sum_{i=1}^n \partial_i f\,dM^{(i)}(t)$ is the stochastic differential of a continuous local martingale [since $M^{(1)},\ldots,M^{(n)}$ are continuous local martingales], we get that

$$
\mathcal{M}(t)\equiv\int_0^t\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \partial_i\partial_j f\,d[M^{(i)},M^{(j)}](t)-\int_0^t\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(\sigma\sigma^T)_{i,j}\,\partial_i\partial_j f\,dt
$$

is a continuous local martingale. Further, $\mathcal{M}$ has bounded variation, since absolutely continuous, so that $[\mathcal{M}](t)=0$. Now Theorem 19.14 gives $\mathcal{M}(t)=0$, so that $d[M^{(i)},M^{(j)}](t)=(\sigma\sigma^T)_{i,j}(t,X(t))\,dt$. Using Theorem 19.14 again, it follows that

$$
M^{(i)}(t)=\sum_{k=1}^n\sigma_{i,k}(r,X(r))\,dB_k(r)\qquad\text{for}\quad i=1,\ldots,n.
$$

Combining this with the definition of $M^{(1)},\ldots,M^{(n)}$, we get (19.4). □

Definition 19.16 *The local martingale problem associated with the generator $\mathcal{A}_t$ in* (19.5) *is well-posed at zero [well-posed], if the local martingale problem for $\mathcal{A}_{(\cdot)}$ [for $\mathcal{A}_{\cdot+s}$] has a solution X such that $X(0)=_{\text{distribution}}X_0$, for each $\mathbb{R}^n$-valued random variable X_0 [and each $s\geq 0$], which has uniquely determined* `fidi's` (any other such solution must have the same `fidi's` as X).

Corollary 19.17 *Consider the generator $\mathcal{A}_t$ in* (19.5). *The SDE*

$$
dX(t)=\mu(t,X(t))\,dt+\sigma(t,X(t))\,dB(t)\qquad\text{for}\quad t\geq 0,\qquad X(0)=X_0,
$$

has a weak solution that is unique, for each $\mathbb{R}^n$-valued random variable X_0, iff. the local martingale problem associated with $\mathcal{A}_t$ is well-posed at zero.

Proof. Theorem 19.15 and a logical exercise. □

Corollary 19.18 *Consider the generator $\mathcal{A}_t$ in* (19.5). *The SDE*

$$
dX(t)=\mu(t,X(t))\,dt+\sigma(t,X(t))\,dB(t)\qquad\text{for}\quad t\geq s,\qquad X(s)=X_0,
$$

has a weak solution $\{X(t)\}_{t\geq s}$ that is unique, for each $\mathbb{R}^n$-valued random variable X_0 and each $s\geq 0$, iff. the local martingale problem associated with $\mathcal{A}_t$ is well-posed.

Proof. Corollary 19.17 applied at each $s\geq 0$. □

20 Twentieth Lecture

20.1 Martingale Problems

Instead of finding X such that $M^f(t) \equiv f(X(t)) - f(X(0)) - \int_0^t (\mathcal{A}_r f)(X(r))\,dr$ is a continuous local martingale for each $f \in \mathbb{C}^2(\mathbb{R}^n)$, it is more convenient to search for X such that $M^f(t)$ is a continuous martingale for $f \in \underline{\mathbb{C}_0^2(\mathbb{R}^n)} \equiv \{g \in \mathbb{C}^2(\mathbb{R}^n) : g(x) = 0$ for x outside a bounded set$\}$. This causes little loss of generality:

Corollary 20.1 *Let the* $\mathbb{R}^n$*-valued process* $\{X(t)\}_{t\geq s} = \{X(t+s)\}_{t\geq 0} = \{\hat{X}(t)\}_{t\geq 0}$ *be a weak solution to the multidimensional SDE* (19.4)*, with generator* $\mathcal{A}_t$*, where* σ *is locally bounded* (σ has locally bounded components)*. For functions* $f \in \mathbb{C}_0^2(\mathbb{R}^n)$

$$f(\hat{X}(t)) - f(\hat{X}(0)) - \int_0^t (\mathcal{A}_{r+s} f)(\hat{X}(r))\,dr, \quad t \geq 0, \qquad \textit{is a continuous martingale.}$$

Proof. Since the process under consideration is a continuous local martingale, by Corollary 19.12, that is locally bounded, by the assumptions on f and σ together with an inspection of Exercise 108, it is a (continuous) martingale by Theorem 12.9. □

Definition 20.2 *Consider the generator* $\mathcal{A}_t$ *in* (19.5)*, and pick an* $s \geq 0$*. A continuous adapted* $\mathbb{R}^n$*-valued process* $\{X(t)\}_{t\geq 0}$ *is a* <u>*solution to the martingale problem*</u> *associated with* $\mathcal{A}_{\cdot+s}$*, if for each* $f \in \mathbb{C}_0^2(\mathbb{R}^n)$

$$f(X(t)) - f(X(0)) - \int_0^t (\mathcal{A}_{r+s} f)(X(r))\,dr, \quad t \geq 0, \qquad \textit{is a continuous martingale.}$$

Definition 20.3 *The martingale problem associated with the generator* $\mathcal{A}_t$ *in* (19.5) *is* <u>*well-posed at zero*</u> *[*<u>*well-posed*</u>*], if the martingale problem for* $\mathcal{A}_t$ *[for* $\mathcal{A}_{t+s}$*] has a solution* X *such that* $X(0) =_{\text{distribution}} X_0$*, for each* $\mathbb{R}^n$*-valued random variable* X_0 *[and each* $s \geq 0$*], and which has uniquely determined* `fidi's`.

Corollary 20.4 *Consider the generator* $\mathcal{A}_t$ *given by* (19.5)*, where* σ *is locally bounded. Pick an* $s \geq 0$ *(and an* $\mathbb{R}^n$*-valued random variable* X_0*). The SDE*

$$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dB(t) \qquad \textit{for} \quad t \geq s$$

has a weak solution $\{X(t)\}_{t\geq s}$ *(with* $X(s) =_{\text{distribution}} X_0$*), iff.* $\{\hat{X}(t)\}_{t\geq 0} = \{X(t+s)\}_{t\geq 0}$ *solves the martingale problem for* $\mathcal{A}_{\cdot+s}$ *(with* $\hat{X}(0) =_{\text{distribution}} X_0$*).*

Proof. The implication to the right is Corollary 20.1, in the special case $f \in \mathbb{C}_0^2(\mathbb{R}^n)$. For that to the left, let X solve the martingale problem associated with $\mathcal{A}_{\cdot+s}$. By

Theorem 19.15, it is enough to show that X solves the local martingale problem

$$M(t) \equiv f(X(t)) - f(X(0)) - \int_0^t (\mathcal{A}_{r+s} f)(X(r))\,dr \qquad \text{is a continuous local martingale}$$

for $f \in \mathbb{C}^2(\mathbb{R}^n)$. Pick $f^{(k)}(x)$ in $\mathbb{C}_0^2(\mathbb{R}^n)$ that agree with $f(x)$ for $\|x\| \leq k$, so that

$$M^{(k)}(t) \equiv f^{(k)}(X(t)) - f^{(k)}(X(0)) - \int_0^t (\mathcal{A}_{r+s} f^{(k)})(X(r))\,dr \quad \text{is a continuous martingale}$$

(since X solves the martingale problem). Define the stopping times $\tau_k \equiv \inf\{t \geq 0 : \|X(t)\| \geq k\}$ for $k \in \mathbb{N}$. By continuity of X, we have $\tau_k \uparrow \infty$ a.s. as $k \to \infty$. By Theorem 12.6, it is thus enough to show that $\{M(t \wedge \tau_k)\}_{t \geq 0}$ is a martingale.

By construction of τ_k, we have the second of the following equalities

$$\begin{aligned} M(t\wedge\tau_k) &= f(X(t\wedge\tau_k)) - f(X(0)) - \int_0^{t\wedge\tau_k} (\mathcal{A}_{r+s} f)(X(r))\,dr \\ &= f^{(k)}(X(t\wedge\tau_k)) - f(X(0)) - \int_0^{t\wedge\tau_k} (\mathcal{A}_{r+s} f^{(k)})(X(r))\,dr \\ &= f^{(k)}(X(t\wedge\tau_k)) - f^{(k)}(X(0)) - \int_0^{t\wedge\tau_k} (\mathcal{A}_{r+s} f^{(k)})(X(r))\,dr = M^{(k)}(t\wedge\tau_k). \end{aligned}$$

To see that also the third equality holds (the first one being trivial), notice that when $f(X(0)) \neq f^{(k)}(X(0))$, we have $\|X(0)\| > k$, so that $\tau_k = 0$. This gives

$$M(t\wedge\tau_k) = 0 = f^{(k)}(X(t\wedge\tau_k)) - f^{(k)}(X(0)) - \int_0^{t\wedge\tau_k} (\mathcal{A}_{r+s} f^{(k)})(X(r))\,dr.$$

Now $M(t\wedge\tau_k) = M^{(k)}(t\wedge\tau_k)$ is a martingale, by Theorem 12.3, since $M^{(k)}$ is. □

Corollary 20.5 *Let the generator $\mathcal{A}_t$ in (19.5) have σ locally bounded. The SDE*

$$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dB(t) \qquad \textit{for} \quad t \geq 0, \qquad X(0) = X_0,$$

has a weak solution X that is unique, for each $\mathbb{R}^n$-valued random variable X_0, iff. the martingale problem associated with $\mathcal{A}_t$ is well-posed at zero. In this case, solutions to the martingale problem can be chosen with same `fidi's` *as solutions to the SDE.*

Proof. Corollary 20.4 and a logical exercise. □

Corollary 20.6 *Let the generator $\mathcal{A}_t$ in (19.5) have σ locally bounded. The SDE*

$$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dB(t) \qquad \textit{for} \quad t \geq s, \qquad X(s) = X_0,$$

has a weak solution $\{X(t)\}_{t \geq s}$ that is unique, for each $\mathbb{R}^n$-valued random variable X_0 and each $s \geq 0$, iff. the martingale problem associated with $\mathcal{A}_t$ is well-posed. In this case, a solution $\{\hat{X}(t)\}_{t \geq 0}$ to the martingale problem for $\mathcal{A}_{\cdot+s}$ can be chosen so that $\{\hat{X}(t)\}_{t \geq 0} =_{\text{same fidi's}} \{X(t+s)\}_{t \geq 0}$.

20.2 Existence of Solutions to Martingale Problems

For strong solutions to SDE, uniqueness criteria (Theorems 16.8, 17.2 and 19.3) are less demanding (in terms of their hypothesises), than existence criteria (Theorems 17.7 and 19.4). For weak solutions to time homogeneous one-dimensional SDE, the uniqueness criterion (Theorem 19.8) is instead more demanding than the existence criterion (Theorem 19.7). This continues for solutions to (time homogeneous multidimensional) martingale problems, where uniqueness is both more demanding and more difficult than existence (which requires rather little).

The existence proof uses weak convergence of probability measures in an *Euler iteration*. The standard reference for this is [5]. For our special setting, with multidimensional SDE, the following adaption of what can be found there is convenient (the proof of which belongs in a course on weak convergence):

Lemma 20.7 (e.g., [22, Section 2.4.B])⋆ *Let* $\{X^{(1)}(t)\}_{t\geq 0}, \{X^{(2)}(t)\}_{t\geq 0}, \ldots$ *be continuous* $\mathbb{R}^n$*-valued processes* [on our basic probability space $(\Omega,\mathcal{F},\mathbf{P})$], *with*

$$\lim_{\lambda\to\infty}\sup_{k\geq 1}\mathbf{P}\{\|X^{(k)}(0)\|>\lambda\}=0, \tag{20.1}$$

and such that there exist constants $K,\alpha,\beta>0$ *with*

$$\mathbf{E}\Big\{\|X^{(k)}(t)-X^{(k)}(s)\|^{\alpha}\Big\}<K\,|t-s|^{1+\beta} \quad \text{for} \quad s,t\geq 0 \quad \text{and} \quad k\in\mathbb{N}. \tag{20.2}$$

There exist a continuous $\mathbb{R}^n$*-valued process* $\{X(t)\}_{t\geq 0}$ [on $(\Omega,\mathcal{F},\mathbf{P})$], *and a subsequence* $\{k_j\}_{j=1}^{\infty}\subseteq\mathbb{N}$*, such that for each function* $F:\mathbb{C}([0,T])^n\to\mathbb{R}$ *with*

$$|F(f)-F(f_k)|\to 0 \qquad \text{whenever} \quad \sup_{t\in[0,T]}\|f(t)-f_k(t)\|\to 0$$

as $k\to\infty$ [where $f,f_k\in\mathbb{C}([0,T])^n$], *we have*

$$F(\{X_{k_j}(t)\}_{t\in[0,T]})\to_{\text{distribution}}F(\{X(t)\}_{t\in[0,T]}) \qquad \text{as} \quad j\to\infty.$$

Theorem 20.8 (Stroock-Varadhan) *Consider the generator* $\mathcal{A}_t$ *in* (19.5), *where* $\mu:[0,\infty)\times\mathbb{R}^n\to\mathbb{R}^n$ *and* $\sigma:[0,\infty)\times\mathbb{R}^n\to\mathbb{R}_{n|n}$ *are continuous and bounded. For each* $\mathbb{R}^n$*-valued random variable* X_0 *and each* $s\geq 0$*, the martingale problem associated with* $\mathcal{A}_{\cdot+s}$ *has a solution* X *such that* $X(0)=_{\text{distribution}}X_0$.

⋆*Proof for* $s=0$ (adaption of [22, Section 5.4.D]). Define $\{X^{(k)}(t)\}_{t\geq 0}$ recursively by

$$X^{(k)}(t)\equiv X^{(k)}(\tfrac{\ell}{k})+\mu(\tfrac{\ell}{k},X^{(k)}(\tfrac{\ell}{k}))\,(t-\tfrac{\ell}{k})+\sigma(\tfrac{\ell}{k},X^{(k)}(\tfrac{\ell}{k}))\,(B(t)-B(\tfrac{\ell}{k}))$$

for $t\in(\frac{\ell}{k},\frac{\ell+1}{k}]$ and $\ell\in\mathbb{N}$, for each $k\in\mathbb{N}$, where $X^{(k)}(0)\equiv X_0$. Notice that, writing

$$\mu^{(k)}(t)=\mu(\tfrac{\ell}{k},X^{(k)}(\tfrac{\ell}{k})) \qquad \text{and} \qquad \sigma^{(k)}(t)=\sigma(\tfrac{\ell}{k},X^{(k)}(\tfrac{\ell}{k})) \qquad \text{for} \quad t\in(\tfrac{\ell}{k},\tfrac{\ell+1}{k}],$$

we have the (non-diffusion type) SDE

$$X^{(k)}(t) = X_0 + \int_0^t \mu^{(k)}(r)\,dr + \int_0^t \sigma^{(k)}(r)\,dB(r) \qquad \text{for} \qquad t \geq 0: \tag{20.3}$$

This is so by an induction argument, since (20.3) holds trivially for $t=0$, and if (20.3) holds for $t\in[0,\frac{\ell}{k}]$, then we get for $t\in(\frac{\ell}{k},\frac{\ell+1}{k}$

$$\begin{aligned}
&X_0 + \int_0^t \mu^{(k)}(r)\,dr + \int_0^t \sigma^{(k)}(r)\,dB(r)\\
&= \Big(X_0 + \int_0^{\ell/k} \mu^{(k)}(r)\,dr + \int_0^{\ell/k} \sigma^{(k)}(r)\,dB(r)\Big) + \Big(\int_{\ell/k}^t \mu^{(k)}(r)\,dr + \int_{\ell/k}^t \sigma^{(k)}(r)\,dB(r)\Big)\\
&= X^{(k)}(\tfrac{\ell}{k}) + \Big(\mu(\tfrac{\ell}{k},X^{(k)}(\tfrac{\ell}{k}))\,(t-\tfrac{\ell}{k}) + \sigma(\tfrac{\ell}{k},X^{(k)}(\tfrac{\ell}{k}))\,(B(t)-B(\tfrac{\ell}{k}))\Big) = X^{(k)}(t).
\end{aligned}$$

We have (20.1) trivially, since $X^{(k)}(0)=X_0$. Further, (20.2) holds with $\alpha=4$ and $\beta=1$, since [by (20.3) and the inequality $|\sum_{i=1}^n x_i|^m \leq n^{m-1}\sum_{i=1}^n |x_i|^m$ for $m\geq 1$]

$$\begin{aligned}
&\mathbf{E}\Big\{\|X^{(k)}(t)-X^{(k)}(s)\|^4\Big\}\\
&= \mathbf{E}\bigg\{\bigg\|\int_s^t \mu^{(k)}(r)\,dr + \int_s^t \sigma^{(k)}(r)\,dB(r)\bigg\|^4\bigg\}\\
&\leq 8\,\mathbf{E}\bigg\{\bigg\|\int_s^t \mu^{(k)}(r)\,dr\bigg\|^4\bigg\} + 8\,\mathbf{E}\bigg\{\bigg\|\int_s^t \sigma^{(k)}(r)\,dB(r)\bigg\|^4\bigg\}\\
&= 8\,\mathbf{E}\bigg\{\bigg(\sum_{i=1}^n\Big(\int_s^t \mu_i^{(k)}(r)\,dr\Big)^2\bigg)^2\bigg\} + 8\,\mathbf{E}\bigg\{\bigg(\sum_{i=1}^n\Big(\sum_{j=1}^n\int_s^t \sigma_{i,j}^{(k)}(r)\,dB_j(r)\Big)^2\bigg)^2\bigg\}\\
&\leq 8n\sum_{i=1}^n \mathbf{E}\bigg\{\Big(\int_s^t \mu_i^{(k)}(r)\,dr\Big)^4\bigg\} + 8n\sum_{i=1}^n \mathbf{E}\bigg\{\Big(\sum_{j=1}^n\int_s^t \sigma_{i,j}^{(k)}(r)\,dB_j(r)\Big)^4\bigg\}\\
&\leq 8n\sum_{i=1}^n \mathbf{E}\bigg\{\Big(\int_s^t \mu_i^{(k)}(r)\,dr\Big)^4\bigg\} + 8n^4\sum_{i=1}^n\sum_{j=1}^n \mathbf{E}\bigg\{\Big(\int_s^t \sigma_{i,j}^{(k)}(r)\,dB_j(r)\Big)^4\bigg\}\\
&\leq 8n\sum_{i=1}^n\Big((t-s)\sup_{r\geq 0,\,x\in\mathbb{R}^n}|\mu_i(r,x)|\Big)^4 + 24n^4(t-s)^2\sum_{i=1}^n\sum_{j=1}^n\sup_{r\geq 0,\,x\in\mathbb{R}^n}\sigma_{i,j}(r,x)^4,
\end{aligned}$$

by Exercise 109 below (recall the definition of $\mu^{(k)}$ and $\sigma^{(k)}$ in terms of μ and σ).

Define the second order PDO [cf. (19.3)]

$$(A^{(k)}f)(t,x) \equiv \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(\sigma^{(k)}(\sigma^{(k)})^T)_{i,j}(t)\,\frac{\partial^2 f(x)}{\partial x_i\partial x_j} + \sum_{i=1}^n \mu_i^{(k)}(t)\,\frac{\partial f(x)}{\partial x_i}$$

for $f\in\mathbb{C}_0^2(\mathbb{R}^n)$. Directly by application of Lemma 20.6, there exist a continuous process X and a subsequence $\{k_j\}_{j=1}^\infty$, such that

$$\Big(f(X^{(k_j)}(t)) - f(X^{(k_j)}(s)) - \int_s^t(\mathcal{A}_r f)(X^{(k_j)}(r))\,dr\Big)g(\{X^{(k_j)}(r)\}_{r\in[0,s]})$$

converges in distribution as $j\to\infty$, to

$$\Big(f(X(t)) - f(X(s)) - \int_s^t(\mathcal{A}_r f)(X(r))\,dr\Big)g(\{X(r)\}_{r\in[0,s]}),$$

for $f\in\mathbb{C}_0^2(\mathbb{R}^n)$, $0\leq s<t$ and $g:\mathcal{C}([0,s])^n\to\mathbb{R}$ with $|g(f_k)-g(f)|\to 0$ as $\sup_{r\in[0,s]}$

$\|f_k(r)-f(r)\|\to 0$. In fact, by continuity and boundedness of σ and μ, also

$$\Big(f(X^{(k_j)}(t))-f(X^{(k_j)}(s))-\int_s^t (A^{(k_j)}f)(r,X^{(k_j)}(r))\,dr\Big)g(\{X^{(k_j)}(r)\}_{r\in[0,s]}) \tag{20.4}$$

converges in distribution to this same limit. It is enough to prove that

$$M(t)\equiv f(X(t))-f(X(s))-\int_s^t (\mathcal{A}_r f)(X(r))\,dr,\quad t\geq 0,$$

is a martingale wrt. itself $\mathbb{G}=\{\mathcal{G}_t\}_{t\geq 0}\equiv\{\sigma(X(s):0\leq s\leq t)\}_{t\geq 0}$, that is

$$\mathbf{E}\{M(t)-M(s)|\mathcal{G}_s\}=\mathbf{E}\Big\{f(X(t))-f(X(s))-\int_s^t (\mathcal{A}_r f)(X(r))\,dr\,\Big|\,\mathcal{G}_s\Big\}=0$$

for $0\leq s<t$. This is the same thing as

$$\mathbf{E}\Big\{\Big(f(X(t))-f(X(s))-\int_s^t (\mathcal{A}_r f)(X(r))\,dr\Big)I_\Lambda\Big\}=0\qquad\text{for}\quad \Lambda\in\mathcal{G}_s.$$

Because of the choice of filtration $\mathbb{G}$, and a standard approximation, this holds if

$$\mathbf{E}\Big\{\Big(f(X(t))-f(X(s))-\int_s^t (\mathcal{A}_r f)(X(r))\,dr\Big)g(\{X(r)\}_{r\in[0,s]})\Big\}=0, \tag{20.5}$$

with g as above. The random variables in (20.4) are bounded, and thus uniformly integrable (cf. Exercise 77). Hence (11.6) shows that

$$\mathbf{E}\Big\{\Big(f(X^{(k_j)}(t))-f(X^{(k_j)}(s))-\int_s^t (A^{(k_j)}f)(r,X^{(k_j)}(r))\,dr\Big)g(\{X^{(k_j)}(r)\}_{r\in[0,s]})\Big\} \tag{20.6}$$

converges to the left-hand side of (20.5). However, since it is locally bounded,

$$M^{(k)}(t)\equiv f(X^{(k)}(t))-f(X^{(k)}(0))-\int_0^t (A^{(k)}f)(r,X^{(k)}(r))\,dr\qquad\text{is a martingale,}$$

wrt. the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ associated with B, by (20.3) together with Corollary 19.11 and Theorem 12.9. Since $g(\{X^{(k)}(r)\}_{r\in[0,s]})$ is adapted to $\mathcal{G}_s\subseteq\mathcal{F}_s$, this gives

$$\begin{aligned}&\mathbf{E}\Big\{\Big(f(X^{(k)}(t))-f(X^{(k)}(s))-\int_s^t (A^{(k)}f)(r,X^{(k)}(r))\,dr\Big)g(\{X^{(k)}(r)\}_{r\in[0,s]})\Big\}\\ &=\mathbf{E}\Big\{\mathbf{E}\Big\{f(X^{(k)}(t))-f(X^{(k)}(s))-\int_s^t (A^{(k)}f)(r,X^{(k)}(r))\,dr\,\Big|\,\mathcal{F}_s\Big\}g(\{X^{(k)}(r)\}_{r\in[0,s]})\Big\}=0.\end{aligned}$$

Hence the expectation (20.6) is zero, which establishes (20.5). □

Exercise 109 Verify the facts used in the proof of Theorem 20.8, that $|\sum_{i=1}^n x_i|^m\leq n^{m-1}\sum_{i=1}^n|x_i|^m$ and $\mathbf{E}\{(\int_s^t\sigma_{i,j}^{(k)}\,dB_j)^4\}\leq 3\,(t-s)^2\sup_{r,x}\sigma_{i,j}(r,x)^4$.

The following result where originally proved without the use of martingales:

Corollary 20.9 (Skorohod) *For each $\mathbb{R}^n$-valued random variable X_0 and each $s\geq 0$, and for each choice of continuous and bounded coefficients $\mu:[0,\infty)\times\mathbb{R}^n\to\mathbb{R}^n$ and $\sigma:[0,\infty)\times\mathbb{R}^n\to\mathbb{R}_{n|n}$, the following multidimensional SDE has a weak solution*

$$dX(t)=\mu(t,X(t))\,dt+\sigma(t,X(t))\,dB(t)\qquad\textit{for}\quad t\geq s,\qquad X(s)=X_0.$$

Proof. Theorem 20.8 together with Corollary 20.4. □

20.3 PDO and PDE

The material in this section has been collected from [14].

A linear partial differential operator (PDO) P of order m takes the form

$$(Pu)(x) = \sum_{\substack{\alpha_1,\ldots,\alpha_n \in \mathbb{N} \\ \alpha_1+\ldots+\alpha_n \leq m}} a_{\alpha_1,\ldots,\alpha_n}(x) \frac{\partial^{\alpha_1}\ldots\partial^{\alpha_n} u(x)}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}} \qquad \text{for functions} \quad u:\mathbb{R}^n\to\mathbb{R}.$$

Here $m\in\mathbb{N}$, $m\geq 1$, is assumed to be chosen as small as possible, that is, $\sum_{\alpha_1+\ldots+\alpha_n\leq m} \neq \sum_{\alpha_1+\ldots+\alpha_n\leq m-1}$. The functions $a_{\alpha_1,\ldots,\alpha_n}:\mathbb{R}^n\to\mathbb{R}$ are called coefficients. If desired, the function u may be replaced with (generalized to) a (Schwarz) distribution.

When $n=1$, the PDO reduces to an ordinary differential operator (ODO).

Associated with the PDO is the characteristic polynomial

$$p(x,\xi) = \sum_{\substack{\alpha_1,\ldots,\alpha_n \in \mathbb{N} \\ \alpha_1+\ldots+\alpha_n = m}} a_{\alpha_1,\ldots,\alpha_n}(x)\, \xi_1^{\alpha_1}\cdot\ldots\cdot\xi_n^{\alpha_n} \qquad \text{for} \quad \xi\in\mathbb{C}^n.$$

Notice that only terms corresponding to the highest order derivatives appear here.

Definition 20.10 *Consider a PDO P in an open region $D\subseteq\mathbb{R}^n$, together with a function (distribution) $f:D\to\mathbb{R}$. The equation* [to find a function (distribution) $u:\mathbb{R}^n\to\mathbb{R}$ such that]

$$(Pu)(x) = f(x) \qquad \textit{for} \quad x\in D$$

is a partial differential equation (PDE).

For a PDO P with constant coefficients [the functions $a_{\alpha_1,\ldots,\alpha_n}(x)$ are constants (do not depend on $x\in\mathbb{R}^n$)], it is known that, the PDE

$$(Pu)(x) = f(x) \qquad \text{has a distribution solution} \quad u(x), \tag{20.7}$$

for each function

$$f \in \mathbb{C}_0^\infty(\mathbb{R}^n) \equiv \Big\{ f\in\mathbb{C}^\infty(\mathbb{R}^n) : f(x)=0 \text{ for } x \text{ outside a bounded set}\Big\},$$

where

$$\mathbb{C}^\infty(\mathbb{R}^n) \equiv \Big\{ (g:\mathbb{R}^n\to\mathbb{R}) : g \text{ has continuous partial derivatives of all orders}\Big\}.$$

A *fundamental solution* to a PDO P, is a function (distribution) $\mathcal{E}$ that sort of inverts P, so that $P\mathcal{E}$ is the unity operator, for convolutions, the Dirac distribution $\delta(x)$. Here is a first definition (that will be more worked upon later):

Definition 20.11 *For a PDO P with constant coefficients, a fundamental solution in an open region $D\subseteq\mathbb{R}^n$, is a distribution e which solves the PDE*

$$(Pe)(x) = \delta(x) \qquad \textit{for} \quad x\in D.$$

Definition 20.12 *For a PDO P a fundamental kernel in an open region $D \subseteq \mathbb{R}^n$, is a distribution $E(\cdot,\cdot)$ which solves the PDE*

$$(PE(\cdot,y))(x) = \delta(x-y) \qquad \textit{for} \quad x,y \in D.$$

Exercise 110 Show that if e is a fundamental solution to a PDO P with constant coefficients, then $E(x,y) = e(x-y)$ is a fundamental kernel to P. Show (at least formally) that if E is a fundamental kernel to a PDO P, in an open region $D \subseteq \mathbb{R}^n$, then the PDE $(Pu)(x) = f(x)$ for $x \in D$, has solution $u(x) = \int_D E(x,y)\,f(y)\,dy$.

For PDE with variable coefficients, such simple results as (20.7) are not available. Now PDE's are classified into categories, which are studied separately, by different methods. Next, we introduce the classical such categories, for second order PDO.

Definition 20.13 *A second order PDO is of elliptic [degenerate elliptic] type in a region $D \subseteq \mathbb{R}^n$, if the characteristic polynomial satisfies*

$$p(x,\xi) > 0 \qquad [p(x,\xi) \geq 0] \qquad \textit{for all} \quad \xi \in \mathbb{R}^n \backslash \{0\}, \qquad \textit{for each} \quad x \in D.$$

Definition 20.14 *A second order PDO is of parabolic type in a region $D \subseteq \mathbb{R}^n$, if the characteristic polynomial satisfies*

$$p(x,\xi) = 0 \qquad \textit{for some} \quad \xi \in \mathbb{R}^n \backslash \{0\}, \qquad \textit{for each} \quad x \in D.$$

Definition 20.15 *A second order PDO P is of hyperbolic type in the j'the coordinate, in a region $D \subseteq \mathbb{R}^n$, if $a_{j,j}(x) = 1$ (so that $\frac{\partial^2}{\partial x_j^2}$ appears in P), and for every $(\xi_1, \ldots, \xi_{j-1}, \xi_{j+1}, \ldots, \xi_n) \in \mathbb{R}^{n-1} \backslash \{0\}$, the characteristic polynomial satisfies*

$$p(x, \xi_1, \ldots, \xi_{j-1}, \lambda, \xi_{j+1}, \ldots, \xi_n) = 0 \qquad \textit{for two distinct} \quad \lambda \in \mathbb{R}, \qquad \textit{for each} \quad x \in D.$$

Example 20.16 The Laplacian $\Delta = \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$ is elliptic. The heat operator $\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x_1^2} - \ldots - \frac{\partial^2}{\partial x_{n-1}^2}$ is parabolic. The wave *operator* $\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \ldots - \frac{\partial^2}{\partial x_{n-1}^2}$ is hyperbolic. The PDO $\mathcal{A}_t$ in (19.5) is (possibly degenerate) elliptic for each $t \geq 0$. #

A second order PDO P is of mixed type in a region D, if D can be divided into sub-regions, each of which in which P is of one of the above three types.

Example 20.17 *Tricomi's PDO* $\frac{\partial^2}{\partial x^2} - x\,\frac{\partial^2}{\partial y^2}$ in $\mathbb{R}^2$ is of mixed type, since it is elliptic for $x < 0$, parabolic for $x = 0$, and hyperbolic for $x > 0$. #

Exercise 111 Explain the labels "elliptic", "parabolic" and "hyperbolic".

*Remark 20.18 An important class of PDO are those of *hypoelliptic* type, where

$$\text{the } symbol \qquad \mathcal{P}(x,\xi) \equiv \sum_{\substack{\alpha_1,\ldots,\alpha_n \in \mathbb{N} \\ \alpha_1+\ldots+\alpha_n \le m}} a_{\alpha_1,\ldots,\alpha_n}(x)\,\xi_1^{\alpha_1}\cdot\ldots\cdot\xi_n^{\alpha_n} \qquad \text{for} \quad \xi\in\mathbb{C}^n,$$

satisfies

$$\mathcal{P}(x,\xi^{(k)}) = 0 \quad \text{and} \quad \|\xi^{(k)}\|\to\infty \quad \Rightarrow \quad \|\Im(\xi^{(k)})\|\to\infty.$$

This includes, for example, elliptic PDO, the parabolic heat PDO, and all ODO.

For a PDO P with symbol $\mathcal{P}$, we have (formally)

$$(Pu)(x) = \frac{1}{(2\pi)^n}\int_{\xi\in\mathbb{R}^n} \mathrm{e}^{-i\langle x,\xi\rangle}\,\mathcal{P}(x,i\xi)\left(\int_{y\in\mathbb{R}^n} \mathrm{e}^{i\langle y,\xi\rangle}\,u(y)\,dy\right)d\xi. \tag{20.8}$$

A *pseudodifferential operator* is a linear operation P on functions, such that (20.8) holds, with the symbol $\mathcal{P}(x,i\xi)$ no longer necessarily a polynomial, but "polynomial-like" (e.g., a fractional polynomial). (How polynomial-like is a matter of choice.) #

In Section 21.3 we consider *uniqueness and representation of solutions*, by means of stochastic calculus, for two basic PDE problems, which we now introduce.

Definition 20.19 *A <u>Cauchy problem,</u> is a PDE*

$$(Pu)(x) = f(x) \qquad \textit{for} \quad x\in D, \qquad \textit{in an open region} \quad D\subseteq\mathbb{R}^n,$$

together with certain <u>initial conditions</u> on the value of the solution u and some of its derivatives, at some point in closure(D) *[or more general subset of* closure(D)*].*

Definition 20.20 *A <u>Dirichlet problem,</u> is a PDE*

$$(Pu)(x) = f(x) \qquad \textit{for} \quad x\in D, \qquad \textit{in an open region} \quad D\subseteq\mathbb{R}^n,$$

together with certain <u>boundary conditions</u> on the value of the solution u and some of its derivatives on boundary(D).

It is far from unique what "Dirichlet problem" and "Cauchy problem" mean, and our definitions are non-technical, and much more general than is practice.

The PDE problems we will consider are the following two, both very important:

(1) *The Cauchy problem for a general nonnegative second order parabolic PDE*

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} + (\mathcal{A}_t u(t,\cdot))(t,x) + k(t,x)\,u(t,x) = g(t,x) & \text{for} \quad (t,x)\in(0,T)\times\mathbb{R}^n \\ u(T,x) = f(x) & \text{for} \quad x\in\mathbb{R}^n \end{cases},$$

where $k,g:(0,T)\times\mathbb{R}^n\to\mathbb{R}$ and $f:\mathbb{R}^n\to\mathbb{R}$ are given functions, and $\mathcal{A}_t$ is the

PDO given by (19.5). Moreover, we find a fundamental solution to this PDO.

(2) *The Dirichlet problem for general second order (possibly degenerate) elliptic PDE*

$$\begin{cases} (\mathcal{A}u)(x) + k(x)\,u(x) = g(x) & \text{for} \quad x \in D \\ u(x) = f(x) & \text{for} \quad x \in \text{boundary}(D) \end{cases},$$

where $k, g : \mathbb{R}^n \to \mathbb{R}$ and $f : \mathbb{R}^n \to \mathbb{R}$ are given functions, $D \subseteq \mathbb{R}^n$ is an open bounded set. Further, $\mathcal{A}$ is the (possibly degenerate) elliptic PDO given by

$$(\mathcal{A}_t f)(x) = \underline{(\mathcal{A}f)(x)} \equiv \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\sigma\sigma^T)_{i,j}(x)\,\frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{n}\mu_i(x)\,\frac{\partial f(x)}{\partial x_i}, \tag{20.9}$$

where $\sigma : \mathbb{R}^n \to \mathbb{R}_{n|n}$ and $\mu : \mathbb{R}^n \to \mathbb{R}^n$ are measurable functions. This means that [cf. (19.5)] $\mathcal{A}$ is the generator of the time homogeneous multidimensional SDE

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t) \qquad \text{for} \quad t \geq 0.$$

Notice that, taking $v(t,x) = u(T-t,x)$, **(1)** becomes the ordinary heat equation

(1′) *the Cauchy problem for general second order parabolic PDE of heat type*

$$\begin{cases} -\dfrac{\partial v(t,x)}{\partial t} + (\mathcal{A}_t v(t,\cdot))(t,x) + k(t,x)\,v(t,x) = g(t,x) & \text{for} \quad (t,x) \in (0,T) \times \mathbb{R}^n \\ v(0,x) = f(x) & \text{for} \quad x \in \mathbb{R}^n \end{cases}.$$

21 Twentyfirst Lecture

21.1 Uniqueness for Solutions to Martingale Problems

Theorem 21.1 (Stroock-Varadhan) *Consider the generator* $\mathcal{A}_t$ *in* (19.5). *Suppose that, for each* $f\in\mathbb{C}_0^\infty(\mathbb{R}^n)$ *and each pair* $s,T\geq 0$, *the Cauchy problem*

$$\frac{\partial g(t,x)}{\partial t}+(\mathcal{A}_{t+s}g)(x)=0 \qquad \text{for} \quad (t,x)\in(0,T)\times\mathbb{R}^n, \qquad g(T,\cdot)=f, \tag{21.1}$$

has a solution $g\in\mathbb{C}_B((0,T]\times\mathbb{R}^n)\cap\mathbb{C}^{1,2}((0,T)\times\mathbb{R}^n)$. *Given an* $\mathbb{R}^n$*-valued random variable* X_0 *and an* $s\geq 0$, *a solution* X *to the martingale problem for* $\mathcal{A}_{\cdot+s}$ *such that* $X(0)=_{\text{distribution}}X_0$, *has uniquely determined* `fidi's`.

The proof of Theorem 21.1 makes crucial use of *regular conditional probabilities.*

Lemma 21.2 (e.g., [22, Section 5.3.C] and [30, Theorem V.8.1])* *Let* $\{X(t)\}_{t\geq 0}$ *be a continuous* $\mathbb{R}^n$*-valued stochastic process. Given a* σ*-algebra* $\mathcal{G}\subseteq\sigma(X)$, *there exists a* $[0,1]$*-valued stochastic process* $\{Q(A)\}_{A\in\sigma(X)}$, *called a regular conditional probability, such that* $Q(\omega,\cdot)$ *is a probability measure on* $\sigma(X)$ *for each* $\omega\in\Omega$, *that satisfies*

$$\mathbf{P}\{A|\mathcal{G}\}=Q(A) \qquad \textit{for all} \quad A\in\sigma(X), \qquad \textit{with probability one} \quad (\textit{wp. 1}).$$

Proof of Theorem 21.1. First we show that, given a constant $x\in\mathbb{R}^n$, a solution $\{X(t)\}_{t\geq 0}$ to the martingale problem for $\mathcal{A}_{\cdot+s}$, with $X(0)=_{\text{distribution}}x$, has unique one-dimensional marginal distributions. By Exercise 112 below, this holds if

$$\mathbf{E}\{f(X(T))\} \quad \text{is uniquely determined for each choice of } T>0 \text{ and } f\in\mathbb{C}_0^\infty(\mathbb{R}^n).$$

To show this, let g solve the Cauchy problem (21.1). By Corollary 19.11,

$$g(t,X(t))-g(0,X(0))-\int_0^t\Big(\partial_1 g(\tau,X(\tau))+(\mathcal{A}_{\tau+s}g)(\tau,X(\tau))\Big)\,d\tau = g(t,X(t))-g(0,X(0))$$

is a continuous local martingale. Since this process is locally bounded, it is in fact a martingale, by Theorem 12.9. It follows that

$$\mathbf{E}\{f(X(T))\}-g(0,x)=\mathbf{E}\{g(T,X(T))\}-g(0,X(0))\}=\mathbf{E}\{g(0,X(0))-g(0,X(0))\}$$

is zero, so that $\mathbf{E}\{f(X(T))\}=g(0,x)$ only depends on X through x.

By Exercise 112, to show uniqueness for `fidi's`, it is enough to show that

$$\mathbf{E}\Big\{\prod_{i=1}^m f_i(X(t_i))\Big\} \quad \text{does not depend on the particular solution } \{X(t),\mathcal{F}_t\}_{t\geq 0} \tag{21.2}$$

to the martingale problem for $\mathcal{A}_{\cdot+r}$ with $X(0)=_{\text{distribution}}X_0$, for $0\leq t_1<\ldots<t_m$,

$f_1,\ldots,f_m\in\mathbb{C}_0^\infty(\mathbb{R}^n)$ and $m\in\mathbb{N}$. Assume that (21.2) holds for $m=k$, and consider the case $m=k+1$. Choose a regular conditional probability Q such that

$$\mathbf{P}\{B|X(t_1),\ldots,X(t_k)\}=Q(B)\qquad\text{for all}\quad B\in\sigma(X),\qquad\text{wp. 1.}$$

Writing $Z(t)\equiv X(t+t_k)$, $\{Z(t),\mathcal{F}_{t+t_k}\}_{t\geq 0}$ solves the martingale problem associated with $\mathcal{A}_{\cdot+t_k+r}$ wp. 1, because we have

$$\begin{aligned}&\int_\Lambda\Big(f(Z(t))-f(Z(s))-\int_s^t(\mathcal{A}_{\tau+t_k+r}f)(Z(\tau))\,d\tau\Big)\,dQ\\
&=\mathbf{E}\Big\{I_\Lambda\Big(f(X(t+t_k))-f(X(s+t_k))-\int_{s+t_k}^{t+t_k}(\mathcal{A}_{\tau+r}f)(X(\tau))\,d\tau\Big)\Big|X(t_1),\ldots,X(t_k)\Big\}\\
&=\mathbf{E}\Big\{I_\Lambda\mathbf{E}\Big\{f(X(t+t_k))-f(X(s+t_k))-\int_{s+t_k}^{t+t_k}(\mathcal{A}_{\tau+r}f)(X(\tau))\,d\tau\Big|\mathcal{F}_{s+t_k}\Big\}\\
&\qquad\Big|X(t_1),\ldots,X(t_k)\Big\}=0\qquad\text{for }\Lambda\in\mathcal{F}_{s+t_k}\text{ and }0\leq s<t,\qquad\text{wp. 1,}\end{aligned}\tag{21.3}$$

since X solves the martingale problem. Further $Z(0)=_{\text{distribution}}X(t_k)$, wp. 1, since

$$Q(\{Z(0)\in\cdot\})=\mathbf{P}\{X(t_k)\in\cdot|X(t_1),\ldots,X(t_k)\}=I_{(\cdot)}(X(t_k))\qquad\text{wp. 1.}\tag{21.4}$$

Hence it follows from the first part of the proof, that

$$\begin{aligned}\mathbf{E}\Big\{\prod_{i=1}^{k+1}f_i(X(t_i))\Big\}&=\mathbf{E}\Big\{\mathbf{E}\Big\{f_{k+1}(X(t_{k+1}))\Big|X(t_1),\ldots,X(t_k)\Big\}\prod_{i=1}^k f_i(X(t_i))\Big\}\\
&=\mathbf{E}\Big\{\Big(\int_\Omega f_{k+1}(Z(t_{k+1}-t_k))\,dQ\Big)\prod_{i=1}^k f_i(X(t_i))\Big\}\\
&=\mathbf{E}\Big\{g(0,X(t_k))\prod_{i=1}^k f_i(X(t_i))\Big\},\end{aligned}$$

where g solves the Cauchy problem (21.1), with $s=t_k+r$ and $T=t_{k+1}-t_k$. Here the right-hand side does not depend on the particular solution X to the martingale problem with $X(0)=_{\text{distribution}}X_0$, by the assumption that (21.2) holds for $m=k$. □

Remark 21.3 (e.g., [36, Theorem 3.2.1])* The Cauchy problem (21.1) has a solution with the properties required in Theorem 21.1, if for example, $\mathcal{A}_t$ is strongly elliptic, with μ and σ bounded and satisfying a Hölder condition, in each strip $[0,T]\times\mathbb{R}^n$. This means that, to each $T>0$, there exist constants $K_T,\alpha_T>0$ such that

$$\begin{cases}\|\mu(s,x)-\mu(t,y)\|\leq K_T\|(s,x)-(t,y)\|^{\alpha_T} & \text{for}\quad (s,x),(t,y)\in[0,T]\times\mathbb{R}^n\\
\|\sigma(s,x)-\sigma(t,y)\|\leq K_T\|(s,x)-(t,y)\|^{\alpha_T} & \text{for}\quad (s,x),(t,y)\in[0,T]\times\mathbb{R}^n\\
\sum_{i=1}^n\sum_{i=1}^n(\sigma\sigma^T)_{i,j}(s,x)\,\xi_i\xi_j\geq K_T^{-1}\|\xi\|^2 & \text{for}\quad \xi\in\mathbb{R}^n\text{ and }(s,x)\in[0,T]\times\mathbb{R}^n\end{cases}.\ \#$$

EXERCISE 112 Prove that, for two random variables Y and Z in $\mathbb{R}^n$, we have

$$Y=_{\text{distribution}}Z\quad\Leftrightarrow\quad\mathbf{E}\Big\{\prod_{i=1}^n f_i(Y_i)\Big\}=\mathbf{E}\Big\{\prod_{i=1}^n f_i(Z_i)\Big\}\quad\text{for all }f_1,\ldots,f_n\in\mathbb{C}_0^\infty(\mathbb{R}).$$

Corollary 21.4 *Consider the generator $\mathcal{A}_t$ in* (19.5)*, where μ and σ are continuous and bounded. Suppose that, for each $f \in \mathbb{C}_0^\infty(\mathbb{R}^n)$ and each pair $s, T \geq 0$, the Cauchy problem* (21.1) *has a solution with the properties specified in Theorem 21.1. The martingale problem for $\mathcal{A}_t$ is well-posed.*

Proof. Theorem 20.8 together with Theorem 21.1. □

By the general discussion of martingale problems in Section 20.1, Theorem 21.1 and Corollary 21.4 (together with Remark 21.3), have obvious corollaries concerning uniqueness (and existence) of solutions to multidimensional time homogeneous SDE.

21.2 Feynman-Kac Formula for parabolic Cauchy Problem

The results in this section, and the next, are important in mathematics, and bring analysis and probability together. We have collected them from [22].

Let the functions $\mu:[0,\infty)\times\mathbb{R}^n \to \mathbb{R}^n$ and $\sigma:[0,\infty)\times\mathbb{R}^n \to \mathbb{R}_{n|n}$ have measurable locally bounded components. Consider the n-dimensional diffusion type SDE

$$dX(s) = \mu(s, X(s))\,ds + \sigma(s, X(s))\,dB(s) \qquad \text{for} \quad s \geq t, \qquad X(t) = x, \tag{21.5}$$

with the generator $\mathcal{A}_t$ given by (19.5). Consider the Cauchy problem

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} + (\mathcal{A}_t u)(t,x) + k(t,x)\,u(t,x) = g(t,x) & \text{for} \quad (t,x) \in (0,T)\times\mathbb{R}^n \\ u(T,x) = f(x) & \text{for} \quad x \in \mathbb{R}^n \end{cases}, \tag{21.6}$$

where $k, g \in \underline{\mathbb{C}([0,T]\times\mathbb{R}^n)} \equiv \{(\tilde{g} : [0,T]\times\mathbb{R}^n \to \mathbb{R}) : \tilde{g} \text{ is continuous}\}$ and $f \in \underline{\mathbb{C}(\mathbb{R}^n)}$ (similarly defined).

Theorem 21.5 (Feynman-Kac Formula) *Let the SDE* (21.5) *have a weak solution $\{X^{t,x}(s)\}_{s\geq t}$ that is unique, for each $x \in \mathbb{R}^n$ and $t > 0$* (cf. Corollary 21.4)*. Let the Cauchy problem* (21.6) *have a solution $u \in \mathbb{C}((0,T]\times\mathbb{R}^n) \cap \mathbb{C}^{1,2}((0,T)\times\mathbb{R}^n)$. We have*

$$u(t,x) = \mathbf{E}\bigg\{ f(X^{t,x}(T))\, \exp\bigg\{\int_t^T k(r, X^{t,x}(r))\,dr\bigg\} - \int_t^T g(s, X^{t,x}(s))\, \exp\bigg\{\int_t^s k(r, X^{t,x}(r))\,dr\bigg\} ds \bigg\} \quad \textit{for} \quad (t,x) \in (0,T]\times\mathbb{R}^n,$$

provided that the mean is well-defined (and in particular such solutions u are unique).

Proof. By application of *Itô's formula*, together with (21.5) and (21.6), we obtain

$$d\bigg(u(s, X(s))\, \exp\bigg\{\int_t^s k(r, X(r))\,dr\bigg\}\bigg)$$

$$= \Big(\partial_1 u\, ds + \sum_{i=1}^n \partial_{x_i} u\, dX_i(s) + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \partial_{x_i}\partial_{x_j} u\, d[X_i, X_j](s) + u\, k\, ds\Big) \exp$$

$$= \Big(\partial_1 u\, ds + \sum_{i=1}^n \partial_{x_i} u \Big(\mu_i\, ds + \sum_{j=1}^n \sigma_{i,j}\, dB_j(s)\Big) + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \partial_{x_i}\partial_{x_j} u\, (\sigma\sigma^T)_{i,j}\, ds + u\, k\, ds\Big) \exp$$

$$= \Big(\sum_{i=1}^n \sum_{j=1}^n \partial_{x_i} u\, \sigma_{i,j}\, dB_j(s) + g\, ds\Big) \exp,$$

so that the integral over $[t,T]$ of the left-hand side

$$u(T, X(T)) \exp\Big\{\int_t^T k(r, X(r))\, dr\Big\} - u(t, X(t)) \exp\Big\{\int_t^t k(r, X(r))\, dr\Big\}$$
$$= f(X(T)) \exp\Big\{\int_t^T k(r, X(r))\, dr\Big\} - u(t,x)$$

is equal to the integral over $[t,T]$ of the right-hand side

$$\int_t^T \sum_{i=1}^n \sum_{j=1}^n \partial_{x_i} u(s, X(s)) \exp\Big\{\int_t^s k(r, X(r))\, dr\Big\}\sigma_{i,j}(s, X(s))\, dB_j(s)$$
$$+ \int_t^T g(s, X(s)) \exp\Big\{\int_t^s k(r, X(r))\, dr\Big\}\, ds.$$

Rearranging, this shows that

$$u(t,x) = f(X(T)) \exp\Big\{\int_t^T k(r, X(r))\, dr\Big\}$$
$$- \int_t^T \sum_{i=1}^n \sum_{j=1}^n \partial_{x_i} u(s, X(s)) \exp\Big\{\int_t^s k(r, X(r))\, dr\Big\}\sigma_{i,j}(s, X(s))\, dB_j(s)$$
$$- \int_t^T g(s, X(s)) \exp\Big\{\int_t^s k(r, X(r))\, dr\Big\}\, ds.$$

Since the left-hand side is non-random, and the difference between the first and third term on the right-hand side have finite means, the second term on the right-hand side must also have finite mean. That mean is zero, by Exercise 113 below. □

Recall from Section 20.3, that the Cauchy problem (21.6) can be transform to a conventional heat equation for $v(t,x) = u(T-t, x)$, with initial value $v(0,x) = f(x)$.

EXERCISE 113 Show that Itô integrals wrt. BM are symmetric random variables.

*Remark 21.6 See [22, Section 5.7] for additional conditions on the functions μ, σ, g and f, that ensure that the expectation in Theorem 21.5 is finite. #

21.3 Solutions to Elliptic Dirichlet Problems

Let $\mu : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}_{n|n}$ be measurable functions, with σ locally bounded. Consider the n-dimensional time homogeneous diffusion type SDE

$$dX(t) = \mu(X(t))\, dt + \sigma(X(t))\, dB(t) \qquad \text{for} \quad s \geq 0, \qquad X(0) = x, \tag{21.7}$$

with time homogeneous generator $\mathcal{A}$ given by (20.9). Consider the Dirichlet problem

$$\begin{cases} (\mathcal{A}u)(x) - k(x)\,u(x) = g(x) & \text{for} \quad x \in D \\ u(x) = f(x) & \text{for} \quad x \in \text{boundary}(D) \end{cases}, \tag{21.8}$$

where $k : \text{closure}(D) \to [0, \infty)$ and $g, f \in \mathbb{C}(\text{closure}(D))$ are given functions, and $D \subseteq \mathbb{R}^n$ is an open and bounded set.

Theorem 21.7 *Let the SDE* (21.7) *have a weak solution* $\{X^x(s)\}_{s \geq 0}$ *that is uni-que, for each* $x \in \mathbb{R}^n$. *Let the Dirichlet problem* (21.8) *have a solution* $u \in \mathbb{C}(\text{closure}(D)) \cap \mathbb{C}^2(D)$. *We have*

$$u(x) = \mathbf{E}\bigg\{ f(X^x(\tau_D)) \exp\Big\{ -\int_0^{\tau_D} k(r, X^x(r))\, dr \Big\}$$
$$- \int_0^{\tau_D} g(s, X^x(s)) \exp\Big\{ -\int_0^s k(r, X^x(r))\, dr \Big\} ds \bigg\} \quad \textit{for} \quad x \in \text{closure}(D),$$

provided that $\tau_D \equiv \inf\{t \geq 0 : X(t) \notin D\}$ *has finite mean* $\mathbf{E}\{\tau_D\} < \infty$.

Proof. By application of *Itô's formula*, together with (21.7) and (21.8), we obtain

$$d\Big(u(X(t)) \exp\Big\{ -\int_0^t k(X(r))\, dr \Big\} \Big)$$
$$= \Big(\sum_{i=1}^n \partial_i u\, dX_i(t) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \partial_i \partial_j u\, d[X_i, X_j](t) - u\, k\, dt \Big) \exp$$
$$= \Big(\sum_{i=1}^n \partial_i u \Big(\mu_i\, dt + \sum_{j=1}^n \sigma_{i,j}\, dB_j(t) \Big) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \partial_i \partial_j u\, (\sigma \sigma^T)_{i,j}\, dt - u\, k\, dt \Big) \exp$$
$$= \Big(\sum_{i=1}^n \sum_{j=1}^n \partial_i u\, \sigma_{i,j}\, dB_j(t) + g\, dt \Big) \exp,$$

so that the integral over $[0, t]$ of the left-hand side

$$u(X(t)) \exp\Big\{ -\int_0^t k(X(r))\, dr \Big\} - u(X(0)) \exp\Big\{ -\int_0^0 k(X(r))\, dr \Big\}$$
$$= u(X(t)) \exp\Big\{ -\int_0^t k(X(r))\, dr \Big\} - u(x)$$

is equal to the integral over $[0, t]$ of the right-hand side

$$\int_0^t \sum_{i=1}^n \sum_{j=1}^n \partial_i u(X(s)) \exp\Big\{ -\int_0^s k(X(r))\, dr \Big\} \sigma_{i,j}(X(s))\, dB_j(s)$$
$$+ \int_0^t g(X(s)) \exp\Big\{ -\int_0^s k(X(r))\, dr \Big\}\, ds.$$

Rearranging, taking $t = \tau_D$, and using that $u(X(\tau_D)) = f(X(\tau_D))$, this gives

$$u(x) = f(X(\tau_D)) \exp\Big\{ -\int_0^{\tau_D} k(X(r))\, dr \Big\}$$

$$- \int_0^{\tau_D} \sum_{i=1}^n \sum_{j=1}^n \partial_i u(X(s)) \exp\Big\{-\int_0^s k(X(r))\,dr\Big\}\sigma_{i,j}(X(s))\,dB_j(s)$$
$$- \int_0^{\tau_D} g(X(s)) \exp\Big\{-\int_0^s k(X(r))\,dr\Big\}\,ds. \tag{21.9}$$

The processes $\{\partial_i u(X(t))\}_{t\in[0,\tau_D]}$ and $\{\sigma_{i,j}(X(t))\}_{t\in[0,\tau_D]}$ are bounded, because $\partial_i u$ and $\sigma_{i,j}$ are locally bounded, and $X(t) \in \mathrm{closure}(D)$ (which is bounded) for $t \in [0,\tau_D]$ (since X is continuous). From this it follows that

$$\sum_{j=1}^n \partial_i u(X(t)) \exp\Big\{-\int_0^t k(X(r))\,dr\Big\}\,\sigma_{i,j}(X(t))\,I_{\{t\le\tau_D\}} \in E_\infty \qquad \text{for} \quad i=1,\ldots,n,$$

since (recalling that the function k is positive)

$$\mathbf{E}\bigg\{\int_0^\infty\Big(\sum_{j=1}^n \partial_i u(X(s)) \exp\Big\{-\int_0^s k(X(r))\,dr\Big\}\,\sigma_{i,j}(X(s))\,I_{\{s\le\tau_D\}}\Big)^2 ds\bigg\}$$
$$\le \mathbf{E}\{\tau_D\} \sup_{x\in\mathrm{closure}(D)} \Big(\sum_{j=1}^n \partial_i u(x)\,\sigma_{i,j}(x)\Big)^2 < \infty.$$

Hence the Itô integral the second term on the right-hand of (21.9) has finite mean, which must be zero. It follows that the difference between the first and third term on the right hand side of (21.9) has finite mean, which must be $u(x)$. □

*Remark 21.8 By [22, Lemma 5.7.4], we have $\mathbf{E}\{\tau_D\} < \infty$ if $\mathcal{A}$ is strongly elliptic in closure(D), so that there exists a constant $K>0$ such that

$$\sum_{i=1}^n \sum_{i=1}^n (\sigma\sigma^T)_{i,j}(x)\,\xi_i\xi_j \ge K\,\|\xi\|^2 \qquad \text{for} \quad \xi\in\mathbb{R}^n \quad \text{and} \quad x\in\mathrm{closure}(D). \quad \#$$

22 Twentysecond Lecture

22.1 Markov Properties of SDE

In the following, we encounter (strong) Markov processes with values in $\mathbb{R}^n$. There is nothing strange with this, or new for that matter, and all notation and results from Lectures 3-5 carry over to the multidimensional, with only obvious modifications.

The usefulness of the martingale formulation is illustrated by the next result.

Theorem 22.1 (Markov Property of SDE) *Let the martingale problem for the generator $\mathcal{A}_t$ in (19.5) be well-posed, with σ locally bounded. A solution to the martingale problem for $\mathcal{A}_\cdot$ is a Markov process wrt. itself, with transition probabilities*

$$\mathbf{P}\{X(t+s)\in\cdot\,|X(s)=x\} = \mathbf{P}\{\hat{X}^{s,x}(t)\in\cdot\} = \mathbf{P}\{X^{s,x}(t+s)\in\cdot\},$$

for $(t,x,s)\in(0,\infty)\times\mathbb{R}^n\times[0,\infty)$, where $\{\hat{X}^{s,x}(t)\}_{t\geq 0}$ solves the martigale problem for $\mathcal{A}_{\cdot+s}$, with $\hat{X}^{s,x}(0)=_{\text{distribution}}x$, and $\{X^{s,x}(t)\}_{t\geq s}$ is a weak solution to SDE

$$dX(t)=\mu(t,X(t))\,dt+\sigma(t,X(t))\,dB(t) \qquad \text{for} \quad t\geq s, \qquad X(s)=x.$$

In the time homogeneous case when $\mathcal{A}_t=\mathcal{A}$ for $t\geq 0$ [cf. (20.9)], *X is a time homogeneous Markov process.*

For the transition probabilities, we use *regular pointwise conditional probabilities*:

Lemma 22.2 (e.g., [22, Section 5.3.C])* *Let $\{X(t)\}_{t\geq 0}$ be a continuous $\mathbb{R}^n$-valued stochastic process. Given an $r\geq 0$, there exists a function $Q:\mathbb{R}^n\times\sigma(X)\to[0,1]$, called a regular pointwise conditional probability, with the following properties*

(1) $Q(x,\cdot):\sigma(X)\to[0,1]$ *is a probability measure on $\sigma(X)$ for each $x\in\mathbb{R}^n$;*

(2) $Q(x,\{X(r)=x\})=1$*;*

(3) $\mathbf{P}\{A|\,X(r)=x\}=Q(x,A)$ *for all* $A\in\sigma(X)$.

Recall that, in general, the pointwise conditional probability $\mathbf{P}\{A|\,X(s)=x\}$ is not uniquely defined as a function of x, so the statement of the lemma means that $Q(x,A)$ is one such pointwise conditional probability, of perhaps many possible.

Proof of Theorem 22.1. Let $\{X(t),\mathcal{F}_t\}_{t\geq 0}$ solve the martingale problem for $\mathcal{A}_\cdot$. We may assume that $\mathcal{F}_t=\sigma(X(s):0\leq s\leq t)$ (see Exercise 115 below). Pick an $\hat{r}\geq 0$. Choose regular conditional probabilities Q_1 and Q_2 (cf. Lemma 21.2), such that

$$\mathbf{P}\{A|\,X(\hat{r})\}=Q_1(A) \quad \text{and} \quad \mathbf{P}\{A|\,\mathcal{F}_{\hat{r}}\}=Q_2(A) \quad \text{for all} \quad A\in\sigma(X), \quad \text{wp. 1.}$$

Writing $Z(t)\equiv X(t+\hat{r})$, we get the Markov property if $\{Z(t),\mathcal{F}_{t+\hat{r}}\}_{t\geq 0}$ solves the martingale problem for $\mathcal{A}_{\cdot+\hat{r}}$, with $Z(0)=_{\text{distribution}}X(\hat{r})$, under the probabilities Q_1 and Q_2, wp. 1, since well-posedness of the martingale problem then gives

$$Q_1(\{X(t+\hat{r})\in\cdot\}) = Q_1(\{Z(t)\in\cdot\}) = Q_2(\{Z(t)\in\cdot\}) = Q_2(\{X(t+\hat{r})\in\cdot\}) \quad \text{wp. 1.}$$

This is the same thing as the Markov property, by the choice of Q_1 and Q_2.

We get the martingale properties required, with $Z(0)=_{\text{distribution}}X(\hat{r})$, under both Q_1 and Q_2, wp. 1, by the arguments (21.3) and (21.4), with $t_k=\hat{r}$ and $r=0$.

Pick a regular pointwise conditional probability $Q(\cdot,\cdot)$ such that (cf. Lemma 22.2)

$$\mathbf{P}\{A|X(\hat{r})=x\} = Q(x,A) \qquad \text{for} \qquad A\in\sigma(X).$$

By Theorem 20.6, writing $Z(t)\equiv X(t+\hat{r})$, it is enough to prove that $\{Z(t),\mathcal{F}_{t+\hat{r}}\}_{t\geq 0}$ solves the martingale problem for $\mathcal{A}_{\cdot+\hat{r}}$, with $Z(0)=_{\text{distribution}}x$, under the probability $Q(x,\cdot)$, for almost all $(dF_{X(\hat{r})})$ $x\in\mathbb{R}^n$. Because then we have

$$\mathbf{P}\{\hat{X}^{\hat{r},x}(t)\in\cdot\} = Q(x,\{Z(t)\in\cdot\}) = \mathbf{P}\{X(t+\hat{r})\in\cdot|X(\hat{r})=x\} = P(\cdot,t,x,\hat{r}),$$

by well-posedness of the martingale problem, for almost all $(dF_{X(\hat{r})})$ $x\in\mathbb{R}^n$. Values of $P(\cdot,t,x,\hat{r})$ for x in a null-set $(dF_{X(\hat{r})})$ are unessential, since corresponding values of $P(\cdot,t,x,X(\hat{r}))=\mathbf{P}\{X(t+\hat{r})\in\cdot\,|X(\hat{r})\}$ only occur for ω in a null-set $(d\mathbf{P})$, and are thus "swallowed" by the natural ambiguity of conditional probabilities.

Now recall from above that, since Z is a martingale under Q_1, we have

$$\mathbf{E}\Big\{I_\Lambda\Big(f(Z(t))-f(Z(s))-\int_s^t(\mathcal{A}_{\tau+\hat{r}}f)(Z(\tau))\,d\tau\Big)\Big|X(\hat{r})\Big\} = 0 \qquad \text{a.s.} \tag{22.1}$$

for $f\in\mathbb{C}_0^2(\mathbb{R}^n)$, $0\leq s<t$ and $\Lambda\in\mathcal{F}_{s+\hat{r}}$. This makes necessary that

$$\begin{aligned}\mathbf{E}\Big\{I_\Lambda\Big(f(Z(t))-f(Z(s))-\int_s^t(\mathcal{A}_{\tau+\hat{r}}f)(Z(\tau))\,d\tau\Big)\Big|X(\hat{r})=x\Big\}\\ = \int_\Lambda\Big(f(Z(t))-f(Z(s))-\int_s^t(\mathcal{A}_{\tau+\hat{r}}f)(Z(\tau))\,d\tau\Big)\,dQ(x,\cdot) = 0\end{aligned} \tag{22.2}$$

for almost all $(dF_{X(\hat{r})})$ $x\in\mathbb{R}^n$ [to not contradict (22.1)]. By properties of $\mathbb{C}_0^2(\mathbb{R}^n)$ (separability), continuity of Z and the integral, and properties of $\mathcal{F}_{s+\hat{r}}$ ("countably determined", by continuity of X), together with *Dominated Convergence*, we have (22.2) simultaneously for all $f\in\mathbb{C}_0^2(\mathbb{R}^n)$, $0\leq s<t$ and $\Lambda\in\mathcal{F}_{s+\hat{r}}$, for almost all $(dF_{X(\hat{r})})$ $x\in\mathbb{R}^n$ (see Exercise 114 below). This is the required martingale property.

The fact that $Z(0)=_{\text{distribution}}x$, is the same thing as

$$Q(x,\{Z(0)=x\}) = Q(x,\{X(\hat{r})=x\} = 1,$$

which is property (2) in Lemma 22.2.

The statement about time homogeneity follows by inspection of the above proof. □

*EXERCISE 114 Show that if (22.2) holds for each choice of $f \in \mathbb{C}_0^2(\mathbb{R}^n)$, $0 \le s < t$ and $\Lambda \in \mathcal{F}_{s+r}$, for almost all $(dF_{X(r)})$ $x \in \mathbb{R}^n$, then (22.2) holds for all $f \in \mathbb{C}_0^2(\mathbb{R}^n)$, $0 \le s < t$ and $\Lambda \in \mathcal{F}_{s+r}$, for almost all $(dF_{X(r)})$ $x \in \mathbb{R}^n$.

EXERCISE 115 Why is the extension of (22.2) in Exercise 114 vital for proof of Theorem 22.2. Why can we assume that $\mathcal{F}_t = \sigma(X(s) : 0 \le s \le t)$ in that proof?

*Remark 22.3 (STRONG MARKOV PROPERTY OF SDE) The strong Markov property holds under the hypothesis of Theorem 22.1 (e.g., [22, Theorem 5.4.20] in the time homogeneous case). The proof requires quite advanced *Optional Stopping* techniques. We do not do this, since we get strong Markov "for free" in Corollary 22.6 below, from Theorem 4.21. But see Exercise 118 below. #

22.2 How everything hang together

Definition 22.4 *Consider the generator* $\mathcal{A}_t$ *in* (19.5)*, and pick a* $T > 0$*. A* *fundamental solution* *to the parabolic PDO*

$$\Big(\frac{\partial}{\partial t} + \mathcal{A}_t\Big) u(t,x) = \frac{\partial u(t,x)}{\partial t} + \sum_{i=1}^n \sum_{i=1}^n (\sigma\sigma^T)_{i,j}(t,x)\, \frac{\partial^2 u(t,x)}{\partial x_i \partial x_j} + \sum_{i=1}^n \mu_i(t,x)\, \frac{\partial u(t,x)}{\partial x_i} \qquad (22.3)$$

in the strip $(0,T] \times \mathbb{R}^n$*, is a family of measurable functions* $\{(\hat{p}(\cdot, \tau, x, t) : \mathbb{R}^n \to [0, \infty)) : (x,t) \in \mathbb{R}^n \times (0,\tau),\ \tau \in (0,T]\}$*, such that, given* $f \in \mathbb{C}_0(\mathbb{R}^n)$ *and* $\tau \in (0,T]$*,*

$$u(t,x) = \int_{\mathbb{R}^n} \hat{p}(y,\tau,x,t)\, f(y)\, dy \qquad \textit{for} \quad (t,x) \in (0,\tau) \times \mathbb{R}^n, \qquad u(\tau,\cdot) = f,$$

is of class $\mathbb{C}_B((0,\tau] \times \mathbb{R}^n) \cap \mathbb{C}^{1,2}((0,\tau) \times \mathbb{R}^n)$*, and solves the* *Kolmogorov Backward equation*

$$\frac{\partial u(t,x)}{\partial t} + (\mathcal{A}_t u)(t,x) = 0 \qquad \textit{for} \quad (t,x) \in (0,\tau) \times \mathbb{R}^n, \qquad u(\tau,\,\cdot\,) = f. \qquad (22.4)$$

Remark 22.5 The PDO (22.3) has a fundamental solution, if for example, the requirements listed in Remark 21.3 are satisfied for the one specific T chosen in Definition 22.4, rather than for each $T > 0$ (e.g., [36, Theorem 3.2.1])*. The strong ellipticity imposed in Remark 21.3 can be dispensed with if $(\sigma\sigma^T)(t,x)$ and $\mu(t,x)$ have two bounded continuous x-derivatives, and $\hat{p}$ is replaced with a measure $\hat{P}(\cdot, \tau, x, t)$ such that $u(t,x) = \int f(\cdot)\, d\hat{P}(\cdot, \tau, x, t)$ satisfies (22.4) (e.g., [36, Section 3.2])*. #

The following result is one of the most important in stochastic processes. It shows why the *Feynman-Kac formula* is of fundamental importance, not only as a result on parabolic PDE, but for the whole business of diffusion type SDE.

Corollary 22.6 *Consider the generator $\mathcal{A}_t$ in (19.5), where σ is locally bounded, and assume that the martingale problem associated with $\mathcal{A}_t$ is well-posed. Suppose that, for each $T \geq 0$, the PDO (22.3) has a fundamental solution $\hat{p}(\cdot,\cdot,\cdot,\cdot)$ in the strip $(0,T]\times\mathbb{R}^n$. For each choice of an $\mathbb{R}^n$-valued random variable X_0, the multidimensional SDE*

$$dX(t) = \mu(t,X(t))\,dt + \sigma(t,X(t))\,dB(t) \qquad \text{for} \quad t\geq 0, \qquad X(0) = X_0, \tag{22.5}$$

has weak solution that is unique and a strong Markov process, with transition density

$$p(y,t,x,s) = \frac{d}{dy}\mathbf{P}\{X(t+s)\leq y\,|\,X(s)=x\} = \hat{p}(y,t+s,x,s). \tag{22.6}$$

Proof. By assumption, given $s,T>0$, the function

$$g(t,x) = \int_{\mathbb{R}^n} \hat{p}(y,T+s,x,t+s)\,f(y)\,dy \qquad \text{for} \quad (t,x)\in(0,T)\times\mathbb{R}^n, \qquad g(T,\cdot) = f,$$

is of class $\mathbb{C}((0,T]\times\mathbb{R}^n)\cap\mathbb{C}^{1,2}((0,T)\times\mathbb{R}^n)$, and solves the Cauchy problem (21.1)

$$\frac{\partial g(t,x)}{\partial t} + (\mathcal{A}_{t+s}g)(x) = 0 \qquad \text{for} \quad (t,x)\in(0,T)\times\mathbb{R}^n, \qquad g(T,\cdot)=f.$$

From Theorem 22.1 we have that a solution to the martingale problem for $\mathcal{A}_\cdot$ is a Markov process. By Corollary 20.4, this also goes for a solution to the SDE (22.5). By Theorem 22.1, the transition probabilities are

$$\mathbf{P}\{X(\tau)\in\cdot\,|\,X(t)=x\} = \mathbf{P}\{X^{t,x}(\tau)\in\cdot\} \qquad \text{for} \quad \tau>t, \tag{22.7}$$

where $\{X^{t,x}(\tau)\}_{\tau\geq t}$ is a weak solution to SDE

$$dX(\tau) = \mu(\tau,X(\tau))\,d\tau + \sigma(\tau,X(\tau))\,dB(\tau) \qquad \text{for} \quad \tau\geq t, \qquad X(\tau) = x. \tag{22.8}$$

Applying the *Feynman-Kac formula* to the Cauchy problem (22.4), we get

$$\mathbf{E}\{f(X^{t,x}(\tau))\} = u(t,x) = \int_{\mathbb{R}^n}\hat{p}(y,\tau,x,t)\,f(y)\,dy,$$

so that the random variable $X^{t,x}(\tau)$ has probability density function $\hat{p}(\cdot,\tau,x,t)$. This together with (22.7) gives (22.6).

Notice that the requirement in the *Feynman-Kac formula*, that the SDE (22.8) has a unique solution, is satisfied by the assumed well-posedness of the martingale problem associated with the generator $\mathcal{A}_t$, together with Corollary 20.6.

We get the strong Markov property from Theorem 4.21, since X satisfies property (B) of a Feller process, by inspection of the requirements on $\hat{p}$ in Definition 22.4. □

From the fact that the function u in Definition 22.4 satisfies (22.4), it is tempting to move the derivatives inside the integral, to get the *Kolmogorov Backward equation*

$$\frac{\partial \hat{p}(y,\tau,x,t)}{\partial t} + \mathcal{A}_t \hat{p}(y,\tau,x,t) = \frac{\partial p(y,\tau-t,x,t)}{\partial t} + \mathcal{A}_t p(y,\tau-t,x,t) = 0 \qquad (22.9)$$

for $(t,x) \in (0,\tau)\times\mathbb{R}^n$. Under the conditions in Remark 22.5, it is turns out that this holds (e.g., [22, p. 368])*. Notice that such a solution, under the hypothesis of Corollary 22.6, is unique (up to equivalence of probability densities). ["Backward" refers to that the PDE is in the variables (x,t), that is the past relative to (y,τ).]

22.3 Multidimensional Time Homogeneous SDE

In the time homogeneous case we use the next simplification of Definition 23.4:

Definition 22.7 *Consider the time homogeneous generator $\mathcal{A}$ in (20.9). A fundamental measure to the parabolic PDO*

$$\Big(\frac{\partial}{\partial t} - \mathcal{A}\Big) v(t,x) = \frac{\partial v(t,x)}{\partial t} - \sum_{i=1}^n \sum_{i=1}^n (\sigma\sigma^T)_{i,j}(x) \frac{\partial^2 v(t,x)}{\partial x_i \partial x_j} - \sum_{i=1}^n \mu_i(x) \frac{\partial v(t,x)}{\partial x_i}, \qquad (22.10)$$

is a family of Borel measures $\{\hat{P}(\cdot,t,x)\}_{(t,x)\in[0,\infty)\times\mathbb{R}^n}$, such that given $f\in\mathbb{C}_0(\mathbb{R}^n)$,

$$v(t,x) = \int_{\mathbb{R}^n} f(y)\, d\hat{P}(y,t,x) \qquad \text{for} \quad (t,x)\in[0,\infty)\times\mathbb{R}^n, \qquad v(0,\cdot) = f,$$

is of class $\mathbb{C}_B([0,\infty)\times\mathbb{R}^n)$, with bounded continuous derivatives $\frac{\partial v(t,x)}{\partial t}$ and $\frac{\partial^2 v(t,x)}{\partial x_i \partial x_j}$ on $[0,\infty)\times\mathbb{R}^n$, that solve the time homogeneous Kolmogorov Backward equation

$$\frac{\partial v(t,x)}{\partial t} - (\mathcal{A}v)(x) = 0 \qquad \text{for} \quad (t,x)\in(0,\infty)\times\mathbb{R}^n, \qquad v(0,\ \cdot\ ) = f. \qquad (22.11)$$

Corollary 22.8 *Consider the time homogeneous generator $\mathcal{A}$ in (20.9), where σ is locally bounded, and assume that the martingale problem associated with $\mathcal{A}$ is well-posed. Suppose that the PDO (22.11) has a fundamental measure $\hat{P}(\cdot,\cdot,\cdot)$. For each choice of an $\mathbb{R}^n$-valued random variable X_0, the multidimensional SDE*

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t) \qquad \text{for} \quad t\geq 0, \qquad X(0) = X_0,$$

has weak solution that is unique and a time homogeneous strong Markov process, with transition probability

$$P(\cdot,t,x,s) = \mathbf{P}\{X(t+s)\in\cdot\,|\,X(s)=x\} = \hat{P}(\cdot,t,x).$$

For $\mathcal{A}$ strongly elliptic, with μ and σ bounded satisfying a Hölder condition [see the text after (22.9)], the PDO (22.10) has a fundamental measure with a density function $\hat{p}(\cdot,t,x)$, so that $P(\cdot,t,x,s)$ has a density $p(\cdot,t,x,s) = \hat{p}(\cdot,t,x)$, such that

$$\frac{\partial \hat{p}(y,t,x)}{\partial t} - \mathcal{A}\hat{p}(y,t,x) = \frac{\partial p(y,t,x,0)}{\partial t} - \mathcal{A}p(y,t,x,0) = 0. \qquad (22.12)$$

In Theorem 26.8 below, we give a formula for the fundamental solution (transition density) of a general one-dimensional time homogeneous diffusion process.

Example 22.9 BM B is a one-dimensional time homogeneous diffusion ($dB = dB$) with generator $(\mathcal{A}f)(x) = \frac{1}{2}f''(x)$. By Corollary 22.8, B is a time homogeneous strong Markov process with transition probability (fundamental measure)

$$P(\cdot\,,t,x,s) = P(\cdot\,,t,x,0) = \hat{P}(\cdot\,,t,x),$$

such that the *time homogeneous Kolmogorov Backward equation*

$$\Big(\frac{\partial}{\partial t} - \mathcal{A}\Big)\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot\,,t,x) = \Big(\frac{\partial}{\partial t} - \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big)\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot\,,t,x) = 0$$

holds for $(t,x)\in(0,\infty)\times\mathbb{R}$ and $f\in\mathbb{C}_0(\mathbb{R})$, with boundary condition

$$\lim_{t\downarrow 0}\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot\,,t,x)\,dy = f(x).$$

Rather than attempting to solve this for $\hat{P}$, we recall that (cf. Corollary 3.14)

$$\begin{aligned} P(\cdot\,,t,x,0) = \mathbf{P}\{B(t+s)-B(s)+B(s)\in\cdot\,|\,B(s)=x\} &= \mathbf{P}\{B(t+s)-B(s)\in\cdot-x\} \\ &= \mathbf{P}\{\mathrm{N}(0,t)\in\cdot-x\} \end{aligned}$$

for $(y,t,x,s)\in\mathbb{R}\times(0,\infty)\times\mathbb{R}\times[0,\infty)$. This means that

$$\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot\,,t,x) = \int_{\mathbb{R}} f(\cdot)\,dP(\cdot\,,t,x,s) = \int_{\mathbb{R}} f(y)\,\frac{1}{\sqrt{2\pi t}}\,\mathrm{e}^{-\frac{(y-x)^2}{2t}}\,dy.$$

In particular, as also follows from above general treatment, $\hat{P}(\cdot\,,t,x)$ $[P(\cdot\,,t,x,s)]$ has a density function $\hat{p}(\cdot\,,t,x)$ $[p(\cdot\,,t,x,s)]$ that satisfies (22.12), given by

$$\hat{p}(y,t,x) = p(y,t,x,s) = \frac{1}{\sqrt{2\pi t}}\,\mathrm{e}^{-\frac{(y-x)^2}{2t}} \qquad \textit{for} \quad (t,x)\in(0,\infty)\times\mathbb{R}. \quad \#$$

23 Twentythird Lecture

Smartingales is one huge subject, that is of fundamental importance in many branches of probability. We have seen how it features in stochastic calculus. It is an essential ingredient in many proofs of convergence results. The many powerful inequalities known for smartingales play a prominent role in many a proof.

The "state of the art" literature on smartingales, as for example [20] and [33] (the latter has recently appeared in a third edition), is not really accesible for an average beginner. Instead, one has to prepare first with introductory treatments of the subject, as for example, the few smartingale results collected in these notes (while here anyway), followed by for example, [21, Chapter 6].

Martingales were first used by Lévy 1934, and named by Ville 1939. The modern continuous theory begun with Doob 1953 [11]. See [22, p. 46] on more history.

This lecture lists a few results for smartingales, which are cornerstones in the theory together with the *Doob-Meyer decomposition*. We do not really use the convergence results in Section 23.2, and thus do not prove them. The *Optional Sampling theorem* in Section 23.3, the characterization of what local martingales are martingales in Section 23.4, and the *Burkholder-Davis-Gundy inequalities* in Section 23.5, are all crucially important for us, and proven in detail (since these are not too overhelming).

Warning 23.1 Continuous time smartingale results about "the continuum" (rather than the smartingale at a finite number of times), do in general require some regularity, such as right-continuity. (The proof of *Doob-Kolmogorov inequality* exemplifies how right-continuity comes into play.) For this reason, in many books (e.g., [21, p. 112], [25, p. 170] and [33, p. 62]), it is at some point stated that "right-continuity is assumed in the sequel", or similar, and subsequent results do not repeatedly state this basic assumption, so beware that important results for smartingales in continuous time do in general require right-continuity or even cádlág. #

23.1 What is a Stopping Time?

There are variations in the literature concerning the terminology in connection with stopping times, that it is important to be aware of.

A random time is a $[0,\infty]$-valued random variable. In stochastic calculus (albeit not in these notes), there is need for (at least) three different classes of random times.

Let $\mathbb{F}=\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration. A random time τ is

- a stopping time if $\{\tau\leq t\}\in\mathcal{F}_t$ for each $t\geq 0$;
- an optional time if $\{\tau<t\}\in\mathcal{F}_t$ for each $t\geq 0$;
- a predictable time if $\tau_n\uparrow\tau$ for a sequence stopping times $\{\tau_n\}_{n=1}^{\infty}$ such that

$\tau_n < \tau$ on (the set of $\omega \in \Omega$ with) $\{\tau > 0\}$.

By Exercise 70, stopping times are optional. Further, an optional time τ is a stopping time wrt. $\mathbb{F}^+ = \{\mathcal{F}_t^+\}_{t\geq 0}$, since

$$\{\tau \leq t\} = \bigcap_{n=m}^{\infty} \{\{\tau < t+1/n\} \in \mathcal{F}_{t+1/m} \qquad \text{for each} \quad m \in \mathbb{N}.$$

Essentially for this reason, instead of having both stopping times and optional times, one may consider stopping times wrt. $\mathbb{F}$ and $\mathbb{F}^+$. And there is no need for both stopping times and optional times if all filtrations are right-continuous (cf. Section 2.4).

So far so good. The names we use are also used by, for example, Karatzas & Shreve [22] (though they do not develop stochastic integration far enough to require predictable times), and are the best ones. However, there are several other opinions in the literature, among influential authors. For example,

- Revuz & Yor [33] only have stopping times, by above reason, as do Protter [31];
- Kallenberg [21] only has stopping times (by above reason), and call them both stopping times and optional times;
- Liptser & Shiryayev [27] only have stopping times, but call them Markov times;
- Chung [9] only has stopping times, but call them strictly optional times if they are stopping times wrt. $\mathbb{F}$, and optional times if they are stopping times wrt. $\mathbb{F}^+$.

23.2 When do Smartingales Converge?

For a right-continuous martingale $\{X(t)\}_{t\geq 0}$, the following properties are equivalent

- $\{X(t)\}_{t\geq 0}$ is a uniformly integrable family of random variables;
- $\lim_{t\to\infty} \mathbf{E}\{|X(t)-X_\infty|\} = 0$ for some random variable X_∞ with $\mathbf{E}\{|X_\infty|\} < \infty$;
- $X(t) = \mathbf{E}\{X_\infty | \mathcal{F}_t\}$ for $t \geq 0$, for some random variable X_∞ with $\mathbf{E}\{|X_\infty|\} < \infty$

(by Theorem 11.11). We shall sharpen this result, and generalize it to smartingales.

Theorem 23.2 (Smartingale Convergence) (e.g., [22, Theorem 1.3.15])* *For a right-continuous smartingale, the limit*

$$\lim_{t\to\infty} X(t) \equiv X_\infty \quad \textit{exists a.s. with} \quad \mathbf{E}\{|X_\infty|\} < \infty,$$

under anyone of the following three conditions

(1) *X is a submartingale with* $\sup_{t\geq 0} \mathbf{E}\{X(t)^+\} < \infty$;

(2) *X is a supermartingale with* $\sup_{t\geq 0} \mathbf{E}\{X(t)^-\} < \infty$;

(3) *X is a martingale with* $\lim_{t\to\infty} \mathbf{E}\{|X(t)|\} < \infty$.

The proof goes much in the same way as the corresponding result for discrete time.

In connection with Theorem 23.2, we make the following elementary observations:

(i) By means of considering $-X$, **(1)** implies **(2)**.

(ii) By **(1)** and **(2)**, negative submartingales and positive supermartingales converge.

(iii) To see how **(1)** gives **(3)**, let X be a martingale, so that $|X|$ is a submartingale, by Exercise 55. It follows that

$$\mathbf{E}\{X(t)^+\} \leq \mathbf{E}\{|X(t)|\} \leq \mathbf{E}\big\{\mathbf{E}\{|X(T)|\,|\,\mathcal{F}_t\}\big\} \leq \mathbf{E}\{|X(T)|\} \qquad \text{for} \quad t \in [0,T].$$

Sending $T \to \infty$ and $t \to \infty$ on the right-hand side, in that order, we get

$$\sup_{t\geq 0} \mathbf{E}\{X(t)^+\} \leq \lim_{T\to\infty} \mathbf{E}\{|X(T)|\} = \mathbf{E}\{|X_\infty|\} < \infty.$$

Since also X is a submartingale, **(1)** shows that we have convergence.

(iv) It is the a.s. convergence part of the theorem that has some depth, while the fact that $\mathbf{E}\{|X_\infty|\} < \infty$ is a corollary. To see this, assume **(1)**, and notice that

$$\begin{aligned}\mathbf{E}\{X_\infty^-\} \leq \liminf_{t\to\infty} \mathbf{E}\{X(t)^-\} &= \liminf_{t\to\infty} (\mathbf{E}\{X(t)^+\} - \mathbf{E}\{X(t)\}) \\ &\leq \sup_{t\geq 0} \mathbf{E}\{X(t)^+\} + \sup_{t\geq 0}(-\mathbf{E}\{X(t)\}) < \infty,\end{aligned}$$

by Theorem 23.2 and *Fatou's lemma*, together with the fact that

$$-\mathbf{E}\{X(t)\} = -\mathbf{E}\big\{\mathbf{E}\{X(t)|\mathcal{F}_0\}\big\} \leq -\mathbf{E}\{X(0)\} \qquad \text{for} \quad t \geq 0.$$

Moreover, Theorem 23.2 and *Fatou's lemma* give

$$\mathbf{E}\{X_\infty^+\} \leq \liminf_{t\to\infty} \mathbf{E}\{X(t)^+\} \leq \sup_{t\geq 0} \mathbf{E}\{X(t)^+\} < \infty.$$

(v) For a martingale X, we have

$$\lim_{t\to\infty} \mathbf{E}\{|X(t)|\} < \infty \;\Leftrightarrow\; \lim_{t\to\infty} \mathbf{E}\{X(t)^+\} < \infty \;\Leftrightarrow\; \lim_{t\to\infty} \mathbf{E}\{X(t)^-\} < \infty.$$

This is so because of the fact that

$$\mathbf{E}\{|X(t)|\} = \begin{cases} 2\,\mathbf{E}\{X(t)^-\} + \mathbf{E}\{X(t)\} = 2\,\mathbf{E}\{X(t)^-\} + \mathbf{E}\{X(0)\} \\ 2\,\mathbf{E}\{X(t)^+\} - \mathbf{E}\{X(t)\} = 2\,\mathbf{E}\{X(t)^+\} - \mathbf{E}\{X(0)\} \end{cases}.$$

Definition 23.3 *For a filtration* $\{\mathcal{F}\}_{t\geq 0}$ *we define* $\underline{\mathcal{F}_\infty} \equiv \sigma(\mathcal{F}_t : t \geq 0)$.

Definition 23.4 *A smartingale* $\{X(t), \mathcal{F}_t\}_{t\geq 0}$ *has a* __*last element*__*, if there exists a* (integrable and adapted to $\mathcal{F}_\infty$) *random variable* $X(\infty)$ *such that* $\{X(t), \mathcal{F}_t\}_{t\in[0,\infty]}$ *is a smartingale.*

Theorem 23.5 (SMARTINGALE CONVERGENCE) (e.g., [22, Problems 1.3.19-1.3.20])* *For a right-continuous smartingale* $\{X(t)\}_{t\geq 0}$ *we have* **(1)**$\Rightarrow$**(2)**$\Rightarrow$**(3)**, *and for a right-continuous martingale* **(1)**$\Leftrightarrow$**(2)**$\Leftrightarrow$**(3)**, *for the following three statements*

(1) $\{X(t)\}_{t\geq 0}$ *is a uniformly integrable family of random variables;*

(2) $\lim_{t\to\infty} \mathbf{E}\{|X(t)-X_\infty|\} = 0$ *for some random variable* X_∞ *with* $\mathbf{E}\{|X_\infty|\}<\infty$;

(3) X *has a last element.*

Notice that, by Theorem 23.4, smartingales that converge in mean converge a.s.

23.3 Optional Sampling

EXERCISE 116 Show that, for a right-continuous smartingale $\{X(t)\}_{t\geq 0}$ with a last element, $X(\tau)$ is a well-defined random variable for any (possible non-finite) stopping time τ. (**Hint:** Recall Theorem 4.17.)

Theorem 23.6 (OPTIONAL SAMPLING THEOREM) *Let* $\{X(t)\}_{t\geq 0}$ *be a right-continuous smartingale with a last element, and* $S\leq T$ *two stopping times. The random variable* $X(T)$ (cf. Exercise 116) *is integrable with*

$$X(S)\begin{cases}\leq \mathbf{E}\{X(T)|\mathcal{F}_S\} & \text{if } X \text{ is a submartingale}\\ = \mathbf{E}\{X(T)|\mathcal{F}_S\} & \text{if } X \text{ is a martingale}\\ \geq \mathbf{E}\{X(T)|\mathcal{F}_S\} & \text{if } X \text{ is a supermartingale}\end{cases}.$$

Proof. For a right-continuous adapted process, $\{X(t)\}_{t\geq 0}$, and a stopping time τ, $X(\tau)$ is adapted to $\mathcal{F}_\tau$, since

$$\{X(\tau)\in B\} = \Big(\{X(\tau)\in B\}\cap\{\tau<t\}\Big)\cup\Big(\{X(t)\in B\}\cap\{\tau=t\}\Big)\in\mathcal{F}_t$$

for $t\geq 0$ and $B\in\mathcal{B}(\mathbb{R})$, by the proof of Theorem 4.20 together with Exercise 70.

Corollary 23.7 (OPTIONAL SAMPLING THEOREM) *Let* $\{X(t)\}_{t\geq 0}$ *be a right-continuous smartingale and* $S\leq T$ *two bounded stopping times. The random variable* $X(T)$ *is integrable with*

$$X(S)\begin{cases}\leq \mathbf{E}\{X(T)|\mathcal{F}_S\} & \text{if } X \text{ is a submartingale}\\ = \mathbf{E}\{X(T)|\mathcal{F}_S\} & \text{if } X \text{ is a martingale}\\ \geq \mathbf{E}\{X(T)|\mathcal{F}_S\} & \text{if } X \text{ is a supermartingale}\end{cases}.$$

Proof. Let t_0 be constant such that $T \le t_0$. The process $\hat{X}(t) = X(t)$ for $t \le t_0$, and $\hat{X}(t) = X(t_0)$ for $t \ge t_0$, is a right-continuous smartingale with a last element. Since $\hat{X}(S) = X(S)$ and $\hat{X}(T) = X(T)$, it is enough to show that the corollary holds when X is replaced with $\hat{X}$. This in turn follows from Theorem 23.6. $\square$

EXERCISE 117 Show how *Optional Sampling* gives *Wald's Identity* for bounded stopping times. Show that $\mathbf{E}\{X(\tau)\} = \mathbf{E}\{X(0)\}$ for a martingale $\{X(t)\}_{t\ge 0}$ with a last element, when τ is a stopping time.

We exemplify that *Optional Sampling* is immensely important by showing how it gives the *Optional Stopping theorem* and the *strong Markov property* of Lévy processes. In the next section we see how it feature when judging whether local martingales are martingales. In Section 24.1 we use *Optional Sampling* to prove a famous representation of continuous local martingales as time-changed BM's. Let's start with

Theorem 23.8 *A right-continuous adapted process $\{X(t)\}_{t\ge 0}$ is a martingale iff. $\mathbf{E}\{X(\tau)\} = \mathbf{E}\{X(0)\}$ for each bounded stopping time τ.*

Proof. The implication to the right is remark (i) after Theorem 23.6. For that to the left, pick $0 \le s < t$ and $\Lambda \in \mathcal{F}_s$. For the random time $T = s I_\Lambda(\omega) + t I_{\Lambda^c}(\omega)$ we have

$$\{T \le r\} = \begin{cases} \emptyset \in \mathcal{F}_r & \text{for} \quad r \in [0,s) \\ \Lambda \in \mathcal{F}_s \subseteq \mathcal{F}_r & \text{for} \quad r \in [s,t) \\ \Omega \in \mathcal{F}_r & \text{for} \quad r \in [t,\infty) \end{cases} \in \mathcal{F}_r \qquad \text{for} \quad r \ge 0.$$

Hence it is a bounded stopping time, as is obviously t. It follows that

$$\begin{aligned} \mathbf{E}\{I_\Lambda X(t)\} = \mathbf{E}\{X(t)\} - \mathbf{E}\{I_{\Lambda^c} X(t)\} &= \mathbf{E}\{X(T)\} - \mathbf{E}\{I_{\Lambda^c} X(t)\} \\ &= \mathbf{E}\{I_\Lambda X(s)\} + \mathbf{E}\{I_{\Lambda^c} X(t)\} - \mathbf{E}\{I_{\Lambda^c} X(t)\} \\ &= \mathbf{E}\{I_\Lambda X(s)\}. \quad \square \end{aligned}$$

Example 23.9 (Optional Stopping again) For a right-continuous martingale $\{X(t)\}_{t\geq 0}$, together with a stopping time τ, *Optional Sampling* gives

$$\mathbf{E}\{X(T\wedge\tau)\} = \mathbf{E}\Big\{\mathbf{E}\{X(T\wedge\tau)|\mathcal{F}_0\}\Big\} = \mathbf{E}\{X(0)\}$$

for each bounded stopping time T, since $0 \leq T\wedge\tau$ are bounded stopping times. Hence Theorem 23.8 shows that $X(t\wedge\tau)$ is a martingale.

Consider a right-continuous local martingale $\{X(t)\}_{t\geq 0}$ with localizing sequence $\{\tau_n\}_{n=1}^{\infty}$, so that $X(t\wedge\tau_n)$ is a matingale for $n\in\mathbb{N}$. For a stopping time τ, we have

$$\mathbf{E}\{X(T\wedge\tau\wedge\tau_n)\} = \mathbf{E}\Big\{\mathbf{E}\{X(T\wedge\tau\wedge\tau_n)|\mathcal{F}_0\}\Big\} = \mathbf{E}\{X(0)\}$$

(by *Optional Sampling*) for each bounded stopping time T, since $0 \leq T\wedge\tau\wedge\tau_n$ are bounded stopping times. Hence Theorem 23.8 gives that $X(t\wedge\tau\wedge\tau_n)$ is a martingale, so that $\{X(t\wedge\tau)\}_{t\geq 0}$ is a local martingale (by Theorem 12.6). #

Although *Optional Sampling* is possible for general smartingales only at bounded stopping times, there is a simple trick to circumvent this seemingly serious restriction in many situation. The trick builds on the following simple fact:

Theorem 23.10 *For a random variable* Z *with* $\mathbf{E}\{|Z|\} < \infty$, *together with a stopping time* τ, *we have*

$$\mathbf{E}\{Z|\mathcal{F}_\tau\} = \mathbf{E}\{Z|\mathcal{F}_{\tau\wedge t}\} \quad \text{a.s.} \qquad \text{for} \quad \omega \in \{\omega'\in\Omega : \tau(\omega')\leq t\}.$$

Proof. Pick a $\Lambda\in\mathcal{F}_\tau$. We have $\Lambda\cap\{\tau\leq t\} \in \mathcal{F}_t$, by the definition of $\mathcal{F}_\tau$, while

$$\Big(\Lambda\cap\{\tau\leq t\}\Big)\cap\{\tau\leq s\} = \begin{cases} \Lambda\cap\{\tau\leq s\} \in \mathcal{F}_s & \text{for} \quad s\leq t \\ \Lambda\cap\{\tau\leq t\} \in \mathcal{F}_t & \text{for} \quad s\geq t \end{cases} \in \mathcal{F}_s,$$

so that $\Lambda\cap\{\tau\leq t\} \in \mathcal{F}_\tau$. Hence $\Lambda\cap\{\tau\leq t\} \in \mathcal{F}_{\tau\wedge t}$, by Exercise 118 below, and so

$$\int_\Lambda I_{\{\tau\leq t\}}\mathbf{E}\{Z|\mathcal{F}_\tau\}\,d\mathbf{P} = \int_\Lambda I_{\{\tau\leq t\}} Z\,d\mathbf{P} = \int_\Lambda I_{\{\tau\leq t\}}\mathbf{E}\{Z|\mathcal{F}_{\tau\wedge t}\}\,d\mathbf{P}. \quad \Box$$

Exercise 118 Show that $\mathcal{F}_{S\wedge T} = \mathcal{F}_S\cap\mathcal{F}_T$ for stopping times S and T. (For this reason, one uses the notation $\underline{\mathcal{F}_S\wedge\mathcal{F}_T} \equiv \mathcal{F}_{S\wedge T}$.)

Let Z be a random variable with $\mathbf{E}\{|Z|\}<\infty$, and let τ be a finite (not necessarily bounded) stopping time. To prove that $\mathbf{E}\{Z|\mathcal{F}_\tau\} = Y$ a.s., for some random variable Y, it is enough to check that

$$\mathbf{E}\{Z|\mathcal{F}_\tau\} = Y \quad \text{a.s.} \qquad \text{for} \quad \omega \in \{\tau\leq n\}, \quad \text{for each} \quad n\in\mathbb{N}, \tag{23.1}$$

since this gives

$$\mathbf{P}\Big\{\mathbf{E}\{Z|\mathcal{F}_\tau\}\neq Y\Big\} = \mathbf{P}\Big\{\bigcap_{n=1}^{\infty}\Big\{\mathbf{E}\{Z|\mathcal{F}_\tau\}\neq Y,\,\tau\leq n\Big\}\Big\} = 0.$$

By Theorem 23.10, (23.1) in turn holds iff.

$$\mathbf{E}\{Z|\mathcal{F}_{\tau\wedge n}\} = Y \quad \text{a.s.} \qquad \text{for} \quad \omega\in\{\tau\le n\}, \quad \text{for each} \quad n\in\mathbb{N}.$$

Here $\tau\wedge n$ is a bounded stopping time, and this is the announced trick!

*Example 23.11 (Strong Markov Property of Lévy Processes again) Let $\{X(t)\}_{t\ge0}$ be a right-continuous Lévy process and T a finite stopping time. To show that $\{X(t+T)-X(T)\}_{t\ge0}$ is independent of $\mathcal{F}_T$, with same `fidi's` as $\{X(t)-X(0)\}_{t\ge0}$, by the aruments presented at the beginning of the proof of *Lévy's characterization of BM*, it is enough to show that

$$\mathbf{E}\{e^{i\theta(X(T+t)-X(T+s))}|\mathcal{F}_{T+s}\} = \mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\} \qquad \text{for} \quad \theta\in\mathbb{Q} \quad \text{and} \quad 0\le s<t.$$

To that end, it is in fact enough show that

$$\mathbf{E}\{e^{i\theta(X(T+t)-X(T+s))}|\mathcal{F}_{T+s}\} = \mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\} \qquad \text{for} \quad \omega\in\{T+s<n\},$$

for each $n\in\mathbb{N}$. By *the trick* (Theorem 23.10), this holds iff.

$$\mathbf{E}\{e^{i\theta(X(T+t)-X(T+s))}|\mathcal{F}_{(T+s)\wedge n}\} = \mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\} \qquad \text{for} \quad \omega\in\{T+s<n\},$$

for $n\in\mathbb{N}$, which of course is same thing as

$$\mathbf{E}\{e^{i\theta(X(((T+s)\wedge n)+(t-s))-X((T+s)\wedge n)}|\mathcal{F}_{(T+s)\wedge n}\} = \mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\}$$

for $\omega\in\{T+s<n\}$ and $n\in\mathbb{N}$. Since (by the fact that X is a Lévy process)

$$\mathbf{E}\{e^{i\theta(X(((T+s)\wedge n)+(t-s))-X((T+s)\wedge n)}|\mathcal{F}_{(T+s)\wedge n}\} = \mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\}$$

for $\omega\in\{T+s\ge n\}=\{(T+s)\wedge n=n\}$, this in turn follows provided that

$$\mathbf{E}\{e^{i\theta(X(((T+s)\wedge n)+(t-s))-X((T+s)\wedge n)}|\mathcal{F}_{(T+s)\wedge n}\} = \mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\}.$$

Hence it is enough to show that, for a bounded stopping time τ and $r>0$,

$$\mathbf{E}\{e^{i\theta(X(\tau+r)-X(\tau))}|\mathcal{F}_\tau\} = \mathbf{E}\{e^{i\theta(X(r)-X(0))}\}.$$

From the proof of Theorem 4.23, we have that

$$\mathbf{E}\{e^{i\theta(X(t-s)-X(0))}\} = \Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{t-s}, \tag{23.2}$$

and so it is enough to prove that

$$\mathbf{E}\{e^{i\theta X(t)}\}\Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{-t} \qquad \text{is a martingale,} \tag{23.3}$$

since *Optional Sampling* at the bounded stopping times $\tau\le\tau+t$ then gives

$$\begin{aligned}
&\mathbf{E}\{e^{i\theta(X(\tau+r)-X(\tau))}|\mathcal{F}_\tau\}\\
&= e^{-i\theta X(\tau)}\Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{\tau+r}\mathbf{E}\Big\{e^{i\theta X(\tau+r)}\Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{-(\tau+r)}\Big|\mathcal{F}_\tau\Big\}\\
&= e^{-i\theta X(\tau)}\Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{\tau+r}e^{i\theta X(\tau)}\Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{-\tau}\\
&= \Big(\mathbf{E}\{e^{i\theta(X(1)-X(0))}\}\Big)^{r} = \mathbf{E}\{e^{i\theta(X(r)-X(0))}\}
\end{aligned}$$

[recall that $X(\tau)$ is adapted to $\mathcal{F}_\tau$, by remark (iii) after Theorem 23.6]. However, in view of (23.2), (23.3) is a straightforward generalization of Exercise 30. #

*EXERCISE 119 Prove the strong Markov property for multidimensional diffusion processes under the hypothesis of Theorem 22.1 (cf. Remark 22.3).

23.4 When are Local Martingales Martingales?

In general, local martingales are much more general processes that martingales, even in the presence of integrability. In fact, not even a uniformly integrable local martingale is necessarily a martingale (even if construction of such examples are not trivial). Very useful exact conditions for when a (uniformly integrable) local martingale is a (uniformly integrable) martingale are available:

Definition 23.12 *A right-continuous adapted process $\{X(t)\}_{t\geq 0}$ is of (Dirichlet) class DL, if for each choice of a constant $t_0\geq 0$, the following family of random variables is uniformly integrable*

$$\{X(\tau):\tau \text{ is a stopping time such that } \tau\leq t_0\}.$$

Definition 23.13 *A right-continuous adapted process $\{X(t)\}_{t\geq 0}$ is of (Dirichlet) class D, if the following family of random variables is uniformly integrable*

$$\{X(\tau):\tau \text{ is a finite stopping time}\}.$$

Theorem 23.14 *A right-continuous local martingale $\{X(t)\}_{t\geq 0}$ is a martingale iff. it is of class* DL.

Proof. [$\Leftarrow$] Let X be of class DL with localizing sequence $\{\tau_n\}_{n=1}^\infty$, and pick $0\leq s<t$. Clearly, $X(\tau_n\wedge t)\rightarrow X(t)$ and $X(\tau_n\wedge s)\rightarrow X(s)$ a.s. as $n\rightarrow\infty$. By assumption, $\{X(\tau_n\wedge t)\}_{n=1}^\infty$ is uniformly integrable (since $\tau_n\wedge t\leq t$). Hence (11.5) gives $\mathbf{E}\{|X(\tau_n\wedge t)-X(t)|\}\rightarrow 0$ and $\mathbf{E}\{|X(\tau_n\wedge s)-X(s)|\}\rightarrow 0$. Using Exercise 68 [and that $X(\tau_n\wedge t)$ is a martingale], it follows that (in the sense of convergence in mean)

$$\mathbf{E}\{X(t)|\mathcal{F}_t\}\leftarrow \mathbf{E}\{X(\tau_n\wedge t)|\mathcal{F}_t\}=X(\tau_n\wedge s)\rightarrow X(s).$$

[$\Rightarrow$] Let X be a martingale and pick $t_0\geq 0$. Since $|X|$ is a submartingale, we have

$$\mathbf{E}\Big\{|X(t_0)|\Big|\mathcal{F}_T\Big\}\geq |X(T)| \qquad \text{for stopping times } \quad T\leq t_0,$$

by Corollary 23.5. Hence *Absolute Continuity of the Integral* gives

$$\int_{\{|X(T)|>y\}} |X(T)|\,d\mathbf{P} \le \int_{\{|X(T)|>y\}} \mathbf{E}\Big\{|X(t_0)|\Big|\mathcal{F}_T\Big\}\,d\mathbf{P} = \int_{\{|X(T)|>y\}} |X(t_0)|\,d\mathbf{P}$$
$$\le \int_{\{\sup_{t\in[0,t_0]}|X(t)|>y\}} |X(t_0)|\,d\mathbf{P},$$

which goes to zero uniformly for stopping times $T\le t_0$, as $y\to\infty$, since (7.1) says

$$\mathbf{P}\Big\{\sup\nolimits_{t\in[0,t_0]}|X(t)|>y\Big\} \le \mathbf{E}\{|X(t_0)|\}/y \to 0 \qquad \text{as} \quad y\to\infty. \quad \square$$

Theorem 23.15 *A right-continuous local martingale* $\{X(t)\}_{t\ge 0}$ *is a uniformly integrable martingale iff. it is of class* D.

Proof. $\boxed{\Leftarrow}$ Let X be of class D. Since X is of class DL, X is a martingale, by Theorem 23.14. Further, X is uniformly integrable, by assumption, since

$$\{X(t): t\ge 0\} \subseteq \{X(\tau): \tau \text{ is a finite stopping time}\}.$$

$\boxed{\Rightarrow}$ Let X be a uniformly integrable martingale. Since $|X|$ is a uniformly integrable submartingale, *Smartingale Convergence* shows that $|X|$ has a last element $|X(\infty)|$. Applying *Optional Sampling*, it follows that

$$\mathbf{E}\Big\{|X(\infty)|\Big|\mathcal{F}_T\Big\} \ge |X(T)| \qquad \text{for stopping times} \quad T.$$

Now *Absolute Continuity of the Integral* gives

$$\int_{\{|X(T)|>y\}} |X(T)|\,d\mathbf{P} \le \int_{\{|X(T)|>y\}} \mathbf{E}\Big\{|X(\infty)|\Big|\mathcal{F}_T\Big\}\,d\mathbf{P} = \int_{\{|X(T)|>y\}} |X(\infty)|\,d\mathbf{P}$$
$$\le \int_{\{\sup_{t\in[0,\infty)}|X(t)|>y\}} |X(\infty)|\,d\mathbf{P},$$

which goes to zero uniformly for finite stopping times T, as $y\to\infty$, since

$$\mathbf{P}\Big\{\sup_{t\in[0,\infty)}|X(t)|>y\Big\} = \lim_{n\to\infty}\mathbf{P}\Big\{\sup_{t\in[0,n]}|X(t)|>y\Big\} \le \limsup_{n\to\infty}\frac{\mathbf{E}\{|X(n)|\}}{y} \le \frac{\mathbf{E}\{|X(\infty)|\}}{y}$$

goes to zero as $y\to\infty$, by (7.1), and since the mean of a submartingale is increasing.
$\square$

Example 23.16 (Theorem 12.9 again) A right-continuous local martingale $\{X(t)\}_{t\ge 0}$, such that $|X(t)|\le Z_{t_0}$ a.s. for each $t\in[0,t_0]$, for some random variable Z_{t_0} with $\mathbf{E}\{Z_{t_0}\}<\infty$, for each constant $t_0\ge 0$, is a martingale. Because, for any stopping time $T\le t_0$, we have $|X(T)|\le Z_{t_0}$, so that

$$\int_{\{|X(T)|>y\}} |X(T)|\,d\mathbf{P} \le \int_{\{Z_{t_0}>y\}} Z_{t_0}\,d\mathbf{P} \to 0 \qquad \text{as} \quad y\to\infty$$

uniformly for such T, by *Absolute Continuity of the Integral*. #

Example 23.17 (Completion of Proof of Girsanov's Theorem) Let τ be a stopping time, and $\{Z(t)\}_{t\geq 0}$ and $\{W(t)\}_{t\geq 0}$ right-continuous adapted processes. Assume that $W(\tau\wedge t)$ is a bounded martingale, $Z(t)$ a martingale, and $W(\tau\wedge t)Z(t)$ a local martingale. Since $Z(t)$ is of class DL, by Theorem 23.14, and $W(\tau\wedge t)$ is bounded, also $W(\tau\wedge t)Z(t)$ must be of class DL (by inspection of the definition of uniform integrability). Hence $W(\tau\wedge t)Z(t)$ is a martingale, by Theorem 23.14. #

23.5 Burkholder-Davis-Gundy Inequalities

The following famous inequality we make crucial use of several times:

Theorem 23.18 (Burkholder-Davis-Gundy inequalities) *To each constant* $\alpha > 0$, *there exist constants* $0 < k_\alpha \leq K_\alpha < \infty$, *such that*

$$k_\alpha\,\mathbf{E}\{[X](T)^\alpha\} \leq \mathbf{E}\Big\{\sup_{t\in[0,T]}|X(t)-X(0)|^{2\alpha}\Big\} \leq K_\alpha\,\mathbf{E}\{[X](T)^\alpha\} \qquad \textit{for} \quad T\geq 0,$$

for every continuous local martingale $\{X(t)\}_{t\geq 0}$.

Despite the importance of this result, the proof is a bit special, and does not really forward ones understanding of stochastic calculus. Therefore it is *-marked.

**Proof* (after [21, Chapter 15]). By Section 15.1, we have $X(t)^2 = M(t) + [X](t)$, where M is a continuous local martingale. Let $\{\tau_n\}_{n=1}^\infty$ be a localizing sequence of stopping times for M, so that $\tau_n\to\infty$ a.s. as $n\to\infty$, and $M(\tau_n\wedge t)$ is a martingale for $n\in\mathbb{N}$. Let $\tilde{\tau}_n \equiv \inf\{t\geq 0 : [X](t)\geq n \text{ or } X(t)^2\geq n\}$, and set $\hat{\tau}_n = \tau_n\wedge\tilde{\tau}_n$. The process $\hat{M}(t) = M(\hat{\tau}_n\wedge t) = M(\tau_n\wedge(t\wedge\tilde{\tau}_n))$ is a martingale, by *Optional Stopping*, and is bounded, and thus uniformly integrable (by Theorem 12.9), since $\hat{M} = Y - Z$, with $Y(t)\equiv X(\hat{\tau}_n\wedge t)^2$ and $Z(t)\equiv[X](\hat{\tau}_n\wedge t)$ bounded. It is enough to show that

$$\mathbf{E}\{\sup\nolimits_{t\in[0,T]} Y(t)^\alpha\} \leq K_\alpha\,\mathbf{E}\{\sup\nolimits_{t\in[0,T]} Z(t)^\alpha\}, \tag{23.4}$$

because by *Monotone Convergence* and since Z is increasing, this gives

$$\mathbf{E}\Big\{\sup_{t\in[0,T]}|X(t)|^{2\alpha}\Big\} = \lim_{n\to\infty}\mathbf{E}\Big\{\sup_{t\in[0,T]}Y(t)^\alpha\Big\} \leq \lim_{n\to\infty}K_\alpha\,\mathbf{E}\Big\{\sup_{t\in[0,T]}Z(t)^\alpha\Big\}$$
$$= \lim_{n\to\infty}K_\alpha\,\mathbf{E}\{Z(T)^\alpha\} = K_\alpha\,\mathbf{E}\{[X](T)^\alpha\}.$$

Let $\tau_r \equiv \inf\{t\geq 0 : Y(t)\geq r\}$. The process $N^{(r)}(t) \equiv \hat{M}(t) - \hat{M}(\tau_r\wedge t)$ is a uni- formly integrable martingale, by *Optional Stopping* (since it is bounded). Since

$N^{(r)}(t) \geq M(t) - Y(\tau_r \wedge t) \geq Y(t) - Z(t) - r,$ we have

$$\begin{aligned}
&2^{-\alpha}\mathbf{E}\Big\{\sup_{t\in[0,T]} Y(t)^\alpha\Big\} - c^{-\alpha}\mathbf{E}\Big\{\sup_{t\in[0,T]} Z(t)^\alpha\Big\} \\
&= \int_0^\infty \mathbf{P}\Big\{\sup_{t\in[0,T]} (Y(t)/2)^\alpha > s\Big\}\,ds - \int_0^\infty \mathbf{P}\Big\{\sup_{t\in[0,T]} (Z(t)/c)^\alpha > s\Big\}\,ds \\
&= \int_0^\infty \mathbf{P}\Big\{\sup_{t\in[0,T]} Y(t) > 2s^{1/\alpha}\Big\}\,ds - \int_0^\infty \mathbf{P}\Big\{\sup_{t\in[0,T]} Z(t) > cs^{1/\alpha}\Big\}\,ds \\
&= \int_0^\infty \alpha\, r^{\alpha-1}\,\mathbf{P}\Big\{\sup_{t\in[0,T]} Y(t) > 2r\Big\}\,dr - \int_0^\infty \alpha\, r^{\alpha-1}\,\mathbf{P}\Big\{\sup_{t\in[0,T]} Z(t) > cr\Big\}\,dr \\
&\leq \int_0^\infty \alpha\, r^{\alpha-1}\,\mathbf{P}\Big\{\sup_{t\in[0,T]} Y(t) > 2r,\ \sup_{t\in[0,T]} Z(t) \leq cr\Big\}\,dr \\
&\leq \int_0^\infty \alpha\, r^{\alpha-1}\,\mathbf{P}\Big\{\tau_r < \infty,\ \sup_{t\in[0,T]} (Y(t)-Z(t)) > (2-c)r,\ \inf_{t\in[0,T]} (Y(t)-Z(t)) \geq -cr\Big\}\,dr \\
&\leq \int_0^\infty \alpha\, r^{\alpha-1}\,\mathbf{P}\Big\{\tau_r < \infty,\ \sup_{t\in[0,T]} N^{(r)}(t) > (1-c)r,\ \inf_{t\in[0,T]} N^{(r)}(t) \geq -(c+1)r\Big\}\,dr \\
&= \int_0^\infty \alpha\, r^{\alpha-1}\,\mathbf{P}\Big\{\tau_r < \infty, S^{(r)} < T^{(r)}\Big\}\,dr \qquad \text{for a constant} \quad c \in (0,1),
\end{aligned}$$

where $S^{(r)} \equiv \inf\{t \geq 0 : N^{(r)}(t) \leq -(c+1)r\}$ and $T^{(r)} \equiv \inf\{t \geq 0 : N^{(r)}(t) \geq (1-c)r\}$. Notice that $N(S) = -(c+1)r$ for $S < \infty$ and $N(T) = (1-c)r$ for $T < \infty$, so that we cannot have $S = T$ when $S \wedge T < \infty$. Further, we have $N(S \wedge T) \leq (1-c)r$ for $S \wedge T = \infty$. Since $N = 0$ on $\{\tau_r = \infty\}$, *Optional Sampling* shows that

$$\begin{aligned}
0 &= \mathbf{E}\{N(0)\} \\
&= \mathbf{E}\{N(S\wedge T)\} \\
&= \mathbf{E}\Big\{I_{\{\tau_r<\infty\}} N(S\wedge T)\Big\} \\
&= (1-c)r\,\mathbf{P}\{\tau_r < \infty, S > T\} - (c+1)r\,\mathbf{P}\{\tau_r < \infty, S < T\} + \mathbf{E}\{I_{\{\tau_r<\infty,\, S=T\}} N(S\wedge T)\} \\
&\leq (1-c)r\,\mathbf{P}\{\tau_r < \infty, S \geq T\} - (c+1)r\,\mathbf{P}\{\tau_r < \infty, S < T\} \\
&= (1-c)r\,\mathbf{P}\{\tau_r < \infty\} - 2r\,\mathbf{P}\{\tau_r < \infty, S < T\},
\end{aligned}$$

so that

$$\mathbf{P}\{S < T, \tau_r < \infty\} \leq (1-c)\,\mathbf{P}\{\tau_r < \infty\}/2 = (1-c)\,\mathbf{P}\Big\{\sup_{t\in[0,T]} Y(t) > r\Big\}\Big/2.$$

Putting things together, we conclude that

$$\begin{aligned}
2^{-\alpha}\mathbf{E}\Big\{\sup_{t\in[0,T]} Y(t)^\alpha\Big\} - c^{-\alpha}\mathbf{E}\Big\{\sup_{t\in[0,T]} Z(t)^\alpha\Big\} &\leq \int_0^\infty \alpha\, r^{\alpha-1}\,\frac{1-c}{2}\,\mathbf{P}\Big\{\sup_{t\in[0,T]} Y(t) > r\Big\}\,dr \\
&= \frac{1-c}{2}\,\mathbf{E}\Big\{\sup_{t\in[0,T]} Y(t)^\alpha\Big\}.
\end{aligned}$$

Taking $c = 1 - 2^{-\alpha}$ and rearranging, we get (23.4), with $K_\alpha = 2^{\alpha+1}/(1-2^{-\alpha})^\alpha$. $\square$

*EXERCISE 120 We only proved the right *Burkholder-Davis-Gundy* inequlity above. Explain how the left inequality follows, simply by inspection of that proof, together with a change in the definitions of Y, Z and $\hat{M}$, to $Y(t)\equiv[X](\hat{\tau}_n\wedge t)$, $Z(t)\equiv X(\hat{\tau}_n\wedge t)^2$ and $\hat{M}(t)=-M(\tau_n\wedge t)=Y(t)-Z(t)$.

Corollary 23.19 *For a continuous local martingale $\{X(t)\}_{t\geq 0}$ such that $[X](T)=0$ a.s., we have $X(t)=X(0)$ a.s. for $t\in[0,T]$.*

Proof. Burkholder-Davis-Gundy inequality gives

$$\mathbf{E}\Big\{\sup_{t\in[0,T]}|X(t)-X(0)|^2\Big\}\leq K_1\,\mathbf{E}\{[X](T)\}=0. \quad \square$$

EXERCISE 121 Show that a continuous local martingale $\{X(t)\}_{t\geq 0}$, such that $X(0)=0$ and $\mathbf{E}\{\sqrt{[X](t)}\}<\infty$ for $t\geq 0$, is a martingale. What happens if the requirement that $X(0)=0$ is replaced with $\mathbf{E}\{|X(0)|\}<\infty$?

24 Twentyfourth Lecture

24.1 Change of Time for Local Martingales

The next famous result builds on the martingale results in Lecture 23. It shows that continuous local martingales are BM's, when runned under a suitable clock.

Theorem 24.1 (DAMBIS-DUBINS-SCHWARZ) *Let $\{M(t)\}_{t\geq 0}$ be a continuous local martingale with $M(0)=0$, wrt. a right-continuous filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Assume that $[M](t)\to\infty$ a.s. as $t\to\infty$, and write $T(t)\equiv\inf\{s\geq 0:[M](s)>t\}$. The process*

$$B(t)\equiv M(T(t)),\quad t\geq 0,\quad \textit{is BM wrt. the filtration}\quad \{\mathcal{F}_{T(t)}\}_{t\geq 0}.$$

Further, $[M](t)$ is a stopping time wrt. this filtration, and

$$M(t)=B([M](t))\qquad \textit{for}\quad t\geq 0,\quad \textit{with probability one.}$$

Proof. Since $[M]$ is continuous and adapted (cf. Definition 15.2), we have

$$\{T(t)<s\}=\{[M](s)>t\}\in\mathcal{F}_s\qquad \text{for}\quad s,t\geq 0,$$

so that $T(t)$ is an optional time. Since the filtration is right-continuous, it follows that $T(t)$ is a stopping time (see Section 23.1). Further, we have $\{[M](t)>s\}\in\mathcal{F}_{T(s)}$ for $s\geq 0$, so that $[M](t)$ is a stopping time as specified, since (by above identity)

$$\{[M](t)>s\}\cap\{T(s)\leq t\}=\{T(s)<t\}\cap\{T(s)\leq t\}=\{T(s)<t\}\in\mathcal{F}_t\qquad \text{for}\quad t\geq 0.$$

Pick a $t>0$, and let $M^{(t)}(r)\equiv M(r\wedge T(t))$ for $r\geq 0$. Notice that $B(t)-B(s)=M^{(t)}(T(t))-M^{(t)}(T(s))$ for $0\leq s<t$. Further, $\{M^{(t)}(r)\}_{r\geq 0}$ is a continuous local martingale, by *Optional Stopping* for local martingales (Example 23.9), with

$$[M^{(t)}](s)=[M(\cdot\wedge T(t))](s)=[M](s\wedge T(t))\leq[M](T(t))=t\qquad \text{for}\quad s\geq 0,$$

[by the definition of $T(t)$ and continuity of M]. Hence *Burkholder-Davis-Gundy inequality* together with *Fatou's lemma* show that

$$\mathbf{E}\Big\{\sup_{r\geq 0}|M^{(t)}(r)|^{2\alpha}\Big\}\leq\liminf_{s\to\infty}\mathbf{E}\Big\{\sup_{r\in[0,s]}|M^{(t)}(r)|^{2\alpha}\Big\}\leq\lim_{s\to\infty}K_\alpha\,\mathbf{E}\{[M^{(t)}](s)^{1/2}\}\leq K_\alpha t^\alpha$$

for $\alpha>0$. taking $\alpha=1/2$, it follows that $M^{(t)}$ is of class D, since

$$\int_{\{|M^{(t)}(\tau)|>y\}}|M^{(t)}(\tau)|\,d\mathbf{P}\leq\int_{\{\sup_{r\geq 0}|M^{(t)}(r)|>y\}}\sup\nolimits_{r\geq 0}|M^{(t)}(r)|\,d\mathbf{P}\to 0$$

uniformly for finite stopping times τ as $y\to\infty$, by *Absolute Continuity of the Integral* [applied to the integrable random variable $\sup_{r\geq 0}|M^{(t)}(r)|$]. Hence $\{M^{(t)}(r)\}_{r\geq 0}$ is a uniformly integrable martingale, by Theorem 23.15, and *Optional Sampling* gives

$$\mathbf{E}\{B(t)|\mathcal{F}_{T(s)}\}=\mathbf{E}\{M^{(t)}(T(t))|\mathcal{F}_{T(s)}\}=M^{(t)}(T(s))=B(s).$$

Hence B is a martingale wrt. $\{\mathcal{F}_{T(t)}\}_{t\geq 0}$, and if B is continuous, we get that B is BM, by *Levy's Characterization of BM*, since we know that $[B](t) = [M](T(t)) = t$.

The process $[M]$ is increasing, and discontinuities of the process T appears exactly at intervals of constancy of $[M]$. By Corollary 23.19, such intervals coincide with intervals of constancy of M. From this we get that the composition $B = X \circ T$ is continuous (although T not necessarily is).

It is a technical exercise, to conclude from the previous paragraph, that $B([M](t)) = M(T([M](t))) = M(t)$. This concludes the proof of the theorem. □

EXERCISE 122 Explain graphically why it is that $B(t)=M(T(t))$ is continuous.

EXERCISE 123 Discuss how *Dambis-Dubins-Schwarz theorem* can be extended to the case when not necessarily $M(0)=0$.

EXERCISE 124 Show how *Dambis-Dubins-Schwarz theorem* together with *Wald's Identity* gives *Novikov's Criterion*. (**Hint:** Exercise 101.)

With the hypothesis of Theorem 24.1, let X be progressively measurable such that

$$\int_0^t X(r)^2 \, d[M](r) < \infty \quad \text{a.s.} \qquad \text{for each} \quad t \geq 0,$$

so that $\int_0^t X\, dM$ is well-defined for $t\geq 0$. Since, with the notation of Theorem 24.1, $M(T(t))=B(t)$ and $M(t)=B([M](t))$, writing $Y(t)=X(T(t))$, we expect that

$$X(T(t))\, dM(T(t)) = Y(t)\, B(t) \qquad \text{and} \qquad X(t)\, dM(t) = Y([M](t))\, dB([M](t)).$$

Of course, the exact meaning of this statement is that

$$\int_0^{T(t)} X\, dM = \int_0^t Y\, B \qquad \text{and} \qquad \int_0^t X\, dM = \int_0^{[M](t)} Y\, dB. \tag{24.1}$$

[A rigorous proof of (24.1) is a bit technical (e.g., [22, pp. 176-178])*. We only use (24.1) in Example 24.2 below, the outcome of which is reached by other methods in Example 25.1. Hence we need not worry about a proof of (24.1).]

Theorem 24.1, together with (24.1), can be used to solve SDE:

Example 24.2 Consider a one-dimensional diffusion type SDE without drift

$$dX(t) = \sigma(X(t))\, dB(t) \qquad \text{for} \quad t \geq 0,$$

where $\sigma : \mathbb{R} \to (0,\infty)$ is continuous and bounded away from zero. By *Engelbert-Schmidt* theorem, the SDE has a unique weak solution, and clearly

$$[X](t) = \int_0^t \sigma(X(r))^2 \, dr \to \infty \quad \text{a.s.} \qquad t \to \infty.$$

Writing $\tilde{B}(t) \equiv X(T(t))$, where $T(t) \equiv \inf\{s\geq 0 : [X](s) > t\}$, Theorem 24.1 shows that $\tilde{B}(t)=X(T(t))$ is BM wrt. $\{\mathcal{F}_{T(t)}\}_{t\geq 0}$, and that $\tilde{B}([X](t))=X(t)$. Notice that

$$dB(T(t)) = \frac{1}{\sigma(X(T(t)))}\, dX(T(t)) = \frac{1}{\sigma(\tilde{B}(t))}\, d\tilde{B}(t)$$

(by the remarks after Theorem 24.1), so that (by same remarks)

$$dB(t) = \int \frac{1}{\sigma(\tilde{B}([X](t)))}\, d\tilde{B}([X](t)) \qquad \text{and} \qquad B(t) = \int_0^{[X](t)} \frac{1}{\sigma(\tilde{B}(r))}\, d\tilde{B}(r).$$

For the corresponding quadratic variations, this means that the stopping time

$$\tau_t \equiv [X](t) \qquad \text{is given by} \qquad t = [B](t) = \int_0^{\tau_t} \frac{1}{\sigma(\tilde{B}(r))^2}\, d[\tilde{B}]r = \int_0^{\tau_t} \frac{1}{\sigma(\tilde{B}(r))^2}\, dr.$$

We have shown that there exists a BM $\tilde{B}$ such that $X(t) = \tilde{B}(\tau_t)$ solves the SDE. (This solution is a weak solution, since it makes referens to another BM than B.) #

Remark 24.3 There is an extension of Theorem 24.1, Knight's theorem, to multidimensional continuous local martingales (e.g., [22, Theorem 4.13])*. #

24.2 Semimartingales

Until recently, the most general stochastic integrals (and thus SDE) known to mankind, were integrals of predictable processes wrt. semimartingales. This has changed the last few years, but we shall stay with outlining the development before that. These results are known as (due to) the french shool, and referred to as general theory of stochastic processes (a bit ambitiously). (A source on the general theory that should be reasonably accessible to a beginner is [21, Chapters 22-23]*.)

Definition 24.4 *An adapted cádlág process $\{Y(t)\}_{t\geq 0}$, such that $Y = M + A$, with $\{M(t)\}_{t\geq 0}$ a cádlág local martingale and $\{A(t)\}_{t\geq 0}$ an adapted cádlág process with finite variation, is a semimartingale.*

Exercise 125 Explain why we cannot hope to integrate an adapted cádlág (but not continuous) process, wrt. a cádlág (but not continuous) process. Explain why we can hope to integrate an adapted left-continuous process wrt. a cádlág process.

Definition 24.5 *A stochastic process $\{X(t)\}_{t\geq 0}$ that is measurable wrt. the σ-algebra on $\Omega\times[0,\infty)$ generated by all adapted left-continuous processes* [i.e. the smallest σ-algebra on $\Omega\times[0,\infty)$ that makes $Y:\Omega\times[0,\infty)\to\mathbb{R}$ Borel-measurable, for every adapted and left-continuous process Y], *is a predictable process.*

Now left-continuous processes come into play, wich are referred to as càg (= "*continu à gauche*"). We have the following quite easy result:

Theorem 24.6 *Adapted càg processes are predictable. If $\{X(t)\}_{t\geq 0}$ is adapted càdlàg, then* $\underline{X_-(t)} \equiv X(t^-)$ *is predictable. If $\{X(t)\}_{t\geq 0}$ is càdlàg predictable, then* $\underline{\Delta X} \equiv X - X_-$ *is predictable. Predictable processes are progressively measurable.*

EXERCISE 126 Prove Theorem 24.6.

The following result, interesting from a general point of view, is also quite easy to prove, by appropriate optional stopping, but not important enough for us to do that:

Theorem 24.7 (e.g., [20, Lemma 4.24] and [21, Proposition 22.16])* *A cádlág local martingale is predictable iff. it is continuous a.s. A semimartingale is continuous iff. it is the sum of a continuous local martingale and a continuous finite variation process.*

The integral of a predictable process X wrt. a semimartingale $Y = M + A$,

$$\int_0^t X(r)\, dY(r) = \int_0^t X(r)\, dM(r) + \int_0^t X(r)\, dA(r),$$

consists of an integral wrt. the càdlàg local martingale M (a generalization of the integral wrt. continuous local martingales from Section 16.1), together with a signed Lebesgue-Stieltjes integral wrt. the finite variation process A. The integral is done in the same way as in Lectures 7-11, by first integrating simple processes, secondly square-integrable ones, and lastly general (locally bounded) predictable processes.

Definition 24.8 *A simple predictable process $\{X(t)\}_{t\geq 0}$ is given by*

$$X(t) = X(0) I_{\{0\}}(t) + \sum_{i=1}^{n} I_{(t_{i-1},t_i]}(t) X_{t_{i-1}} \qquad \text{for} \quad t \geq 0,$$

where $0 = t_0 < t_1 < \ldots < t_n < \infty$ are constants and $X(0), X_{t_0}, \ldots, X_{t_{n-1}}$ bounded random variables adapted to $\mathcal{F}_0, \mathcal{F}_0, \ldots, \mathcal{F}_{t_{n-1}}$, respectively.

Definition 24.9 *The Itô integral of a simple predictable process $\{X(t)\}_{t\geq 0}$ wrt. a semimartingale $\{Y(t)\}_{t\geq 0}$ is defined by [$\int_0^0 X(r)\, dY(r) = 0$ and]*

$$\int_0^t X(r)\, dY(r) \equiv \sum_{i=1}^{n} X_{t_{i-1}} (Y(t_i) - Y(t_{i-1})) + X_{t_m} (Y(t) - Y(t_m)) \qquad \text{for} \quad t \in (t_m, t_{m+1}].$$

Definition 24.10 *A measurable stochastic process $\{X(t)\}_{t\geq 0}$ is locally bounded, if there exists localizing sequence of stopping times $0 \leq \tau_1 \leq \tau_2 \leq \ldots$, with $\lim_{n\to\infty} \tau_n = \infty$ a.s., such that the process $\{X(t \wedge \tau_n)\}_{t\geq 0}$ is bounded for each $n \in \mathbb{N}$.*

Example 24.11 A cág process X is locally bounded [take $\tau_n = \inf\{t \geq 0 : |X(t)| \geq n\}$]. In particular, for a semimartingale X, X_- is locally bounded (since cág), and predictable (by Theorem 24.6), which is essential in Definition 24.14 below. Simple predictable processes are locally bounded (since bounded). #

Example 24.12 A cádlág Lévy process with locally bounded means is a semimartingale, by Exercise 41. In particular a PP is a semimartingale, which alternatively follows directly from that it is a finite variation process (because increasing). #

Now we are in shape to face the announced Itô integral wrt. semimartingales:

Theorem 24.13 (e.g., [20, pp. 46-47])* *For a semimartingale* $\{Y(t)\}_{t\geq 0}$, *the map*

$$[0,\infty) \times \Big\{\text{“simple predictable processes”}\Big\} \ni (t, X) \to \int_0^t X\, dY$$

has a unique extension to an <u>Itô integral</u>

$$[0,\infty) \times \Big\{\text{“locally bounded predictable processes”}\Big\} \ni (t, X) \to \underline{\int_0^t X\, dY}$$

with the following properties

(1) $\{\int_0^t X\, dY\}_{t\geq 0}$ *is a semimartingale;*

(2) $\int_0^t (aX_1 + bX_2)\, dY = a\int_0^t X_1\, dY + b\int_0^t X_2\, dY$ *for constants* $a, b \in \mathbb{R}$;

(3) *If* Y *is a (continuous) local martingale, then so is* $\{\int_0^t X\, dY\}_{t\geq 0}$;

(4) *If* Y *has finite variation, then* $\{\int_0^t X\, dY\}_{t\geq 0}$ *has finite variation and coincides with the corresponding Lebesgue-Stieltjes integral;*

(5) $(\Delta \int_0^\cdot X\, dY)(t) = X(t)\Delta Y(t)$, *so that* $\{\int_0^t X\, dY\}_{t\geq 0}$ *is continuous when* Y *is;*

(6) *If* $X, Z, X_1, X_2, \ldots$ *are locally bounded predictable processes such that* $\lim_{k\to\infty} X_k(t) = X(t)$ *a.s., with* $|X_k(t)| \leq |Z(t)|$ *for* $k \in \mathbb{N}$ *and* $t \geq 0$, *we have*

$$\operatorname*{P-lim}_{k\to\infty} \sup_{s\in[0,t]} \left| \int_0^s X_k\, dY - \int_0^s X\, dY \right| = 0 \qquad \text{for each} \quad t \geq 0.$$

The Itô integral wrt. semimartingales keeps the connection to “elementary Itô integrals” (approximating sums), as long as the integrated predictable process is cág:

Theorem 24.14 (e.g., [20, Proposition 4.44])* *For the Itô integral of a cág adapted process* $\{X(t)\}_{t\geq 0}$ *wrt. a semimartingale* $\{Y(t)\}_{t\geq 0}$, *we have*

$$\int_0^t X\, dY = \text{P-lim}\left\{ \sum_{i=1}^n X(t_{i-1})\,(Y(t_i) - Y(t_{i-1})) : \begin{array}{c} 0 = t_0 < t_1 < \ldots < t_n = t \\ \max_{1\leq i\leq n} t_i - t_{i-1} \to 0 \end{array} \right\}.$$

To give an *Itô formula* for the stochastic integrals wrt. semimartingales, we need quadratic variations and covariations for such. Since the previously developed theory only applies to continuous processes, appropriate generalizations are required:

Definition 24.15 *The covariation between two semimartingales $\{X(t)\}_{t\geq 0}$ and $\{Y(t)\}_{t\geq 0}$ is defined*

$$[X,Y](t) \equiv X(t)Y(t) - X(0)Y(0) - \int_0^t X_-(r)\,dY(r) - \int_0^t Y_-(r)\,dX(r) \qquad \text{for} \quad t\geq 0.$$

The quadratic variation of a semimartingale $\{X(t)\}_{t\geq 0}$ is defined $[X](t) \equiv [X,X](t)$.

The general covariation has properties similar to the previously considered one. However, to use this covariation is more difficult, since it need not be continuous.

Example 24.16 Consider the stochastic differentials (cf. Section 16.1)

$$dX(t) = a_1(t)\,dt + b_1(t)\,dZ_1(t) \qquad \text{and} \qquad dY(t) = a_2(t)\,dt + b_2(t)\,dZ_2(t),$$

for Z_i continuous local martingales, and a_i, b_i progressively measurable with

$$\int_0^T |a_i(r)|\,dr < \infty \qquad \text{and} \qquad \int_0^T b_i(r)^2\,d[X,X](r) < \infty \qquad \text{with probability one.}$$

Such processes X and Y are the most general for which we have previosuly considered the covariation $[X,Y]$ (cf. Sections 15.1 and 16.1): By *Integration by Parts* (Examples 15.10 and 16.1), that covariation coincides with that in Definition 24.15. #

Theorem 24.17 (e.g., [21, Theorem 23.6])* *For semimartingales $\{X(t)\}_{t\geq 0}$ and $\{Y(t)\}_{t\geq 0}$, it holds that*

(1) $\{[X,Y](t)\}_{t\geq 0}$ *is a semimartingale;*

(2) $\{[X,Y](t)\}_{t\geq 0}$ *is a finite variation process a.s.;*

(3) $\{[X](t)\}_{t\geq 0} = \{[X,X](t)\}_{t\geq 0}$ *is an increasing process a.s.;*

(4) $\Delta[X,Y](t) = \Delta X(t)\Delta Y(t)$ *for* $t\geq 0$ *a.s.;*

(5) $[X,Y](t) = \sum_{s\leq t}\Delta X(s)\Delta Y(s)$ *for* $t\geq 0$ *a.s., if* X *or* Y *has finite variation;*

(6) $[\int_0^\cdot Z\,dX, Y](t) = \int_0^t Z\,d[X,Y](t)$ *for* $t\geq 0$ *a.s., if* $\{Z(t)\}_{t\geq 0}$ *is a locally bounded predictable process;*

(7) $[X,Y](t) = \text{P-lim}\left\{\sum_{i=1}^n (X(t_i)-X(t_{i-1}))\,(Y(t_i)-Y(t_{i-1})) : \begin{array}{c} 0=t_0<t_1<\ldots<t_n=t \\ \max_{1\leq i\leq n} t_i-t_{i-1}\to 0 \end{array}\right\}.$

Since $[X,Y]$ has finite variation, it has a well-defined and finite continuous part

$$\underline{[X,Y]^c(t)} \equiv [X,Y](t) - \sum_{s\le t} \Delta[X,Y](t) = [X,Y](t) - \sum_{s\le t} \Delta X(t)\Delta Y(t) \qquad \text{for} \quad t\ge 0.$$

Theorem 24.18 (Itô's Formula) (e.g., [20, Theorem 4.57])* *Let* $\{X(t)\}_{t\ge0} = \{(X_1(t),\dots,X_n(t))\}_{t\ge0}$ *be an* n*-dimensional process with semimartingale components* $X_1,\dots,X_n$. *For a function* $f\in\mathbb{C}^2(\mathbb{R}^n)$, $\{f(X(t))\}_{t\ge0}$ *is a semimartingale and*

$$f(X(t)) = f(X(0)) + \sum_{i=1}^n \int_0^t \partial_i f(X_-)\,dX_i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^t \partial_i\partial_j f(X_-)\,d[X_i,X_j]^c(t)$$
$$+ \sum_{s\le t}\Big(f(X(s)) - f(X_-(s)) - \sum_{i=1}^n \partial_i f(X_-(s))\,\Delta X_i(s)\Big).$$

Example 24.19 By Definition 24.15, a PP $\{N(t)\}_{t\ge0}$ has quadratic variation

$$\begin{aligned}[N](t) &= N(t)^2 - N(0)^2 - 2\int_0^t N_-(r)\,dN(r)\\ &= N(t)^2 - N(0)^2 - 2\sum_{k=N(0)}^{N(t)-1} k\\ &= N(t)^2 - N(0)^2 - (N(t)-1)\,N(t) + (N(0)-1)\,N(0) = N(t) - N(0),\end{aligned}$$

which we also get from (7) in Theorem 24.17. Since $[N]^c(t)=0$, *Itô's formula* gives

$$f(N(t)) = f(N(0)) + \int_0^t f'(N_-)\,dN + \sum_{s\le t}\Big(f(N(s)) - f(N_-(s)) - f'(N_-(s))\,\Delta N(s)\Big).$$

In fact, this identity holds trivially, just by inspection (more or less).

In general, one establishes *Itô's formula* for semimartingales by decomposing them into a continuous part, to which the previous Itô formula applies, and to a jump part, for which Itô's formula is more or less trivial, in the above fashion. #

Remark 24.20 The reader may now think he/she has seen every kind of martingale there is. But two such have not been mentioned; special semimartingales, which are semimartingales where the component with finite variation can be chosen predictable, and quasimartingales, which are processes that can be obtained by localizing special martingales (so that their local martingale components become martingales and finite variation components integrable). It turns out that a process is a special semimartingale iff. it is the difference between two cádlág nonnegative local supermartingales, and thus a quasimartingale iff. the difference between two nonnegative cádlág supermartingales.

By Exercise 41, cádlág Lévy processes with locally bounded means are special semimartingales (since their finite variations components are nonrandom continuous, and thus predictable), as well as quasimartingales (since "already localized"). #

25 Twentyfifth Lecture

25.1 Application of Martingale Problems

Example 25.1 (Adaption of [25, Theorem 7.34])* The result of Example 24.2 can be obtained in the following direct way, which does not involve random change of time (but more computional details beacuse of that):

Since $X = B$ solves the SDE $dX = dB$, with generator $(\mathcal{A}f)(x) = \frac{1}{2}f''(x)$,

$$f(B(t)) - f(B(0)) - \int_0^t \frac{f''(B(r))}{2}\,dr, \quad t \geq 0, \qquad \text{is a martingale for } f \in \mathbb{C}_B^2(\mathbb{R}),$$

wrt. a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Take $\sigma: \mathbb{R} \to (0,\infty)$ continuous and bounded, and let

$$\text{the random time } \tau_t \text{ be given by} \qquad t = \int_0^{\tau_t} \frac{1}{\sigma(B(r))^2}\,dr \qquad \text{for } t \geq 0.$$

By elementary calculus, $\tau \in \underline{\mathbb{C}^1(\mathbb{R})} \equiv \{\tilde{g}: \mathbb{R} \to \mathbb{R} : \tilde{g} \text{ is continuously differentiable}\}$, with derivative $\tau_t' = \frac{d}{dt}\tau_t = \sigma(B(\tau_t))^2$ wp. 1. The increasing random times $\{\tau_t\}_{t\geq 0}$ are bounded, by the form of τ', and stopping times, by observing that

$$\{\tau_t \leq s\} = \{\tau_t > s\}^c = \left\{\int_0^s \frac{1}{\sigma(B(r))^2}\,dr < t\right\}^c \in \mathcal{F}_s.$$

Using *Optional Sampling*, we hence obtain

$$\mathbf{E}\left\{f(B(\tau_t)) - f(B(0)) - \int_0^{\tau_t} \frac{f''(B(r))}{2}\,dr \,\middle|\, \mathcal{F}_{\tau_s}\right\} = f(B(\tau_s)) - f(B(0)) - \int_0^{\tau_s} \frac{f''(B(r))}{2}\,dr$$

for $0 \leq s < t$, so that the following process is a martingale wrt. $\{\mathcal{F}_{\tau_t}\}_{t\geq 0}$

$$f(B(\tau_t)) - f(B(0)) - \int_0^{\tau_t} \frac{f''(B(r))}{2}\,dr = f(B(\tau_t)) - f(B(\tau_0)) - \int_0^{\tau_t} \frac{f''(B(r))}{2}\,dr.$$

Here, by elementary calculus, writing $X(t) \equiv B(\tau_t)$, the right-hand side is equal to

$$f(X(t)) - f(X(0)) - \int_0^t \frac{f''(B(\tau_r))}{2}\,\tau_r'\,dr = f(X(t)) - f(X(0)) - \int_0^t \frac{\sigma(X(r))^2}{2} f''(X(r))\,dr.$$

Hence also this process is a martingale, so that X solves the martingale problem for the generator $(\mathcal{A}f)(x) = \frac{1}{2}\sigma(x)^2 f''(x)$, and is a weak solution to $dX = \sigma(X)\,dB$. #

By Example 18.4 and Remark 18.5, together with Example 25.1, we can solve a general time homogeneous one-dimensional SDE with bounded continuous coefficients:

Example 25.2 Let $dX = \mu(X)/\sigma(X)^2\,dt + dB$ (cf. Example 18.4 and Remark 18.5), with generator $(\mathcal{A}f)(x) = \frac{\mu(x)}{\sigma(x)^2} f'(x) + \frac{1}{2}f''(x)$, where $\mu: \mathbb{R} \to (0,\infty)$ is measurable and $\sigma: \mathbb{R} \to (0,\infty)$ bounded continuous with μ/σ^2 bounded. The process

$$M^f(t) \equiv f(X(t)) - f(X(0)) - \int_0^t (\mathcal{A}f)(X(r))\,dr \qquad \text{is a martingale for } f \in \mathbb{C}_B^2(\mathbb{R}),$$

wrt. a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Take $\sigma: \mathbb{R} \to (0,\infty)$ continuous and bounded, and let the random time τ_t be given by $\quad t=\int_0^{\tau_t}\frac{1}{\sigma(X(r))^2}\,dr \quad$ for $t\geq 0$,

so that $\tau_t'=\sigma(X(\tau_t))^2$. We get that the increasing random times $\{\tau_t\}_{t\geq 0}$ are bounded and stopping times exactly as in Example 25.1 (since X is continuous and adapted), and that $\{M^f(\tau_t)\}_{t\geq 0}$ is a martingale wrt. $\{\mathcal{F}_{\tau_t}\}_{t\geq 0}$. Again, we can rewrite

$$M^f(\tau_t)=f(Y(t))-f(X(0))-\int_0^t(\mathcal{A}f)(Y(r))f''(Y(r))\,\sigma(Y(r))^2\,dr,$$

where now $Y(t)=X(\tau_t)$. Hence Y solves the martingale problem for $(\mathcal{A}f)(x)=\mu(x)f'(x)+\frac{1}{2}\sigma(x)^2f''(x)$, and thus the SDE $dY=\mu(Y)\,dt+\sigma(Y)\,B$. #

Example 25.3 (Bessel Equation) (Adaption of [25, Section 6.10])* Let B be n-dimensional BM, and put $R(t)=\|B(t)\|^2$. By *Itô's formula*, we have

$$\begin{aligned}dR(t)=d(B_1(t)^2+\ldots+B_n(t)^2)&=2\sum_{i=1}^n B_i(t)\,dB_i(t)+\sum_{i=1}^n\sum_{j=1}^n d[B_i,B_j](t)\\&=2\sum_{i=1}^n B_i(t)\,dB_i(t)+n\,dt.\end{aligned}$$

We may rewrite this as a multidimensional non-diffusiontype SDE

$$d\overline{R}(t)=d(R(t),\ldots,R(t))=\mu\,dt+\sigma(t)\,dB(t)$$

where $\mu_i=n$ and $(\sigma(t))_{i,j}=2\,B_j(t)$. For the corresponding PDO [cf. (19.3)]

$$\begin{aligned}(Ag)(t,x)&=\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(\sigma\sigma^T)_{i,j}(t)\,\frac{\partial^2 g(t,x)}{\partial x_i\partial x_j}+\sum_{i=1}^n\mu_i(t)\,\frac{\partial g(t,x)}{\partial x_i}\\&=2\,R(t)\sum_{i=1}^n\sum_{j=1}^n\frac{\partial^2 g(t,x)}{\partial x_i\partial x_j}+n\sum_{i=1}^n\frac{\partial g(t,x)}{\partial x_i},\end{aligned}$$

Corollary 19.11 shows that the following process is a continuous local martingale

$$g(t,\overline{R}(t))-g(0,\overline{R}(0))-\int_0^t\Big(\partial_1 g(r,\overline{R}(r))+(Ag)(r,\overline{R}(r))\Big)\,dr,\quad t\geq 0,$$

for $g\in\mathbb{C}^{1,2}([0,\infty)\times\mathbb{R}^n)$. Taking $g(t,x)=f(x_1)$, it follows that the following process is a continuous local martingale

$$f(R(t))-f(R(0))-\int_0^t(\mathcal{A}f)(R(r))\,dr,\qquad\text{for}\quad f\in\mathbb{C}^2(\mathbb{R}),$$

where $(\mathcal{A}f)(x)=2\,x\,f''(x)+n\,f'(x)$. Hence the Bessel process R solves the

one-dimensional SDE $\qquad dR(t)=2\sqrt{R(t)}\,dB_1(t)+n\,dt.$

[The corresponding Kolmogorov backward equation is

$$\partial_t f(t,x)+(\mathcal{A}f)(t,x)=D_t f(t,x)+2\,x\,D_x^2 f(t,x)+n\,D_x f(t,x)=0.$$

By Laplace transforming $F(\lambda,x)=\int_0^\infty e^{-\lambda t}f(t,x)\,dt$, it becomes a Bessel type ODE

$$\lambda\,F(\lambda,x)-F(0,x)+2\,x\,D_x^2 F(\lambda,x)+n\,D_x F(\lambda,x)=0.]\quad \#$$

EXERCISE 127 Discuss possibly negative values of R in the Bessel SDE.

25.2 Stationary One-Dimensional Diffusion Processes

In the remainder of these notes, we do explicit calculations for one-dimensional diffusions. A classic and rich introductory source to such material is [24, pp. 157-397].

Definition 25.4 *A stationary distribution for a Markov process* $\{X(t)\}_{t\geq 0}$, *is a probability distribution* ν *on* $\mathbb{R}$ *such that*

$$\int_{\mathbb{R}} \mathbf{P}\{X(t+s)\in\cdot\,|\,X(s)=x\}\,d\nu(x) = \nu(\cdot) \qquad \text{for} \quad t>0 \quad \text{and} \quad s\geq 0.$$

Theorem 25.5 *Let* $\{X(t)\}_{t\geq 0}$ *be a Markov process with stationary distribution* ν. *If* $X(s) =_{\text{distribution}} \nu$, *then we have* $X(t) =_{\text{distribution}} \nu$ *for each* $t\geq s$.

Proof. When $X(s) =_{\text{distribution}} \nu$, we have

$$\begin{aligned}\mathbf{P}\{X(t)\in\cdot\} &= \int_{\mathbb{R}} \mathbf{P}\{X(t)\in\cdot\,|\,X(s)=x\}\,dF_{X(s)}(x)\\ &= \int_{\mathbb{R}} \mathbf{P}\{X(t)\in\cdot\,|\,X(s)=x\}\,d\nu(x) = \nu(\cdot) \qquad \text{for} \quad t\geq s. \quad \square\end{aligned}$$

EXERCISE 128 Let $\{X(t)\}_{t\geq 0}$ be a time homogeneous Markov process. Show that X is a stationary process $\{X(t+h)\}_{t\geq 0} =_{\text{same fidi's}} \{X(t)\}_{t\geq 0}$ for each constant $h\geq 0$, iff. $X(0) =_{\text{distribution}} \nu$ where ν is a stationary distribution.

Warning 25.6 Some Markov literature call time homogeneous Markov processes stationary, contrary to non-Markov literature, where stationarity means translation invariance of the `fidi's` (see Exercise 128). However, from a Markovian point of view, the notation is natural, since the really important feature of a Markov process is the transition probability, which cannot separate stationarity from time homogeneity. #

The "evolution" of a time-homogeneous Markov processes is studied through that of its transition probability $P(\cdot,t,\cdot,0)$. To avoid technical problems (e.g., to change order between derivatives and integrals), it is helpful to Laplace transform time:

Definition 25.7 *Consider a time homogeneous Feller process* $\{X(t)\}_{t\geq 0}$ *with transition probability* $P(\cdot,t,x,0) = \mathbf{P}\{X(t+s)\in\cdot\,|\,X(s)=x\}$. *The resolvent is the family of operators* $\{(G_\lambda:\mathbb{C}_B(\mathbb{R})\to\mathbb{C}_B(\mathbb{R})) : \lambda>0\}$ *given by*

$$(G_\lambda f)(x) = \int_0^\infty e^{-\lambda t}\,\mathbf{E}\{f(X(t+s))|X(s)=x\}\,dt = \int_0^\infty e^{-\lambda t}\Big(\int_{\mathbb{R}} f(\cdot)\,dP(\cdot,t,x,0)\,dy\Big)dt.$$

EXERCISE 129 Show formally, that for each $\lambda>0$, the resolvent $(G_\lambda f)(x)$ for a one-dimensional SDE $dX(t)=\mu(X)\,dt+\sigma(X)\,dB$ satisfies the ODE

$$\mu(x)D_x(G_\lambda f)(x)+\frac{1}{2}\sigma(x)^2D_x^2(G_\lambda f)(x)-\lambda\,(G_\lambda f)(x)=-f(x)\qquad\text{for}\quad x\in\mathbb{R}.$$

Define the resolvent of a multidimensional time homogeneous Feller processes. Extend the above ODE to a corresponding PDE in the case of a multidimensional diffusion.

Theorem 25.8 *Consider the time homogeneous one-dimensional SDE*

$$dX(t)=\mu(X(t))\,dt+\sigma(X(t))\,dB(t)\qquad\textit{for}\quad t\geq 0,$$

where $\mu,\,\sigma:\mathbb{R}\to\mathbb{R}$ *are bounded continuous, such that*

$$\int_{\mathbb{R}}\exp\Big\{\int_0^y\frac{2\,\mu(z)\,dz}{\sigma(z)^2}\Big\}\frac{dy}{\sigma(y)^2}<\infty\qquad\textit{and}\qquad\lim_{y\to\pm\infty}\exp\Big\{\int_0^y\frac{2\,\mu(z)\,dz}{\sigma(z)^2}\Big\}=0,\qquad(25.1)$$

and such that the generator $\mu(x)D_x+\frac{1}{2}\sigma(x)^2D_x^2$ *has a fundamental measure. The diffusion process* X *has a stationary distribution with density function*

$$\pi(x)=\frac{1}{\sigma(x)^2}\exp\Big\{\int_0^x\frac{2\,\mu(z)\,dz}{\sigma(z)^2}\Big\}\Big/\Big(\int_{\mathbb{R}}\exp\Big\{\int_0^y\frac{2\,\mu(z)\,dz}{\sigma(z)^2}\Big\}\frac{dy}{\sigma(y)^2}\Big)\qquad\textit{for}\quad x\in\mathbb{R}.$$

Proof. Clearly, π is a probability density function, and we have to show that it is a stationary density. Writing $\hat{P}$ for the fundamental measure, this holds if

$$\begin{aligned}\int_{\mathbb{R}}f(y)\,\pi(y)\,dy&=\int_{\mathbb{R}}f(\cdot)\,d\Big(\int_{\mathbb{R}}\mathbf{P}\{X(t+s)\in\cdot\,|\,X(s)=x\}\,\pi(x)\,dx\Big)\\&=\int_{\mathbb{R}}\Big(\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,\pi(x)\,dx\qquad\text{for}\quad t\geq 0\quad\text{and}\quad f\in\mathbb{C}_0(\mathbb{R})\end{aligned}$$

(recall Exercsie 112). Since the right-hand side is a bounded continuous function of $t\in[0,\infty)$, this in turn holds if the corresponding Laplace transforms agree

$$\frac{1}{\lambda}\int_{\mathbb{R}}f(y)\,\pi(y)\,dy=\int_0^\infty e^{-\lambda t}\Big[\int_{\mathbb{R}}\Big(\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,\pi(x)\,dx\Big]\,dt=\int_{\mathbb{R}}(G_\lambda f)(x)\,\pi(x)\,dx$$

for $\lambda>0$. From defining properties of the fundamental measure $\hat{P}$, we have

$$\begin{aligned}&\lambda\int_{\mathbb{R}}(G_\lambda f)(x)\,\pi(x)\,dx-\int_{\mathbb{R}}f(x)\,\pi(x)\,dx\\&=\int_{\mathbb{R}}\Big[\int_0^\infty\lambda\,e^{-\lambda t}\Big(\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,dt\Big]\,\pi(x)\,dx-\int_{\mathbb{R}}f(x)\,\pi(x)\,dx\\&=\int_{\mathbb{R}}\Big[e^{-\lambda t}\Big(\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\Big]_\infty^0\,\pi(x)\,dx-\int_{\mathbb{R}}f(x)\,\pi(x)\,dx\\&\qquad+\int_{\mathbb{R}}\Big[\int_0^\infty e^{-\lambda t}\Big(D_t\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,dt\Big]\,\pi(x)\,dx\\&=\int_{\mathbb{R}}\Big[\int_0^\infty e^{-\lambda t}\Big(\big(\mu(x)D_x+\tfrac{1}{2}\sigma(x)^2D_x^2\big)\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,dt\Big]\,\pi(x)\,dx\\&=\int_0^\infty e^{-\lambda t}\Big(\int_{\mathbb{R}}\Big(\big(\mu(x)D_x+\tfrac{1}{2}\sigma(x)^2D_x^2\big)\int_{\mathbb{R}}f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,\pi(x)\,dx\Big)\,dt\end{aligned}$$

$$= \int_0^\infty e^{-\lambda t}\Big[\int_{\mathbb{R}}\Big(\mu(x)D_x\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,\pi(x)\,dx\Big]\,dt$$
$$+\int_0^\infty e^{-\lambda t}\Big[\Big(\tfrac{1}{2}\sigma(x)^2 D_x\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,\pi(x)\Big]_{-\infty}^{\infty}\,dt$$
$$-\int_0^\infty e^{-\lambda t}\Big[\int_{\mathbb{R}}\Big(D_x\big(\tfrac{1}{2}\sigma(x)^2\pi(x)\big)\Big)\Big(D_x\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot,t,x)\Big)\,dx\Big]\,dt.$$

This is zero, by (25.1) together with the fact that $\mu(x)\pi(x)-D_x(\frac{1}{2}\sigma(x)^2\pi(x))=0$. □

EXERCISE 130 Explain to what extent the requirement that $\int_{\mathbb{R}} f(\cdot)\,d\hat{P}(\cdot,t,x)$ and its derivatives are bounded in Definition 22.7 is used in the proof of Theorem 25.8. How can it be relaxed, just by inspection of that proof?

Example 25.9 (LANGEVIN EQUATION) Pick constants $\alpha,\sigma\in\mathbb{R}$ and $\beta>0$, and consider the *Ornstein-Uhlenbeck process* X in (15.1). It has stationary density

$$\pi(y)=\frac{C}{\sigma^2}\exp\Big\{\int_0^y\frac{(\alpha-\beta z)\,dz}{\sigma^2/2}\Big\}=\frac{C}{\sigma^2}\exp\Big\{-\frac{(y-\alpha/\beta)^2-\alpha^2/\beta^2}{\sigma^2/\beta}\Big\}=f_{\mathrm{N}(\alpha/\beta,\,\frac{1}{2}\sigma^2/\beta)}(y)$$

for $\sigma\neq 0$ (by Theorem 25.8), and stationary distribution $\mathrm{N}(\alpha/\beta,0)$ for $\sigma=0$ [by (15.2)]. By Exercise 128, X is stationary iff. $X(0)$ has the stationary distribution.

Now start the process X in (15.1) with an initial value $X(0)$ that is independent of B, and has the stationary distribution. By (15.2), we have

$$X(t)=\alpha/\beta+e^{-\beta t}\Big(X(0)-\alpha/\beta+\int_0^t\sigma\,e^{\beta r}\,dB(r)\Big)\qquad\text{for}\quad t\geq 0.$$

Hence X is a Gaussian process [since a linear combination of process values is the sum of the Gaussian random variable $X(0)$ multiplied by a constant, and an independent Itô integral of a non-random function, which is the limit in probability of Gaussian distributed Riemann-Stieltjes type approximating sums]. The covariance function is

$$r(t-s)\equiv e^{-\beta(s+t)}\,\mathbf{Cov}\Big\{X(0)+\int_0^s\sigma\,e^{\beta r}\,dB(r),\,X(0)+\int_0^t\sigma\,e^{\beta r}\,dB(r)\Big\}$$
$$=e^{-\beta(s+t)}\Big(\mathbf{Var}\{X(0)\}+\int_0^{s\wedge t}\sigma^2\,e^{2\beta r}\,dr\Big)=\frac{\sigma^2}{2\,\beta}\,e^{-\beta|t-s|}.$$

Conversely, let $\{X(t)\}_{t\geq 0}$ be a stationary Gaussian Markov process with continuous covariance function $r(t)\equiv\mathbf{E}\{\hat{X}(s)\hat{X}(s+t)\}$, where $\hat{X}(t)=X(t)-m$ with $m=\mathbf{E}\{X(t)\}$. Since $r(0)\hat{X}(t+s)-r(t)\hat{X}(s)$ is independent of $X(s)$ (calculate the covariance and use Corollary 2.11), the Markov property shows that

$$r(s+t)r(0)=\mathbf{E}\{\hat{X}(0)\,r(0)\hat{X}(s+t)\}$$
$$=\mathbf{E}\Big\{\hat{X}(0)\,\mathbf{E}\Big\{\big(r(0)\hat{X}(s+t)-r(t)\hat{X}(s)\big)+r(t)\hat{X}(s)\,\Big|\,\mathcal{F}_s\Big\}\Big\}$$
$$=\mathbf{E}\Big\{\hat{X}(0)\,\mathbf{E}\Big\{\big(r(0)\hat{X}(s+t)-r(t)\hat{X}(s)\big)+r(t)\hat{X}(s)\,\Big|\,X(s)\Big\}\Big\}$$
$$=\mathbf{E}\Big\{\hat{X}(0)\,\big(0+r(t)\hat{X}(s)\big)\Big\}=r(t)r(s)\qquad\text{for}\quad s,t\geq 0.$$

Now Exercise 131 below gives $r(t) = r(0)\,\mathrm{e}^{-C|t|}$ for some constant $C \geq 0$. Thus *any stationary Gaussian Markov process with continuous covariance is an Ornstein-Uhlenbeck process*, since it must have transition probabilities as well as one-dimensional marginal distributions as such, by the form of r together with Theorem 2.10.

The Ornstein-Uhlenbeck process X in (15.1), started with the stationary distribution, has fundamental measure [since $X(t){-}r(t)X(0)/r(0)$ is independent of $X(0)$]

$$\begin{aligned}\hat{P}(\cdot,t,x) = \mathbf{P}\{X(t)\in\cdot\,|\,X(0)=x\} &= \mathbf{P}\Big\{X(t)-\frac{r(t)}{r(0)}X(0)\in\cdot-\frac{r(t)}{r(0)}x\Big\} \\ &= \mathbf{P}\Big\{\mathrm{N}\Big(\frac{\alpha}{\beta}(1-\mathrm{e}^{-\beta t})+\mathrm{e}^{-\beta t}x,\ \frac{\sigma^2}{2\beta}(1-\mathrm{e}^{-\beta t})\Big)\in\cdot\Big\}. \quad \#\end{aligned}$$

Exercise 131 Let $\{X(t)\}_{t\geq 0}$ be a Gaussian stationary process, with continuous covariance function $r(t) \equiv \mathbf{Cov}\{X(s),X(s{+}t)\}$ that satisfies $r(t{+}s)r(0) = r(t)r(s)$. Show that $r(t) = r(0)\,\mathrm{e}^{-C|t|}$ for some constant $C \geq 0$. (**Hint:** Prove that r is strictly positive and use the *Cauchy functional equation*.)

⋆Exercise 132 Try to calculate $\int_0^\infty \mathrm{e}^{-\lambda t}\,d\hat{P}(dy,t,x)/dy$ for the Ornstein-Uhlenbeck process in Example 25.9, to find an explicit expression for its resolvent.

26 Twentysixth Lecture

26.1 Transformation of One-Dimensional Diffusion Processes

Example 26.1 (TRANSFORMATION OF DIFFUSIONS) Consider the general one-dimensional time homogeneous diffusion type SDE

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t) \qquad \text{for} \quad t \geq 0, \qquad X(0) = x. \tag{26.1}$$

Taking $Y(t) = f(X(t))$ with $f \in \mathbb{C}^2(\mathbb{R})$, Itô's formula gives

$$df(X(t)) = f'(X(t))\Big(\mu(X(t))\,dt + \sigma(X(t))\,dB(t)\Big) + \frac{\sigma(X(t))^2}{2}\,f''(X(t))\,dt. \tag{26.2}$$

Now take $f'(x)\,\sigma(x) = 1$, which is to say that $f(x) = \int_0^x \sigma(y)^{-1}\,dy$, assuming that σ is strictly positive and in $\mathbb{C}^1(\mathbb{R})$ [so that $f \in \mathbb{C}^2(\mathbb{R})$ as required]. Writing $\tilde{\mu}(x) = \frac{\mu(f^{-1}(x))}{\sigma(f^{-1}(x))} - \frac{\sigma'(f^{-1}(x))}{2}$, we obtain

$$\begin{aligned} dY(t) &= \left(f'(X(t))\,\mu(X(t)) + \frac{\sigma(X(t))^2}{2}\,f''(X(t))\right) dt + dB(t) \\ &= \left(\frac{\mu(X(t))}{\sigma(X(t))} - \frac{\sigma'(X(t))}{2}\right) dt + dB(t) \\ &= \left(\frac{\mu(f^{-1}(Y(t)))}{\sigma(f^{-1}(Y(t)))} - \frac{\sigma'(f^{-1}(Y(t)))}{2}\right) dt + dB(t) = \tilde{\mu}(Y(t))\,dt + dB(t), \end{aligned}$$

with $Y(0) = f(x)$. Conversely, if Y is a diffusion given by

$$dY(t) = \tilde{\mu}(Y(t))\,dt + dB(t) \qquad \text{for} \quad t \geq 0, \qquad Y(0) = f(x), \tag{26.3}$$

then $X = f^{-1}(Y)$ solves the SDE (26.1), by a similar application of Itô's formula. #

Definition 26.2 *A time homogeneous one-dimensional SDE*

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t) \qquad \textit{for} \quad t \geq 0,$$

with coefficients $\mu, \sigma : \mathbb{R} \to \mathbb{R}$ *such that* μ/σ^2 *is locally integrable, has*

$$\underline{\textit{scale function}} \qquad \mathbb{R} \ni x \to \underline{p(x)} \equiv \int_0^x \exp\Big\{-\int_0^y \frac{2\,\mu(z)}{\sigma(z)^2}\,dz\Big\}\,dy \in (0, \infty).$$

Example 26.3 (REMOVAL OF DRIFT) Consider the one-dimensional time homogeneous SDE X in (26.1). Take $f \in \mathbb{C}^2(\mathbb{R})$ such that $f'(x)\,\mu(x) + \frac{1}{2} f''(x)\,\sigma(x)^2 = 0$. Assuming that μ/σ^2 is continuous, this holds for the scale function $f = p$. By (26.2), the process $Y(t) = p(X(t))$ satisfies the diffusion SDE with zero drift

$$dY(t) = \sigma(X(t))\,p'(X(t))\,dB(t) = \sigma(p^{-1}(Y(t)))\,\exp\Big\{-\int_0^{p^{-1}(Y(t))} \frac{2\,\mu(z)}{\sigma(z)^2}\,dz\Big\}\,dB(t). \ \#$$

Theorem 26.4 (Removal of Drift) *A time homogeneous one-dimensional SDE*

$$dX(t)=\mu(X(t))\,dt+\sigma(X(t))\,dB(t)\qquad \textit{for}\quad t\geq 0,$$

with scale function p, has a weak solution [strong solution] $\{X(t)\}_{t\geq 0}$, iff. $Y=p(X)$ is a weak solution [strong solution] to the SDE

$$dY(t)=\sigma(p^{-1}(Y(t)))\,\exp\Big\{-\int_0^{p^{-1}(Y(t))}\frac{2\,\mu(z)}{\sigma(z)^2}\,dz\Big\}\,dB(t)\qquad \textit{for}\quad t\geq 0.$$

Proof for $p\in\mathbb{C}^2(\mathbb{R})$. This we have directly from Example 26.3. □

*Remark 26.5** The proof of Theorem 26.4 in Example 26.3 is an application of *Itô's formula* Theorem 14.4 to $p(X)$, which requires $p\in\mathbb{C}^2(\mathbb{R})$. There is an extension of *Itô's formula* to convex functions. This extension uses the <u>local time</u>

$$L^x(t)\equiv|X(t)-x|-|X(0)-x|-\int_0^t \mathrm{sign}_-(X(r)-x)\,dX(r)\qquad \text{for}\quad t\geq 0\quad\text{and}\quad x\in\mathbb{R},$$

of the diffusion (or continuous semimartingale) X. The extended *Itô's formula* reads

$$f(X(t))=f(X(0))+\int_0^t f'_-(X(r))\,dX(r)+\frac{1}{2}\int_{\mathbb{R}}L^x(t)\,d\mu_f(x)\qquad\text{for}\quad t\geq 0,$$

when $f:\mathbb{R}\to\mathbb{R}$ is convex, with increasing and left-continuous left-derivative f'_- and Stieltje measure "second derivative" $\mu_f([x,y))=f'_-(y)-f'_-(x)$ for $[x,y)\subseteq\mathbb{R}$ (e.g., [21, Theorem 19.5]). For a measurable $g:\mathbb{R}\to\mathbb{R}$, we further have

$$\int_0^t g(X(r))\,d[X](r)=\int_{\mathbb{R}}g(x)\,L^x(t)\,dx\qquad\text{for}\quad t\geq 0\tag{26.4}$$

(the integrals are well-defined simultaneously, and coincide when well-defined).

In the particular case when X is BM with $X(0)=0$, we have (e.g., [22, Eq. 3.6.2])

$$L^x(t)=\lim_{\varepsilon\downarrow 0}\frac{1}{4\varepsilon}\int_0^t I_{[0,\varepsilon]}(|X(r)-x|)\,dr\quad\text{a.s.}\tag{26.5}$$

This is what is required for a rigorous treatment of the SDE in Example 17.5

$$dX(t)=\mathrm{sign}(X(t))\,dB(t)\qquad\text{for}\quad t\geq 0,\qquad X(0)=0:$$

Take $f=|\cdot|$, so that $f'_-=\mathrm{sign}_-$ and $\mu_f=2\delta$. By *Itô's formula*, we have

$$\begin{aligned}|X(t)|&=|X(0)|+\int_0^t\mathrm{sign}_-(X(r))\,dX(r)+L^0(t)\\&=|X(0)|+\int_0^t(\mathrm{sign}_-(X(r)))\,(\mathrm{sign}(X(r)))\,dB(r)+\lim_{\varepsilon\downarrow 0}\frac{1}{4\varepsilon}\int_0^t I_{[0,\varepsilon]}(|X(r)|)\,dr,\end{aligned}$$

when X is a strong solution to the SDE, since this implies that X is BM, by Exercise 95 (or by Remark 17.6). Rearranging, it follows that $B(t)$ is adapted to $\sigma(|X(s)|:s\leq t)$, as is crucial in Example 17.5, because

$$\int_0^t \mathrm{sign}_-(X(r))\,dX(r) = B(t) - \int_0^t I_{\{0\}}(X(r))\,dB(r) = B(t),$$

by Corollary 23.19, since (26.5) together with *Dominated Convergence* give

$$\Big[\int_0^\cdot I_{\{0\}}(X)\,dB\Big](t) = \int_0^t I_{\{0\}}(X(r))\,dr \leq \int_0^t I_{[-\varepsilon,\varepsilon]}(X(r))\,dr \to 0 \qquad \text{as} \quad \varepsilon\downarrow 0.$$

In the particular case when $f\in\mathbb{C}^1(\mathbb{R})$, with absolutely continuous derivative f' with derivative (density) f'', (26.4) shows that *Itô's formula* is the "usual"

$$f(X(t)) = f(X(0)) + \int_0^t f'(X(r))\,dX(r) + \frac{1}{2}\int_0^t f''(X(r))\,d[X](r) \qquad \text{for} \quad t\geq 0.$$

In fact, this holds for general (not necessarily convex) $f\in\mathbb{C}^1(\mathbb{R})$ with f' absolutely continuous, because for such f we may write

$$f(x) = \int_0^x f'(y)\,dy = f_1(x) - f_2(x) = \int_0^x F_1(y)\,dy - \int_0^x F_2(y)\,dy \qquad \text{for} \quad x\in\mathbb{R},$$

where F_1 and F_2 are increasing functions, so that f_1 and f_2 are convex, by an easy argument. Using *Itô's formula* separately on the convex components f_1 and f_2 of f, and adding things up (or subtracting, rather), we get *Itô's formula* for f itself. It is this *Itô's formula* which is required for a complete proof of Theorem 26.4, because the scale p is in $\mathbb{C}^1(\mathbb{R})$ with p' absolutely continuous. #

Corollary 26.6 (ENGELBERT-SCHMIDT) *A time homogeneous one-dimensional SDE*

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t) \qquad \textit{for} \quad t\geq 0, \qquad X(0)=X_0,$$

such that σ is strictly positive, and μ/σ^2 and $1/\sigma^2$ locally integrable, [i.e.,

$$\int_B \frac{|\mu(y)|+1}{\sigma(y)^2}\,dy < \infty \qquad \text{for each bounded} \quad B\in\mathcal{B}(\mathbb{R}),]$$

has a weak solution that in addition is unique for every choice of initial value X_0.

Proof. Writing p for the scale, by Theorem 26.4, it is enough to show that

$$dY(t) = \sigma(p^{-1}(Y(t)))\,p'(p^{-1}(Y(t)))\,dB(t) \qquad \text{for} \quad t\geq 0, \qquad Y(0) = p(X_0)\equiv Y_0,$$

has a weak solution that is unique for every initial value Y_0. By earlier *Engelbert-Schmidt theorem* (Theorem 19.8), this holds if (iff.) we have

$$\int_{-\varepsilon}^{\varepsilon} \frac{dy}{\sigma(p^{-1}(x+y))^2\,p'(p^{-1}(x+y))^2} = \int_{p^{-1}(x-\varepsilon)}^{p^{-1}(x+\varepsilon)} \frac{dy}{\sigma(y)^2\,p'(y)} = \infty \qquad \text{for all} \quad \varepsilon>0$$

precisely when $\sigma(p^{-1}(x))\,p'(p^{-1}(x))=0$, for each choice $x\in\mathbb{R}$. However, since p' is locally bounded away from zero, and σ strictly positive and continuous, the integral is never infinite, and σ never zero, and so the condition holds trivially. □

Example 26.7 (Brownian Bridge) Consider the one-dimensional non-homogeneous SDE, with $\mu(t,x)=-x/(1-t)$ and $\sigma(t,x)=1$,

$$d\overline{B}(t)=-\overline{B}(t)/(1-t)\,dt+dB(t)\qquad\text{for}\quad t\in[0,1],\qquad \overline{B}(0)=0.$$

A strong solution $\{\overline{B}(t)\}_{t\in[0,1]}$ to this SDE is given by

$$\overline{B}(t)=(1-t)\int_0^t\frac{1}{1-r}\,dB(r)=\int_0^t\frac{1-t}{1-r}\,dB(r)\qquad\text{for}\quad t\in[0,1],$$

because writing $Y(t)=\int_0^t\frac{1}{1-r}\,dB(r)$, Itô's formula [with $f(t,y)=(1-t)\,y$] gives

$$d\overline{B}(t)=df(t,Y(t))=-Y(t)\,dt+(1-t)\,dY(t)=-\overline{B}(t)/(1-t)\,dt+dB(t).$$

The solution is unique in each interval $[0,1-1/n]$, $n\in\mathbb{N}$, by Theorem 16.8. Hence it is unique in $\cup_{n=1}^{\infty}[0,1-1/n]=[0,1)$, and thus in $[0,1]$, by continuity. Further, X is a zero-mean Gaussian stochastic process (cf. Example 25.9), with covariance function

$$\begin{aligned}\mathbf{Cov}\{\overline{B}(s),\overline{B}(t)\}=\mathbf{E}\Big\{\Big(\int_0^s\frac{1-s}{1-u}\,dB(u)\Big)\Big(\int_0^t\frac{1-t}{1-v}\,dB(v)\Big)\Big\}&=\int_0^{s\wedge t}\frac{1-s}{1-r}\,\frac{1-t}{1-r}\,dr\\&=(s\wedge t)-st\end{aligned}$$

for $s,t\in[0,1]$. Notice the important fact (readily established by means of comparing the covariance functions), that $\{\overline{B}(t)\}_{t\in[0,1]}=_{\text{same fidi's}}\{B^0(t)-tB^0(1)\}_{t\in[0,1]}$. #

We finish the notes with a general formula for the transition density function of a one-dimensional time homogeneous diffusion from [16, p. 98, Eq. 9]. We have found this result very useful in [2]*. (Now it seems that one sees less and less of such quite analytic results, and the purely probabilistic is "modern".) We give a probabilistic proof, that uses virtually every peace of stochastic calculus we have learned.

Theorem 26.8 *Consider the general one-dimensional time homogeneous diffusion type SDE* (26.1)*, with $\mu\in\mathbb{C}^1(\mathbb{R})$ satisfying a global Lipschitz condition, and $\sigma:\mathbb{R}\to(0,\infty)$ two times continuously differentiable* (slightly weaker conditions work)*. Let*

$$G(y)\equiv\Big(\frac{\mu(f^{-1}(y))}{\sigma(f^{-1}(y))}-\frac{\sigma'(f^{-1}(y))}{2}\Big)^2+\frac{d}{dy}\Big(\frac{\mu(f^{-1}(y))}{\sigma(f^{-1}(y))}-\frac{\sigma'(f^{-1}(y))}{2}\Big),$$

where $f(y)\equiv\int_0^y\sigma(z)^{-1}\,dz$. The diffusion has transition density function

$$\begin{aligned}p(y,t,x,s)=\frac{\sqrt{\sigma(x)}}{\sqrt{2\pi t\,\sigma(y)^3}}\exp\Big\{-\frac{(f(y)-f(x))^2}{2t}+\int_x^y\frac{\mu(z)\,dz}{\sigma(z)^2}\Big\}\\\times\mathbf{E}\Big\{\exp\Big[-\frac{t}{2}\int_0^1 G\Big(r\,(f(y)-f(x))+f(x)+\sqrt{t}\,\overline{B}(r)\Big)\,dr\Big]\Big\}.\end{aligned}$$

EXERCISE 133 Calculate the transition density for the Ornstein-Uhlenbeck process in Example 25.9, by means of the formula in Theorem 26.8.

Proof of Theorem 26.8 (for t small). Taking $\tilde{\mu}(x)=\frac{\mu(f^{-1}(x))}{\sigma(f^{-1}(x))}-\frac{\sigma'(f^{-1}(x))}{2}$, we have

$$\begin{aligned}\int_x^{f^{-1}(y)}\frac{\mu(z)\,dz}{\sigma(z)^2}=\int_{f(x)}^{y}\frac{\mu(f^{-1}(z))\,dz}{\sigma(f^{-1}(z))}&=\int_{f(x)}^{y}\tilde{\mu}(z)\,dz+\int_{f(x)}^{y}\frac{\sigma'(f^{-1}(z))}{2}\,dz\\&=\int_{f(x)}^{y}\tilde{\mu}(z)\,dz+\int_x^{f^{-1}(y)}\frac{\sigma'(z)}{2\,\sigma(z)}\,dz\\&=\int_{f(x)}^{y}\tilde{\mu}(z)\,dz+\ln\Bigg(\sqrt{\frac{\sigma(f^{-1}(y))}{\sigma(x)}}\Bigg).\end{aligned}$$

Hence, by Example 26.1, X solves the SDE (26.1), and has the indicated transition density $p(y,t,x,s)$, iff. $Y=f(X)$ solves the SDE (26.3), and has transition density

$$\begin{aligned}\big(D_yf^{-1}(y)\big)\,p(f^{-1}(y),t,x,s)&=\frac{\sqrt{\sigma(x)}}{\sqrt{2\pi t\,\sigma(f^{-1}(y))}}\exp\Big\{-\frac{(y-f(x))^2}{2t}+\int_x^{f^{-1}(y)}\frac{\mu(z)\,dz}{\sigma(z)^2}\Big\}\\&\quad\times\mathbf{E}\Big\{\exp\Big[-\frac{t}{2}\int_0^1G\Big(r\,(y-f(x))+f(x)+\sqrt{t}\,\overline{B}(r)\Big)\,dr\Big]\Big\}\\&=\frac{1}{\sqrt{2\pi t}}\exp\Big\{-\frac{(y-f(x))^2}{2t}+\int_{f(x)}^{y}\tilde{\mu}(z)\,dz\Big\}\\&\quad\times\mathbf{E}\Big\{\exp\Big[-\frac{t}{2}\int_0^1\tilde{G}\Big(r\,(y-f(x))+f(x)+\sqrt{t}\,\overline{B}(r)\Big)\,dr\Big]\Big\}.\end{aligned}$$

Here $\tilde{G}(y)=\tilde{\mu}(y)^2+\tilde{\mu}'(y)$ is the function G in Theorem 26.8, evaluated for the SDE (26.3). Notice that $\tilde{\mu}\in\mathbb{C}^1(\mathbb{R})$, by assumptions imposed on μ and σ. From this we conclude that it is enough to prove the theorem in the particular case when $\sigma(x)=1$, which is the one adressed subsequently:

The generator $\mathcal{A}=\frac{1}{2}D_x^2+\mu(x)D_x$ obviously is strongly elliptic. By *Engelbert-Schmidt theorem* (Corollary 26.6) together with Section 20.1, the martingale problem associated with $\mathcal{A}$ is well-posed. By Remark 22.5 together with Theorem 22.6, a transition density exists, and is given by $p(y,t,x,s)=\hat{p}(y,t{+}s,x,s)$, where $\hat{p}(y,\tau,x,s)$ is a fundamental solution to the PDO $D_t{+}\mathcal{A}$. Further, the *Feynman-Kac formula* is in play. This means that the solution to the Cauchy problem

$$\frac{\partial u(t,x)}{\partial t}+\mathcal{A}u(x)=0\qquad\text{for}\quad(t,x)\in(0,\tau)\times\mathbb{R},\qquad u(\tau,x)=f(x)$$

can be represented in the following two alternative ways

$$u(t,x)=\mathbf{E}\{f(X(\tau-t))\}=\int_{\mathbb{R}}p(y,\tau-t,x,s)\,f(y)\,dy\qquad\text{for}\quad f\in\mathbb{C}_0(\mathbb{R})$$

[where X is a solution to the SDE (26.1), with $\sigma(x)=1$]. By *Girsanov's Theorem* (Theorem 19.1) and Example 18.4, we have that BM B^x solves the SDE (26.1), for $t\in[0,T]$, under the probability measure

$$\tilde{\mathbf{P}}\{A\} = \mathbf{E}\bigg\{I_A \exp\bigg\{\int_0^T \mu(B(r))\,dB(r) - \frac{1}{2}\int_0^T \mu(B(r))^2\,dr\bigg\}\bigg\}.$$

From this, together with *Baye's rule* (Lemma 19.2), we may now conclude that

$$\begin{aligned}\int_{\mathbb{R}} p(y,t,x,s)\,f(y)\,dy &= \mathbf{E}\bigg\{f(B^x(t))\exp\bigg\{\int_0^T \mu(B^x(r))\,dB(r) - \frac{1}{2}\int_0^T \mu(B^x(r))^2\,dr\bigg\}\bigg\}\\ &= \mathbf{E}\bigg\{f(B^x(t))\exp\bigg\{\int_0^t \mu(B^x(r))\,dB(r) - \frac{1}{2}\int_0^t \mu(B^x(r))^2\,dr\bigg\}\bigg\}.\end{aligned}$$

Writing $U(x) = \int_0^x \mu(y)\,dy$, *Itô's formula* shows that

$$dU(B^x(t)) = \mu(B^x(t))\,dB^x(t) + (1/2)\,\mu'(B^x(t))\,dt,$$

so that

$$U(B^x(t)) = U(x) + \int_0^t \mu(B^x(r))\,dB^x(r) + \frac{1}{2}\int_0^t \mu'(B^x(r))\,dr.$$

Inserting in above findings, this gives that

$$\begin{aligned}&\int_{\mathbb{R}} p(y,t,x,s)\,f(y)\,dy\\ &= \mathbf{E}\bigg\{f(B^x(t))\exp\bigg\{U(B^x(t)) - U(x) - \frac{1}{2}\int_0^t \mu(B^x(r))^2\,dr - \frac{1}{2}\int_0^t \mu'(B^x(r))\,dr\bigg\}\bigg\}\\ &= \mathbf{E}\bigg\{f(B^x(t))\exp\bigg\{U(B^x(t)) - U(x) - \frac{1}{2}\int_0^t G(B^x(r))\,dr\bigg\}\bigg\}.\end{aligned}$$

Conditioning on the value of $B^x(t)$, and noticing that $B^x(t)$ is independent of $B^x(r) - (r/t)B^x(t)$, this readily becomes

$$\begin{aligned}&\int_{\mathbb{R}} p(y,t,x,s)\,f(y)\,dy\\ &= \int_{\mathbb{R}} f(y)\,\mathbf{E}\bigg\{\exp\bigg\{U(y) - U(x) - \frac{1}{2}\int_0^t G\Big(B^x(r) - (r/t)B^x(t) + (r/t)y\Big)\,dr\bigg\}\bigg\}p_t(x,y)\,dy\\ &= \int_{\mathbb{R}} f(y)\,\mathbf{E}\bigg\{\exp\bigg\{\int_x^y \mu(z)\,dz - \frac{t}{2}\int_0^1 G\Big(B^0(\hat{r}t) - \hat{r}B^0(t) + (1-\hat{r})x + \hat{r}y\Big)\,d\hat{r}\bigg\}\bigg\}p_t(x,y)\,dy\\ &= \int_{\mathbb{R}} f(y)\,\mathbf{E}\bigg\{\exp\bigg\{\int_x^y \mu(z)\,dz - \frac{t}{2}\int_0^1 G\Big(\sqrt{t}\,(B^0(\hat{r}) - \hat{r}B^0(1)) + (1-\hat{r})x + \hat{r}y\Big)\,d\hat{r}\bigg\}\bigg\}p_t(x,y)\,dy\end{aligned}$$

[using self-similarity of B^0 (Exercise 36) in the last step]. Hence Example 26.7 gives

$$p(y,t,x,s) = \exp\bigg\{\int_x^y \mu(z)\,dz\bigg\}\,\mathbf{E}\bigg\{\exp\bigg\{-\frac{t}{2}\int_0^1 G\Big(\sqrt{t}\,\overline{B}(\hat{r}) + (1-\hat{r})x + \hat{r}y\Big)\,d\hat{r}\bigg\}\bigg\}p_t(x,y),$$

which is the expression for the transition density in the theorem [when $\sigma(x) = 1$]. □

EXERCISE 134 In the proof of Theorem 26.8 we did not check that the process Z in *Girsanov's Theorem* really is a martingale: Show how *Novikov's Criterion* (Corollary 18.7), together with the *Reflection Principle* for BM, give this for $t \in [0,T]$ with $T > 0$ sufficiently small, using that $|\mu(x) - \mu(B^x(t))|^2 \le K^2 B^0(t)^2 \le K^2 \sup_{t\in T} B^0(t)^2$ for $t \in [0,T]$, for some constant $K > 0$ (by Lipschitz continuity of μ).

A *Appendix. An Elementary Construction of BM

Theorem A.1 *For each constant $x\in\mathbb{R}$, BM B^x with $B^x(0)=x$ exist.*

Proof. It is enough to show that $B=B^0$ exists, because then $B+x$ will do as B^x. For this, it is enough to show that there exists a zero-mean Gaussian stochastic process B, with $\mathbf{Cov}\{B(s),B(t)\} = s\wedge t$, that is continuous a.s. Because then we have $\mathbf{Cov}\{B(r),B(t)-B(s)\} = 0$ for $0\le r\le s\le t$, so that increments are uncorrelated with earlier process values, and thus independent of them, by Corollary 2.11. Further, $B(t)-B(s)$ will be $\mathrm{N}(0,t-s)$-distributed, as required.

Let $\xi_1,\xi_2,\ldots$ be independent $\mathrm{N}(0,1)$-distributed random variables, and set

$$B(t) \equiv \underset{n\to\infty}{\mathrm{l.i.m.}} \sum_{k=0}^{n} \frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi t/2)\,\xi_k \qquad \text{for} \quad t\in[0,1]. \tag{A.1}$$

By the Cauchy criterion for mean-square convergence, this limit is well-defined iff.

$$\mathbf{E}\bigg\{\bigg(\sum_{k=0}^{m}\frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi t/2)\,\xi_k - \sum_{\ell=0}^{n}\frac{\sqrt{2}}{\pi}\,\frac{2}{2\ell+1}\,\sin((2\ell+1)\pi t/2)\,\xi_\ell\bigg)^2\bigg\} \to 0$$

as $m,n\to\infty$. However, this holds, since the mean on the left-hand side is

$$\mathbf{E}\bigg\{\bigg(\sum_{k=(m\wedge n)+1}^{(m\vee n)}\frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi t/2)\,\xi_k\bigg)^2\bigg\} = \sum_{k=(m\wedge n)+1}^{(m\vee n)}\frac{8\,\sin((2k+1)\pi t/2)^2}{\pi^2(2k+1)^2}.$$

By symmetry in (A.1), we have $\mathbf{E}\{B(t)\}=0$. For the covariance, we notice that

$$\sum_{k=0}^{\infty}\frac{4}{\pi^2\,(2k+1)^2}\cos((2k+1)\pi t/2) = (1-|t|)/2 \qquad \text{for} \quad t\in[-2,2]: \tag{A.2}$$

By symmetry, it is enough to show (A.2) for $t\in[0,2]$. For such t, (A.2) holds since the left-hand side and right-hand side of (A.2) are continuous functions of t (by basic math), and, according to `Mathematica`, their one-sided Laplace transforms coincide:

```
In[1]:= Simplify[Sum[Integrate[4/(Pi*(2*k+1))^2*Cos[(2*k+1)*Pi*t/2]
          *Exp[-x*t], {t, 0, 2}], {k, 0, Infinity}]]
```

Out[1]= $\dfrac{e^{-2x}\left(1+e^{2x}(-1+x)+x\right)}{2x^2}$

```
In[2]:= Simplify[Integrate[(1/2-t/2)*Exp[-x*t], {t, 0, 2}]]
```

Out[2]= $\dfrac{e^{-2x}\left(1+e^{2x}(-1+x)+x\right)}{2x^2}$

Since $\mathbf{Cov}\{\cdot,\cdot\}$ commutes with mean-square limits, we have

$$\begin{aligned}&\mathbf{Cov}\{B(s),B(t)\}\\ &= \mathbf{Cov}\bigg\{\sum_{k=0}^{\infty}\frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi s/2)\,\xi_k,\ \sum_{\ell=0}^{\infty}\frac{\sqrt{2}}{\pi}\,\frac{2}{2\ell+1}\,\sin((2\ell+1)\pi t/2)\,\xi_\ell\bigg\}\end{aligned}$$

$$= \sum_{k=0}^{\infty} \frac{8\,\sin((2k+1)\pi s/2)\,\sin((2k+1)\pi t/2)}{\pi^2\,(2k+1)^2}$$

$$= \sum_{k=0}^{\infty} \frac{4}{\pi^2\,(2k+1)^2}\Big(\cos((2k+1)\pi\,(s-t)/2) - \cos((2k+1)\pi\,(s+t)/2)\Big),$$

by the identity $2\sin(x)\sin(y) = \cos(x-y) - \cos(x+y)$. Hence (A.2) gives

$$\mathbf{Cov}\{B(s),B(t)\} = (1-|t-s|)/2 - (1-|t+s|)/2 = s\wedge t \qquad \text{for} \quad s,t\in[0,1].$$

Moreover, B is Gaussian, since each linear combination of process values is a mean-square limit of a sequence of univariate Gaussian random variables

$$\sum_{i=1}^{n} a_i B(t_i) \leftarrow \sum_{k=0}^{m} \frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\left(\sum_{i=1}^{n} a_i \sin((2k+1)\pi\,t_i/2)\right)\xi_k \qquad \text{as} \quad m\to\infty$$

for $a_1,\ldots,a_n\in\mathbb{R}$ and $n\in\mathbb{N}$, so that the limit is also univariate Gaussian.

Finally, to prove that B is continuous with probability 1, we notice that

$$\mathbf{P}\left\{\sum_{k=0}^{\infty} \frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi\,t/2)\,\xi_k \;\text{ is continuous for }\; t\in[0,1]\right\}$$

$$\geq \mathbf{P}\left\{\sum_{n=0}^{\infty}\sum_{k=2^n-1}^{2^{n+1}-2} \frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi\,t/2)\,\xi_k \;\text{ converges uniformly for }\; t\in[0,1]\right\}$$

$$\geq 1 - \mathbf{P}\Big\{\sup\nolimits_{t\in[0,1]} |X_n(t)| > 2^{-n/8} \;\text{ for infinitely many }\; n\Big\}, \tag{A.3}$$

where X_n is the zero-mean Gaussian process given by

$$X_n(t) = \sum_{k=2^n-1}^{2^{n+1}-2} \frac{\sqrt{2}}{\pi}\,\frac{2}{2k+1}\,\sin((2k+1)\pi\,t/2)\,\xi_k \qquad \text{for} \quad t\in[0,1].$$

By the elementary identity $\sin(x)-\sin(y) = 2\cos(\frac{x+y}{2})\sin(\frac{x-y}{2})$, together with the fact that $|\sin(x)| \leq |x|$, we readily obtain

$$\begin{aligned}
\mathbf{E}\{(X_n(t)-X_n(s))^2\} &= \sum_{k=2^n-1}^{2^{n+1}-2} \frac{32\,\cos((2k+1)\pi\,(t+s)/4)^2\,\sin((2k+1)\pi\,(t-s)/4)^2}{\pi^2(2k+1)^2}\\
&\leq \sum_{k=2^n-1}^{2^{n+1}-2} \frac{16\,|t-s|^{1/2}}{\pi^{3/2}(2k+1)^{3/2}}\\
&\leq \frac{16\,2^n\,|t-s|^{1/2}}{\pi^{3/2}\,(2^{n+1}-1)^{3/2}}.
\end{aligned}$$

Using that X_n is continuous and symmetric, with $X_n(0)=0$, it follows that

$$\begin{aligned}
s_n &\equiv \mathbf{P}\{\sup\nolimits_{t\in[0,1]} |X_n(t)| > 2^{-n/8}\}\\
&\leq 2\,\mathbf{P}\left\{\bigcup_{k=0}^{\infty}\bigcup_{\ell=0}^{2^k-1}\{X_n(2^{-k}\ell) > 2^{-n/8}\}\right\}\\
&\leq 2\,\mathbf{P}\left\{\bigcup_{k=0}^{\infty}\bigcup_{\ell=0}^{2^k-1}\Big\{X_n(2^{-k}\ell) > 2^{-n/8-1}\Big(1+(1-2^{-1/8})\textstyle\sum_{j=0}^{k} 2^{-j/8}\Big)\Big\}\right\}\\
&= 2\,\mathbf{P}\Big\{X_n(0) > 2^{-n/8-1}\Big(1+(1-2^{-1/8})\Big)\Big\}
\end{aligned}$$

$$+2\sum_{k=1}^{\infty}\mathbf{P}\left\{\bigcup_{\ell=0}^{2^k-1}\left\{X_n(2^{-k}\ell)>2^{-n/8-1}\Big(1+(1-2^{-1/8})\textstyle\sum_{j=0}^{k}2^{-j/8}\Big)\right\},\right.$$

$$\left.\bigcap_{m=0}^{k-1}\bigcup_{\ell=0}^{2^m-1}\left\{X_n(2^{-m}\ell)\le 2^{-n/8-1}\Big(1+(1-2^{-1/8})\textstyle\sum_{j=0}^{m}2^{-j/8}\Big)\right\}\right\}$$

$$\le 0+2\sum_{k=1}^{\infty}\sum_{\ell=0}^{2^{k-1}-1}\mathbf{P}\Big\{X_n(2^{-k}(2\ell+1))>2^{-n/8-1}\Big(1+(1-2^{-1/8})\textstyle\sum_{j=0}^{k}2^{-j/8}\Big),$$

$$X_n(2^{-k+1}\ell)\le 2^{-n/8-1}\Big(1+(1-2^{-1/8})\textstyle\sum_{j=0}^{k-1}2^{-j/8}\Big)\Big\}$$

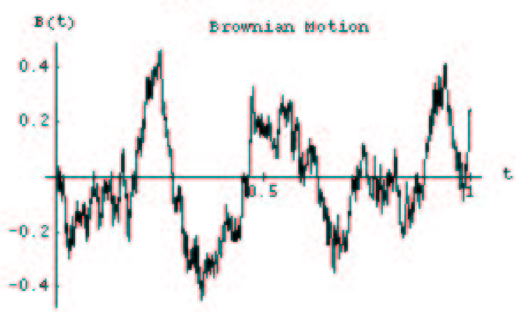
$$\le 2\sum_{k=1}^{\infty}\sum_{\ell=0}^{2^{k-1}-1}\mathbf{P}\Big\{X_n(2^{-k}(2\ell+1))-X_n(2^{-k}2\ell)>2^{-n/8-1}(1-2^{-1/8})2^{-k/8}\Big\}$$

$$=2\sum_{k=1}^{\infty}\sum_{\ell=0}^{2^{k-1}-1}\mathbf{P}\left\{\mathrm{N}(0,1)>\frac{2^{-n/8-1}(1-2^{-1/8})2^{-k/8}}{\sqrt{\mathbf{E}\{(X_n(2^{-k}(2\ell+1))-X_n(2^{-k}2\ell))^2\}}}\right\}$$

$$\le\sum_{k=1}^{\infty}2^k\,\mathbf{P}\left\{\mathrm{N}(0,1)>\frac{2^{-n/8-1}(1-2^{-1/8})2^{-k/8}\,\pi^{3/4}\,(2^{n+1}-1)^{3/4}}{4\,2^{n/2}\,2^{-k/4}}\right\}.$$

Since $\sum_{n=0}^{\infty}s_n<\infty$, the right-hand side of (A.3) is 1 by the Borel-Cantelli lemma.

```
In[1]:= << Statistics`ContinuousDistributions`

In[2]:= xi = N[Table[2*Sqrt[2]/Pi*Random[NormalDistribution[0, 1]], {2000}]];

In[3]:= B = Table[Sum[xi[[k]]*Sin[(2*k+1)*Pi*1/4000]/(2*k+1), {k, 1, 2000}], {1, 1, 2000}];

In[4]:= ListPlot[B, Ticks -> {{{1000, "0.5", 0.02}, {2000, "1", 0.02}}, Automatic},
   AxesLabel -> {"t", "B(t)"}, PlotLabel -> "      Brownian Motion", PlotJoined -> True]
```



Out[4]= - Graphics -

*Remark A.2 By the theory of Banach space-valued Gaussian random elements [$C([0,1])$ in this case], a zero-mean continuous Gaussian stochastic process X, on the unit-interval say, can be represented $X(t)=\sum_{k=1}^{\infty}f_k(t)\,\xi_k$, with uniform convergence and in mean-square, for some continuous functions $\{f_k\}_{k=1}^{\infty}$ and independent $\mathrm{N}(0,1)$-distributed random variables $\{\xi_k\}_{k=1}^{\infty}$. Switching to regard X as a $\mathbb{L}^2([0,1])$-valued Gaussian random element, $\{f_k\}_{k=1}^{\infty}$ can be chosen as $f_k=\sqrt{\lambda_k}\,e_k$, where $\{f_k\}_{k=1}^{\infty}$ are eigen-functions of the covariance operator $\mathbb{L}^2([0,1])\ni f\to\int_0^1\mathbf{E}\{X(\cdot)X(r)\}\,f(r)\,dr\in\mathbb{L}^2([0,1])$, with corresponding eigen-values $\{\lambda_k\}_{k=1}^{\infty}$. In the case of BM, this becomes (A.1) (e.g., [1, Sections III.2-III.3]). #

B ⋆Appendix. An Elementary Construction of PP

Proof of Theorem 4.4. It is enough to show that $N = N^0$ is a PP started at zero, because then $N+x$ will do as N^x. Since N is right-continuous by construction, for this in turn, it is enough to prove that N has `fidi's`

$$\mathbf{P}\Big\{\bigcap_{i=1}^{n}\{N(t_i)=k_i\}\Big\} = \frac{(\lambda t_1)^{k_1}}{k_1!}\bigg(\prod_{i=2}^{n}\frac{(\lambda(t_i-t_{i-1}))^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}\bigg)\,\mathrm{e}^{-\lambda t_n} \tag{B.1}$$

for $k_1\leq\ldots\leq k_n$ in $\mathbb{N}$ and $0<t_1<\ldots<t_n$. Because this gives

$$\begin{aligned}
&\mathbf{P}\Big\{\bigcap_{i=1}^{n}\{N(r_i)=k_i\},\, N(t)-N(s)=\ell\Big\}\\
&\quad=\sum_{k=k_n}^{\infty}\mathbf{P}\Big\{\bigcap_{i=1}^{n}\{N(r_i)=k_i\},\, N(s)=k,\, N(t)=\ell+k\Big\}\\
&\quad=\sum_{k=k_n}^{\infty}\frac{(\lambda r_1)^{k_1}}{k_1!}\bigg(\prod_{i=2}^{n}\frac{(\lambda(r_i-r_{i-1}))^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}\bigg)\frac{(\lambda(s-r_n))^{k-k_n}}{(k-k_n)!}\,\frac{(\lambda(t-s))^{\ell}}{\ell!}\,\mathrm{e}^{-\lambda t}\\
&\quad=\frac{(\lambda r_1)^{k_1}}{k_1!}\bigg(\prod_{i=2}^{n}\frac{(\lambda(r_i-r_{i-1}))^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}\bigg)\,\mathrm{e}^{-\lambda r_n}\,\frac{(\lambda(t-s))^{\ell}}{\ell!}\,\mathrm{e}^{-\lambda(t-s)}\\
&\quad=\mathbf{P}\Big\{\bigcap_{i=1}^{n}\{N(r_i)=k_i\}\Big\}\,\mathbf{P}\{\mathrm{Po}(\lambda(t-s))=\ell\}
\end{aligned}$$

for $0\leq k_1\leq\ldots\leq k_n$, $\ell\geq 1$ and $0\leq r_1\leq\ldots\leq r_n<s<t$, where (sum over all k_i's on both sides) $\mathbf{P}\{\mathrm{Po}(\lambda(t-s))=\ell\} = \mathbf{P}\{N(t)-N(s)=\ell\}$, so that increments are independenent (recall Theorem 1.15), stationary, and Poisson distributed, as required.

Write $T_k=\sum_{i=1}^{k}\xi_i$, and recall that T_k is gamma(k,λ)-distributed, so that

$$\mathbf{P}\{T_k>t\} = \sum_{i=0}^{k-1}\frac{(\lambda t)^i}{i!}\,\mathrm{e}^{-\lambda t} \qquad\text{and}\qquad f_k(t) = \frac{d}{dt}\mathbf{P}\{T_k\leq t\} = \frac{\lambda^k t^{k-1}}{(k-1)!}\,\mathrm{e}^{-\lambda t}$$

for $t\geq 0$ and $k\geq 1$. From this we get (B.1) for $n=1$, since

$$\mathbf{P}\{N(t)=k\} = \mathbf{P}\{N(t)<k+1\}-\mathbf{P}\{N(t)<k\} = \mathbf{P}\{T_{k+1}>t\}-\mathbf{P}\{T_k>t\} = \frac{(\lambda t)^k}{k!}\,\mathrm{e}^{-\lambda t}.$$

Notice that, having dealt with $n=1$ above, (B.1) is equivalent with

$$\mathbf{P}\Big\{\bigcap_{i=1}^{n}\{N(t_i)=k_i\}\Big\} = \mathbf{P}\{N(t_1)=k_1\}\prod_{i=2}^{n}\mathbf{P}\{N(t_i-t_{i-1})=k_i-k_{i-1}\} \tag{B.2}$$

for $k_1\leq\ldots\leq k_n$ and $0<t_1<\ldots<t_n$. Assume now that (B.2) holds for $n=\hat{n}-1\geq 1$, and consider $n=\hat{n}$. To prove (B.2) for $n=\hat{n}$, is the same thing as proving that

$$\begin{aligned}
&\mathbf{P}\Big\{N(t_1)\geq k_1,\,\bigcap_{i=2}^{\hat{n}}\{N(t_i)<k_i\}\Big\}\\
&\quad=\sum_{\ell_1=k_1}^{k_2-1}\sum_{\ell_2=k_1}^{k_2-1}\sum_{\ell_3=\ell_2}^{k_3-1}\cdots\sum_{\ell_{\hat{n}}=\ell_{\hat{n}-1}}^{k_{\hat{n}}-1}\mathbf{P}\{N(t_1)=\ell_1\}\prod_{i=2}^{\hat{n}}\mathbf{P}\{N(t_i-t_{i-1})=\ell_i-\ell_{i-1}\}.
\end{aligned}$$

for $k_1<k_2\leq\ldots\leq k_{\hat{n}}$. However, by repeated integration by parts, we get

$$\mathbf{P}\Big\{N(t_1)\geq k_1,\bigcap_{i=2}^{\hat n}\{N(t_i)<k_i\}\Big\}$$

$$=\mathbf{P}\Big\{T_{k_1}\leq t_1,\bigcap_{i=2}^{\hat n}\{T_{k_i}>t_i\}\Big\}$$

$$=\int_0^{t_1}\mathbf{P}\Big\{\bigcap_{i=2}^{\hat n}\{T_{k_i-k_1}>t_i-t\}\Big\}\,f_{k_1}(t)\,dt$$

$$=\int_0^{t_1}\mathbf{P}\Big\{\bigcap_{i=2}^{\hat n}\{N(t_i-t)<k_i-k_1\}\Big\}\,f_{k_1}(t)\,dt$$

$$=\int_0^{t_1}\sum_{\ell_2=0}^{k_2-k_1-1}\sum_{\ell_3=\ell_2}^{k_3-k_1-1}\ldots\sum_{\ell_{\hat n}=\ell_{\hat n-1}}^{k_{\hat n}-k_1-1}\mathbf{P}\Big\{\bigcap_{i=2}^{\hat n}\{N(t_i-t)=\ell_i\}\Big\}\,f_{k_1}(t)\,dt$$

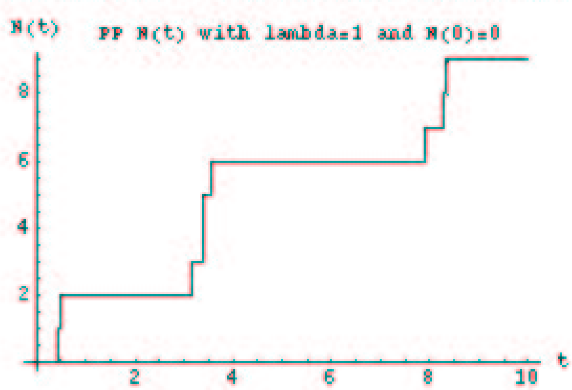
$$=\int_0^{t_1}\sum_{\ell_2=k_1}^{k_2-1}\sum_{\ell_3=\ell_2}^{k_3-1}\ldots\sum_{\ell_{\hat n}=\ell_{\hat n-1}}^{k_{\hat n}-1}\mathbf{P}\{N(t_2-t)=\ell_2-k_1\}\prod_{i=3}^{\hat n}\mathbf{P}\{N(t_i-t_{i-1})=\ell_i-\ell_{i-1}\}\,f_{k_1}(t)\,dt$$

$$=\sum_{\ell_2=k_1}^{k_2-1}\sum_{\ell_3=\ell_2}^{k_3-1}\ldots\sum_{\ell_{\hat n}=\ell_{\hat n-1}}^{k_{\hat n}-1}\lambda^{\ell_2}\mathrm{e}^{-\lambda t_2}\int_0^{t_1}\frac{(t_2-t)^{\ell_2-k_1}}{(\ell_2-k_1)!}\,\frac{t^{k_1-1}}{(k_1-1)!}\,dt\prod_{i=3}^{\hat n}\mathbf{P}\{N(t_i-t_{i-1})=\ell_i-\ell_{i-1}\}$$

$$=\sum_{\ell_2=k_1}^{k_2-1}\sum_{\ell_3=\ell_2}^{k_3-1}\ldots\sum_{\ell_{\hat n}=\ell_{\hat n-1}}^{k_{\hat n}-1}\lambda^{\ell_2}\mathrm{e}^{-\lambda t_2}\sum_{\ell_1=k_1}^{k_2-1}\frac{t_1^{\ell_1}}{\ell_1!}\,\frac{(t_2-t_1)^{\ell_2-\ell_1}}{(\ell_2-\ell_1)!}\prod_{i=3}^{\hat n}\mathbf{P}\{N(t_i-t_{i-1})=\ell_i-\ell_{i-1}\}$$

$$=\sum_{\ell_1=k_1}^{k_2-1}\sum_{\ell_2=k_1}^{k_2-1}\sum_{\ell_3=\ell_2}^{k_3-1}\ldots\sum_{\ell_{\hat n}=\ell_{\hat n-1}}^{k_{\hat n}-1}\mathbf{P}\{N(t_1)=\ell_1\}\prod_{i=2}^{\hat n}\mathbf{P}\{N(t_i-t_{i-1})=\ell_i-\ell_{i-1}\}.\quad\square$$

```
In[1]:= << Statistics`ContinuousDistributions`;

In[2]:= Jump = {N[Random[ExponentialDistribution[1]]]}; While[Last[Jump] ≤ 10,
          Jump = Join[Jump, {Last[Jump] + N[Random[ExponentialDistribution[1]]]}]];

In[3]:= Heavyside[t_] := If[t ≥ 0, 1, 0];

In[4]:= PP[t_] := Sum[Heavyside[t - Jump[[k]]], {k, 1, Length[Jump]}]

In[5]:= Plot[PP[t], {t, 0, 10}, AxesLabel → {"t", "N(t)"}, PlotPoints → 140,
          PlotLabel → "    PP N(t) with lambda=1 and N(0)=0"]
```




```
Out[5]= - Graphics -
```

C Appendix. Solution to Exercise 64

Pick a $t>0$. By assumption, the Lebesgue integral $Y(\omega,t)=\int_0^t X(\omega,r)\,dr$ exists. Writing $\mathrm{Leb}(\cdot)$ for the Lebesgue measure on $\mathbb{R}$, by definition, this means that at least one of the limits $I^{(+)}(\omega)$ and $I^{(-)}(\omega)$, as $n\to\infty$, of the increasing sequences

$$I_n^{(+)}(\omega,t)\equiv\sum_{k=1}^{2^{2n}}2^{-n}(k-1)\,\mathrm{Leb}\Big(\Big\{r\in[0,t]:X(\omega,r)\in[2^{-n}(k-1),2^{-n}k)\Big\}\Big)$$

and

$$I_n^{(-)}(\omega,t)\equiv\sum_{k=-2^{2n}+1}^{0}2^{-n}k\,\mathrm{Leb}\Big(\Big\{r\in[0,t]:X(\omega,r)\in[2^{-n}(k-1),2^{-n}k)\Big\}\Big),$$

are finite. The value of the integral (possibly infinite) is

$$Y(\omega,t)=I^{(+)}(\omega,t)-I^{(-)}(\omega,t)=\lim_{n\to\infty}\Big(I_n^{(+)}(\omega,t)-I_n^{(-)}(\omega,t)\Big).$$

If X is progressively measurable, then we have

$$\Big\{(\omega,r)\in\Omega\times[0,t]:X(\omega,r)\in[2^{-n}(k-1),2^{-n}k)\Big\}\in\mathcal{F}_t\times\mathcal{B}([0,t])\quad\text{for}\quad k\in\mathbb{Z}.\tag{C.1}$$

This is the same thing as saying that the indicator

$$I_{n,k}(\omega,r)\equiv I_{\{(\tilde\omega,\tilde r)\in\Omega\times[0,t]\,:\,X(\tilde\omega,\tilde r)\in[2^{-n}(k-1),2^{-n}k)\}}(\omega,r)$$

is a $\mathcal{F}_t\times\mathcal{B}([0,t])$-measurable function. By *Fubini's Theorem*, it follows that

$$I_n^{(+)}(\omega,t)-I_n^{(-)}(\omega,t)=\int_0^t\Big(\sum_{k=1}^{2^{2n}}2^{-n}(k-1)\,I_{n,k}(\omega,r)-\sum_{k=-2^{2n}+1}^{0}2^{-n}k\,I_{n,k}(\omega,r)\Big)\,dr$$

is a $\mathcal{F}_t$-measurable function (of $\omega\in\Omega$). This in turn shows that the limit $Y(\omega,t)$ is $\mathcal{F}_t$-measurable, that is, adapted.

If X is not progressively measurable, but only measurable, we have

$$\Big\{(\omega,r)\in\Omega\times[0,t]:X(\omega,r)\in[2^{-n}(k-1),2^{-n}k)\Big\}\in\mathcal{F}\times\mathcal{B}([0,t])\qquad\text{for}\quad k\in\mathbb{Z}.$$

Even if X is adapted, this is only a "univariate" property, regarding measurability properties of $X(\omega,t)$, as a function of $\omega\in\Omega$, for each t at a time. We do not have what we need [(C.1)] to be able to get adaptedness for Y by *Fubini's Theorem*.

It is not true in general that Y is adapted, when X is not progressively measurable, but only measurable and adapted. What holds in this case is that there exists a version $\tilde Y$ of Y, that is adapted [the proof of which uses Theorem 8.7 and progressive measurability as above (see the proof of Theorem 11.1)]. And one has to assume that the filtration is augmented (Definition 2.20), to get adaptedness (see Exercise 66).

D *Appendix. Selected further Martingale Proofs

Proof of Theorem 23.6.

References

[1] Adler, R.J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, (IMS).

[2] Albin, J.M.P. (1993). Extremes of diffusions over fixed intervals. *Stochastic Process. Appl.* **48** 211-235.

[3] Baverman, M. (2000). Suprema of compound Poisson processes with light tails. Stochastic Process. Appl. **90** 145-156.

[4] Bertoin, J. (1996). *Lévy Processes*, (Cambridge University Press).

[5] Billingsley, P. (1999). *Convergence of Probability Measures*, second edition (Wiley).

[6] Bingham, N.H., Goldie, C.M. and Teugels J.L. (1987). *Regular Variation*, (Cambridge University Press).

[7] Bourbaki, N. (1987). *Topological Vector Spaces, Chapter 1-5*, (Springer).

[8] Chung, K.L. (1974). *A Course in Probability Theory* 2^{nd} Ed., (Academic).

[9] Chung, K.L. (1982). *Lectures from Markov Processes to Brownian Motion*, (Springer).

[10] Darling, D.A. and Siegert, A.J.F. (1953). The first passage problem for a continuous Markov process. *Ann. Math. Stat.* **24** 624-639.

[11] Doob, J.L. (1953). *Stochastic Processes*, (Wiley).

[12] Douglas, R.G. (1972). *Banach Algebra Techniques in Operator Theory*, (Academic).

[13] Dynkin, E.B. (1960). *Theory of Markov Processes*, (Pergamon).

[14] *Encyclopedic dictionary of mathematics, Vol. I-II*, Translated from the second Japanese edition (S. Iyanaga and Y. Kawada Eds.), Translation reviewed by K.O. May. (MIT Press, 1977).

[15] Freedman, D. (1971). *Brownian Motion and Diffusion*, (Holden-Day).

[16] Gihman, I.I. and Skorohod, A.V. (1972). *Stochastic Differential Equations*, (Springer, New York).

[17] Grimmett, G. and Stirzaker, D. (1982). *Probability and Random Processes*, (Clarendon).

[18] Hewitt, E. and Stromberg, K. (1965). *Real and Abstract Analysis*, (Springer).

[19] Itô, K. and McKean Jr., H.P. (1974). *Diffusion Processes and their Sample Paths*, (Springer).

[20] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*, (Springer).

[21] Kallenberg, O. (1997). *Foundations of Modern Probability*, (Springer).

[22] Karatzas, I. and Shreve, S.E. (1988). *Brownian Motion and Stochastic Calculus*, (Springer).

[23] Karatzas, I. and Shreve, S.E. (1998). *Methods of Mathematical Finance*, (Springer).

[24] Karlin, S. and Taylor, H.M. (1981). *A second course in stochastic processes*, (Academic, New York).

[25] Klebaner, F.C. (1998). *Introduction to Stochastic Calculus with Applications*, (Imperial College Press).

[26] Kwapień, S. and Woyczyński, W.A. (1992). *Random Series and Stochastic Integrals: Single and Multiple*, (Birkhäuser).

[27] Liptser, R.S. and Shiryayev, A.N. (1977). *Statistics of Random Processes I. General Theory*, (Springer).

[28] Mandl, P. (1974). *Analytical Treatment of One-dimensional Markov Processes*, (Springer).

[29] Meyer, P.A. (1966). *Probability and Potentials*, (Blaisdell).

[30] Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces*, (Academic).

[31] Protter, P. (1990). *Stochastic Integration and Differential Equations*, (Springer).

[32] Ray, D. (1956). Stationary Markov processes with continuous paths. *Trans. Amer. Math. Soc.* **82** 452-493.

[33] Revuz, D. and Yor, M. (1991). *Continuous Martingales and Brownian Motion*, (Springer).

[34] Rudin, W. (1974). *Real and Complex Analysis* 2^{nd} Ed., (McGraw-Hill).

[35] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*, (Chapman and Hall).

[36] Stroock, D.W. and Varadhan, S.R.S. (1979). *Multidimensional Diffusion Processes*, (Springer).

[37] Øksendal, B. (1998). *Stochastic Differential Equations* 5^{th} Ed., (Springer).

Index

$\|\cdot\|$ 110
a.s. 6
$\mathcal{A}$ 125
$\mathcal{A}_t$ 114
$\mathcal{B}(\mathbb{R}),\ \mathcal{B}(\mathbb{R}^n)$ 4
B 11
$\overline{B}$ 164
BM 11
B^x 11
$\mathbb{C}([a,b])$ 3
$\mathbb{C}(\mathbb{R}^n)$ 128
$\mathbb{C}([0,T]\times\mathbb{R})$ 128
$\mathbb{C}_B(D)$ 36
$\mathbb{C}_0(\mathbb{R}^n)$ 134
$\mathbb{C}^1(\mathbb{R})$ 155
$\mathbb{C}^{1,2}([0,T]\times\mathbb{R})$ 83
$\mathbb{C}^{1,2}([0,\infty)\times\mathbb{R}^n)$ 113
$\mathbb{C}^2(\mathbb{R})$ 81
$\mathbb{C}^2(\mathbb{R}^n)$ 91
$\mathbb{C}^2_0(\mathbb{R}^n)$ 117
$\mathbb{C}^\infty(\mathbb{R}^n)$ 122
$\mathbb{C}^\infty_0(\mathbb{R}^n)$ 122
$d[X,Y](t)$ 86
$\delta(x)$ 122
Δ 123
$\Delta X(t)$ 151
D 145
DL 145
$dX(t)dY(t)$ 86
$dY(t)$ 46
E 6
$\mathbf{E}\{X|\mathcal{G}\}$ 6
$\mathbf{E}\{X|Y\}$ 7
$\mathbf{E}\{X|Y=y\}$ 22
E_T 52
E_∞ 52, 77
$\exp(\lambda)$ 32
$\mathcal{E}X$ 87
$\lfloor\cdot\rfloor$ 37
$[f,g]([a,b]),\ [f,g](t)$ 3, 86, 153
`fidi's` 9
$\mathcal{F}$ 4
$\mathcal{F}_{t+}$ 19
$\mathcal{F}'_t$ 40
$\mathcal{F}_S\wedge\mathcal{F}_T$ 143
$\mathcal{F}_T$ 35
$\mathcal{F}'_T$ 40
$\mathcal{F}_\infty$ 140
$[g]([a,b]),\ [g](t)$ 3, 79, 85, 153
G_λ 157
$[\cdot]$ 37
$\int_0^t X\,dB$ 44, 52, 56, 66
$\int_s^t X\,dB$ 44, 52, 63, 70
$\int_s^t X\,dY$ 151, 152
$\int_0^t X\circ dB$ 44, 105
$\mathbb{L}^0$ 65
$\mathbb{L}^1$ 71, 74
$\mathbb{L}^2$ 69
$\mathbb{L}_B(\mathbb{R})$ 40
l.i.m. 56
$\cdot^-$ 6
$\mathrm{N}(m,\sigma^2)$ 11
N^x 31
ODE 91
ODO 122
ω 4
Ω 4
P 4
$p(\cdot,\cdot,\cdot,\cdot)$ 24
$\hat{p}(\cdot,\cdot,\cdot,\cdot)$ 135
$P(\cdot,\cdot,\cdot,\cdot)$ 24

$\hat{P}(\cdot,\cdot,\cdot,\cdot)$ 136
$\mathbf{P}\{A|\mathcal{G}\}$ 9
PDE 122
PDO 122
P-lim 65
$.^+$ 6
PP 31
$p_t(x,y)$ 11
P_T 52
P_∞ 52, 77
$\mathbb{R}_{n|n}$ 110
SDE 46, 88
$\sigma(\ldots)$ 4
$\text{sign}(x)$ 95
$\sim$ 27, 48, 81
S_T 52
S_∞ 52, 77
$V_g([a,b]),\ V_g(t)$ 1
$\vee$ 101
$\wedge$ 14
wp. 1 126
$X_-(t)$ 151
$[X,Y]^c(t)$ 154

absolute continuity of the integral 73
absolute continuity wrt. BM 115
absolutely continuous random variable 22
adapted stochastic process 18
almost surely, a.s. 6
arcsine law for BM 30
augmented filtration 18
backward equation for diffusion type SDE 134, 136
Bayes' rule 108
Bessel equation (SDE) 156
Bessel process 156
Black-Scholes Formula 88
Borel set 4
boundary conditions for PDE 124
bounded away from zero 112
bounded first moments of stochastic process 21
bounded second moments of stochastic process 21
bounded stochastic process 75
bounded variation 1
Brownian bridge 164
Brownian motion, BM 11
Brownian motion, multidimensional 109
Brownian motion wrt. filtration 92
Burkholder-Davis-Gundy inequalities 147
càdlàg 19
càg 150
Cameron-Martin formula 105
Cauchy functional equation 32
Cauchy problem for PDE 124, 125
characteristic polynomial of PDO 122
class D 145
class DL 145
closed martingale 72
coefficients of PDO 122
compensator 79
conditional expectation wrt. σ-algebra 6
conditional expectation wrt. random variable 7
conditional expectation, elementary 23
conditional probability 9
conditional probability, regular 126
continuous part of covariation 153
convergence almost surely 21
convergence in mean $(\mathbb{L}^1)$ 71, 74
convergence in mean-square 15, 56
convergence in measure 21, 65

convergence in probability 21, 65
convex function 9
covariance function 14
covariation 3
covariation between Itô integrals 86
covariation between local martingales 86
covariation between semimartingales 153
Dambis-Dubins-Schwarz theorem 148
degenerate elliptic PDO 123
density function of random variable 22
diffusion process 46, 89, 110
diffusion matrix of SDE 113
diffusion type SDE 88
Dirac distribution 122
Dirichlet class D and DL 145
Dirichlet problem for PDE 124, 125
dispersion of diffusion type SDE 88
distribution (generalized function) 122
distribution function of random variable 22
distribution of random variable 22
dominated convergence theorem for conditional expectations 8
Doob-Kolmogorov inequality 51
Doob-Meyer decomposition 79
Doob maximal inequality 51
drift of diffusion type SDE 88
Dynkin system 5
Dynkin system lemma 5
elementary Markov property 41
elliptic PDO (second order) 123
Engelbert-Schmidt theorems 112, 163
Euler iteration 47, 119
event 4
exp(λ)-distribution 32
expectation of positive random variable 6
expectation of random variable 6
exponential distribution, exponential random variable 32
extended real line, $\overline{\mathbb{R}}$ 29
Fatou's lemma for conditional expectations 8
Feller process 36
Feynman-Kac formula 128
filtration 18
finite dimensional distributions, `fidi's` 9
finite variation 1
finite variation over $[a, b]$ 1
Fisk-Stratonovich integral 44, 104
french school 150
fundamental kernel of PDO 123
fundamental solution of PDO 122
fundamental measure to parabolic PDO 136
fundamental solution to parabolic PDO 134
Gaussian distribution, Gaussian random variable 11, 13
Gaussian process 13
general linear SDE 90
general theory (of stochastic processes) 150
generator of multidimensional SDE 114
generator of time homogeneous multidimensional SDE 125
Girsanov's theorem 105, 108, 111
Girsanov transformation 105
Grönwall's lemma 95
heat operator 123
hitting time for BM 28
hyperbolic PDO (second order) 123

Hölder condition 97
Hölder's inequality 51
independent increments 11, 15
independent random variables 4
independent σ-algebras 4
initial conditions for PDE 124
integrable process 21
integration by parts of Itô integrals 84, 92, 153
integration by parts of Stratonovich integrals 105
Itô integral wrt. BM 44
Itô integral wrt. BM of processes in E_T 56
Itô integral wrt. BM of processes in P_T 66
Itô integral wrt. BM of processes in S_T 52
Itô integral wrt. semimartingale 151, 152
Itô's formula 81, 83, 87, 91, 154, 162, 163
Itô's formula for convex functions 162, 163
Itô's formula for local martingales 91
Itô's formula for semimartingales 154
Itô's rule 87
Jensen's inequality for conditional expectations 8
jump discontinuity 2
Kazamaki's criterion 106
Knight's theorem 150
Kolmogorov backward equation 134
Kolmogorov backward equation, time-homogeneous 136
Langevin equation (SDE) 89, 159
Laplacian 123
last element of smartingale 140
Lévy process 25
Lévy's characterization of BM 103, 111
linear PDO 122
linear SDE 90
Lipschitz condition 95
local martingale 74
local martingale problem 114
local smartingale 74
local time 162
localizing sequence of stopping times 74
locally bounded function 32
locally bounded stochastic process 151
locally integrable function 90
Markov process 24
Markov property 24, 40
Markov property of BM 26
Markov property of PP 34
Markov property of SDE 48, 132
Markov process wrt. itself 25
Markov time 139
martingale 18
martingale convergence 139, 141
martingale problem 117
martingale property of BM 21
martingale property of PP 34
martingale wrt. itself 18
mathematical finance ii, 85, 88
measurable process 35
measurable space 4
multidimensional diffusion process 110
multidimensional SDE 113
multidimensional SDE of diffusion type 110
$\mathrm{N}(m, \sigma^2)$-distribution 11
normal distribution, normal random variable 11, 13
normal process 13
Novikov's criterion 107

optional sampling theorem 141, 142
optional stopping theorem for martingales 73, 143
optional stopping theorem for local martingales 143
optional time 138
order of PDO 117
ordinary differential operator, ODO 122
Ornstein-Uhlenbeck process 90, 159
Ornstein-Uhlenbeck process, inhomogeneous 90
parabolic PDO (second order) 123
parameter set of stochastic process 9
partial differential equation, PDE 122
partial differential operator, PDO 122
PDO of mixed type (second order) 124
PDO with constant coefficients 122
Picard-Lindelöf iteration 98
Po(λ)-distribution 31
pointwise conditional expectation 22
Poisson distribution, Poisson random variable 31
Poisson process, PP 31
polarization 3
predictable process 52, 150
predictable time 138
probability density of BM 11
probability measure 4
probability space 4
progressively measurable process 57
quadratic variation 3, 79
quadratic variation of BM 15
quadratic variation of Itô integral 77, 78, 85
quadratic variation of local martingale 85
quadratic variation of martingale 79
quadratic variation of semimartingale 153
quasimartingale 154
random time 138
random variable 4
$\mathbb{R}^n$-valued random variable 4
reflection principle 27
reflection principle for BM 27
regular conditional probability 126
regular pointwise conditional probability 132
regular martingale 72
removal of drift for one-dimensional time homogeneous diffusion 161, 162
resolvent 157
right-continuous filtration 20
scale function of one-dimensional time homogeneous diffusion 161
Schwarz distribution 122
SDE of diffusion type 88
SDE of diffusion type, multidimensional 109
section of set 57
self-similar process 26
semimartingale 150
separable stochastic process 10
separator for stochastic process 10
σ-algebra 4
σ-algebra generated by random variable 4
σ-algebra generated by σ-algebras 4
σ-algebra of the future 40
simple process 52
simple predictable process 151
Skorohod theorem 121
smartingale 18
smartingale convergence 139, 141
smartingale wrt. itself 18
space homogeneity of BM 11

special semimartingale 154
square-integrable martingale 21
square-integrable process 21
stationary distribution 157
stationary increments 11, 15
stationary Gaussian Markov process 160
stationary process 157
stochastic differential equation, SDE 46, 88
stochastic differentials 46, 84
stochastic exponential 88, 92
stochastic process 9
stopping time 35, 138
Stratonovich integral 44, 104
strictly optional time 139
strong Markov process 36
strong Markov property 36, 41
strong Markov property of BM 43
strong Markov property of Lévy process 42, 144
strong Markov property of SDE 134
strong solution to SDE 89, 93
strongly elliptic PDO 127
Stroock-Varadhan theorem 119, 126
submartingale 18
supermartingale 18
time homogeneous Markov process 25
time homogeneous SDE 111
time homogeneous transition probability 25
tranformation of time homogeneous one-dimensional diffusion 161
transition density function of Markov process 24
transition density for multidimensional diffusion 135
transition density for time homogeneous multidimensional diffusion 136
transition density for time homogeneous one-dimensional diffusion 165
transition distribution function of Markov process 24
transition probability of Markov process 24
uniformly integrable family of random variables 71
unique strong solution to SDE 94
unique weak solution to SDE 94
usual conditions 20
version of stochastic process 10
Wald's identity for BM 77, 107
weak Markov process 39
weak solution to SDE 89, 93
well-posed at zero local martingale problem 116
well-posed local martingale problem 116
well-posed at zero martingale problem 117
well-posed martingale problem 117
Wiener process 11
Yamada-Watanabe theorem 97