BROWNIAN MOTION AND ITÔ CALCULUS

CHRISTIAN BAYER

Abstract. The aim of this text is to give an introduction to Itô calculus. It is based on a short course about the subject given by the author at the WK-Summer camp 2006 at the lake Weissensee in Austria. The emphasis lies on the probabilistic basics of the stochastic integration and, in the second part, on the connections with PDEs.

Contents

1. Preliminaries from Probability Theory

Let $\big(\Omega, \mathcal{F}, P\big)$ be a probability space. Mostly for convenience, we will usually assume the probability space to be complete, i. e. given $A \in \mathcal{F}$ with $P(A) = 0$ and

$B \subset A$, then $B \in \mathcal{F}$. Completeness allows us to avoid formulations like "property ... is satisfied outside a set with probability 0" in favor of "property ... is satisfied almost surely (a. s.)", and the assumption does not cost us anything, since it is always possible to pass to the completion of a probability space.

1.1. **Independence.** One of the most important notions of probability is the notion of *independence*, which plays a dominant rôle in statistics and somehow separates probability theory from measure theorem. We start with the elementary definition.

Definition 1.1. Given two measurable sets A and B. The sets are called *independent* if

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

We write $A \perp\!\!\!\perp B$.

This definition is not satisfying, since we are really interested whether two – or more – given random variables are independent or not. By a random variable we understand a measurable function from the measurable space $(\Omega, \mathcal{F})$ to a polish – i. e. seperable, complete, metrizeable – space endowed with its Borel σ-field. Usually, this space will be $\mathbb{R}$ or $\mathbb{R}^d$, but in some situations we will consider random variables taking values in more general polish spaces. Polish spaces are attractive since integration theory can be constructed using the same ideas as in the case of finite-dimensional vector spaces. We do not go into details her and refer to Teichmann[10].

Definition 1.2. Given n random variables $X_1, \dots, X_n$ taking values in some polish space $(H, \mathcal{B}(H))$, where $\mathcal{B}(H)$ denotes the Borel σ-field on H. $X_1, \dots, X_n$ are called independent if for all bounded and measurable functions $f_i : H \to \mathbb{R}$, $i = 1, \dots, n$, we have

$$E\big(f_1(X_1) \cdots f_n(X_n)\big) = E\big(f_1(X_1)\big) \cdots E\big(f_n(X_n)\big).$$

A family $\{X_i, i \in I\}$ is called independent if the elements of each finite subset are independent. Again we use the symbol $\perp\!\!\!\perp$ to indicate independence of random variables.

Remark 1.3. As the following example shows, it is not enough to require that each X_i is (pairwise) independent of each X_j, $i, j = 1, \dots, n$.

Example 1.4. Let X and Y be two independent random variables with $P(X = 1) = P(Y = 1) = \frac{1}{2}$ and $P(X = -1) = P(Y = -1) = \frac{1}{2}$, i. e.

$$P\big(X = (-1)^i,\ Y = (-1)^j\big) = \frac{1}{4}, \quad i,\ j \in \{0, 1\}.$$

Define the random variable $Z = XY$, i. e. $Z(\omega) = X(\omega)Y(\omega)$. Elementary calculations show that both $X \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z$, but the three random variables X, Y, and Z are obviously not independent.

Lemma 1.5. *Let $\mu_i = (X_i)_* P$ be the law of the random variable X_i, $i = 1, \dots, n$, i. e. μ_i is the probability measure on $\mathcal{B}(H)$ satisfying $\mu_i(A) = P(X_i \in A)$ for $A \in \mathcal{B}(H)$. The random variables $X_1, \dots, X_n$ are independent if and only if their joint law μ on $\mathcal{B}(H^n)$, i. e. the law of the random vector $X = (X_1, \dots, X_n)$, is the product measure of the μ_is, in symbols $\mu = \mu_1 \otimes \cdots \otimes \mu_n$.*

Proof. The lemma follows easily using the fact that $\mathcal{B}(H^d)$ is generated by the system of all measurable rectangles $A_1 \times \cdots \times A_n$, $A_i \in \mathcal{B}(H)$, $i = 1, \dots, n$. □

Even this definition is not really satisfactory for us. The notion of independence of σ-fields allows us to say that a random variable is independent of a σ-field, i. e. of every random variable measurable with respect to this σ-field.

Definition 1.6. Let $\{\mathcal{G}_i, i \in I\}$ be a family of σ-fields on Ω, $\mathcal{G}_i \subset \mathcal{F}$, $\forall i \in I$. The σ-fields $\mathcal{G}_i$ are independent if $\forall n \in \mathbb{N}$, $\forall i_1, \dots, i_n \in I$ with $i_j \neq i_l$ for $j \neq 1$ and for all $\mathcal{G}_{i_j}$-measurable random variables X_{i_j}, $j = 1, \dots, n$ we have: the random variables $X_{i_1}, \dots, X_{i_n}$ are independent.

Remark 1.7. Note that a collection of random variables $\{X_i, i \in I\}$ is independent if and only if the collection of the respective generated σ-fields $\{\sigma(X_i), i \in I\}$ is independent. Recall that the σ-field $\sigma(f)$ generated by a function $f : \Omega \to H$ is the smallest σ-field on Ω such that f is Borel-measurable.

1.2. **Conditional Expectations.** The concept of conditional expectations and probabilities is maybe one of the most difficult concepts in probability theory, since often rather subtle problems appear. We begin with the general definition. Note that we work with $\mathbb{R}$-valued random variables in this subsection. The generalization to $\mathbb{R}^d$-valued ones is obvious, as is the generalization to Banach space valued random variables.

Definition 1.8. Given $X \in L^1\big(\Omega, \mathcal{F}, P\big)$ and a sub-σ-field $\mathcal{G} \subset \mathcal{F}$. The *conditional expectation of X with respect to $\mathcal{G}$* is the unique $\mathcal{G}$-measurable random variable $Z \in L^1\big(\Omega, \mathcal{G}, P\big)$ such that

$$E\big(YZ\big) = E\big(YX\big), \quad \forall Y \in L^\infty\big(\Omega, \mathcal{G}, P\big).$$

We write $Z = E\big(X|\mathcal{G}\big)$. (Note that we used the same symbol for the measure P on $\mathcal{F}$ and its restriction to $\mathcal{G}$.)

Proof of existence. Let us first assume that $X \geq 0$ and define the measure μ_X on $(\Omega, \mathcal{G})$ by

$$\mu_X(C) = E(\mathbf{1}_C X), \quad C \in \mathcal{G},$$

where the *indicator function* of the set C is defined by

$$\mathbf{1}_C(\omega) = \begin{cases} 1 & \omega \in C \\ 0 & \omega \notin C \end{cases}.$$

This measure is absolutely continuous with respect to P, therefore the Radon-Nikodym theorem implies existence of a (unique) density, i. e. of a (non-negative) function $Z \in L^1(\Omega, \mathcal{G}, P)$ such that

$$E(YX) = \int_\Omega Y(\omega)\mu_X(d\omega) = \int_\Omega Y(\omega)Z(\omega)P(d\omega) = E(ZY)$$

for $Y \in L^\infty(\Omega, \mathcal{G}, P)$.

The proof in the general case follows by considering the positive and the negative part of X separately. □

Remark 1.9. The conditional expectation is an almost surely defined random variable. It is unique as an element of $L^1(\Omega, \mathcal{G}, P)$.

Remark 1.10 (Conditional expectation as an orthogonal projection). Assume that $\mathcal{G}$ contains all $\mathcal{F}$-measurable sets of P-measure 0. Then we may consider $L^2(\Omega, \mathcal{G}, P) \subset L^2(\Omega, \mathcal{F}, P)$ and it is easy to see that $L^2(\mathcal{G})$ is even a closed subspace of $L^2(\mathcal{F})$. Therefore, there is an orthogonal projection $L^2(\mathcal{F}) \to L^2(\mathcal{G})$, and this orthogonal projection is nothing else but the conditional expectation. In other words, the conditional expectation of an L^2-random variable is the best approximation by a $\mathcal{G}$-measurable random variable in the least square sense.

Proposition 1.11. *The conditional expectation has the following properties.*
(1) Linearity*: $\alpha, \beta \in \mathbb{R}$, $X, Y \in L^1(\Omega, \mathcal{F}, P)$, then*

$$E\big(\alpha X + \beta Y\big|\mathcal{G}\big) = \alpha E\big(X\big|\mathcal{G}\big) + \beta E\big(Y\big|\mathcal{G}\big),$$

where the equality, as usual, is understood in L^1, i. e. a. s.
(2) Monotonicity*: $X, Y \in L^1(\mathcal{F})$, $X \geq Y$ a. s., then $E(X|\mathcal{G}) \geq E(Y|\mathcal{G})$ a. s.*
(3) Jensen's inequality*: $X \in L^1(\mathcal{G})$, $\varphi : \mathbb{R} \to \mathbb{R}$ convex, then $E(\varphi(X)|\mathcal{G}) \geq \varphi\big(E(X|\mathcal{G})\big)$ a. s.*
(4) Continuity*: $E(\cdot|\mathcal{G}) : L^1(\Omega, \mathcal{F}, P) \to L^1(\Omega, \mathcal{G}, P)$ is continuous. More precisely,*

$$E\big(|E(X|\mathcal{G})|\big) \leq E\big(|X|\big), \quad X \in L^1\big(\Omega, \mathcal{F}, P\big).$$

By restriction, we get continuous maps $E(\cdot|\mathcal{G}) : L^p(\Omega, \mathcal{F}, P) \to L^p(\Omega, \mathcal{G}, P)$, $1 < p < \infty$.
(5) For any $X \in L^1(\mathcal{F})$ we have $E\big(E(X|\mathcal{G})\big) = E(X)$.
(6) If $X \in L^1(\mathcal{F})$ is actually $\mathcal{G}$-measurable, then $E(X|\mathcal{G}) = X$ a. s. In contrast, if $X \perp\!\!\!\perp \mathcal{G}$ (i. e. $\sigma(X) \perp\!\!\!\perp \mathcal{G}$, then $E(X|\mathcal{G}) = E(X)$ a. s.
(7) Tower law*: Given another σ-field $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ and $X \in L^1(\mathcal{F})$, then we have the equality (in $L^1(\Omega, \mathcal{H}, P)$)*

$$E\big(E\big(X\big|\mathcal{G}\big)\big|\mathcal{H}\big) = E\big(X\big|\mathcal{H}\big).$$

(8) For $X \in L^p(\mathcal{F})$ and $Y \in L^q(\mathcal{G})$, $1 \leq p \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, we get

$$E\big(XY\big|\mathcal{G}\big) = YE\big(X\big|\mathcal{G}\big), \quad \text{a. s.}$$

Proof. (1) and (2) are easily seen, (3) follows using similar ideas as for Jensen's inequality for the expectation.

(5) is proved by using the $\mathcal{G}$-measurable function $Y \equiv 1$ in the defining property of the conditional expectation, see Definition 1.8. Note that (5) can be interpreted as a special case of the tower law for the trivial σ-field $\mathcal{H} = \{\varnothing, \Omega\}$.

For the proof of (4) note that monotonicity implies that $|E(X|\mathcal{G})| \leq E(|X|\,|\mathcal{G})$ and the above inequality follows by property (5). Continuity of the restriction to L^p follows in the same way, using the Jensen inequality.

The first part of (6) follows immediately from the definition. Given $X \perp\!\!\!\perp \mathcal{G}$, $X \in L^1(\mathcal{F})$, $Y \in L^\infty(\mathcal{G})$ we have, using $X \perp\!\!\!\perp Y$,

$$E(XY) = E(X)E(Y) = E\big(E(X)Y\big).$$

Thus, we can choose $E(X) = E(X|\mathcal{G})$.

For the proof of (8), choose $V \in L^\infty(\mathcal{G})$ and note that

$$E(XYV) = E\big(E(XY|\mathcal{G})V\big)$$

and – using an L^∞-approximation of YV –

$$E(XYV) = E\big(E(X|\mathcal{G})YV\big),$$

and consequently $E(XY|\mathcal{G}) = YE(X|\mathcal{G})$.

For the tower law we need to verify that $E(E(X|\mathcal{G})Y) = E(XY)$, for all $Y \in L^\infty(\mathcal{H})$. Using properties (5) and (8), we can argue as follows:

$$E\big(E(X|\mathcal{G})Y\big) = E\big(E(XY|\mathcal{G})\big) = E(XY). \qquad \square$$

Definition 1.12. The *conditional probability* of a set $A \in \mathcal{F}$ with respect to the σ-field $\mathcal{G}$ is the random variable $P(A|\mathcal{G}) = E(\mathbf{1}_A|\mathcal{G})$.

Note that the conditional probability does not necessarily define a random measure, i. e. it might not be true that for almost all ω the function $\mathcal{F} \to \mathbb{R}$, $A \mapsto P(A|\mathcal{G})(\omega)$ is a probability measure. Indeed, for a given set $A \in \mathcal{F}$ it is certainly true (due to monotonicity of the conditional expectation) that $P(A|\mathcal{G})(\omega) \in [0, 1]$, for almost all ω. If, however, $\mathcal{F}$ contains an uncountably infinite number of sets, we cannot be sure that $\forall A \in \mathcal{F} : P(A|\mathcal{G})(\omega) \in [0, 1]$ for almost all ω, since the intersection of uncountably many sets with probability 1 does not need to have probability 1.

It is always possible to choose the conditional probability of a probability measure on $\mathbb{R}$ (or $\mathbb{R}^d$) as a random measure, and we will do so.

We will also need the notion of the conditional expectation of one random variable conditioned on the event that some other random variable takes a certain value. Given two random variables X and Y on $(\Omega, \mathcal{F}, P)$, it is rather obvious to define the conditional expectation of X given Y by

$$E(X|Y) = E(X|\sigma(Y)).$$

Definition 1.13. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel-measurable, Y_*P-integrable function such that $E(X|Y)(\omega) = h(Y(\omega))$ for almost all ω. Then we define the expectation of X given $Y = y$ by

$$E(X|Y = y) = h(y), \quad y \in \mathbb{R}.$$

Note that such a function h always exists and is Y_*P-a. s. unique.

1.3. **Central Limit Theorem and Law of Large Numbers.** The central limit theorem is one of the most important theorems in probability and statistics. It roughly says that the average of independent random variables is asymptotically normal. In many cases, the central limit theorem is applied as follows: one is given a random variable X which is either explicitly given as an average of many independent random variables – for example, in Monte-carlo approximations – or can at least be interpreted in that way. Then, if the number of independent variables is large enough, one replaces the unknown true distribution of X by a normal distribution.

We make precise what we mean by "asymptotic normality" by formulating one version of the theorem.

Theorem 1.14. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed (i. i. d.) real random variables such that both the expected value $m = E(X_1)$ and the variance $\sigma^2 = E((X_1 - m)^2)$ exist. Let $S_n = \sum_{i=1}^n X_i$. Then*

$$\sqrt{n}\frac{\frac{1}{n}S_n - m}{\sigma} \rightharpoonup \mathcal{N}(0, 1),$$

where $\mathcal{N}(0, 1)$ is a standard normal distribution and the limit is understood in the sense of weak convergence.

Remark 1.15. Recall that a sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ on some polish space $(H, \mathcal{B}(H))$ converges weakly to some probability measure μ on $(H, \mathcal{B}(H))$ if for all continuous, bounded $f : H \to \mathbb{R}$ we have

$$\int_H f(x)\mu_n(dx) \to \int_H f(x)\mu(dx).$$

A sequence of random variables converges weakly to some distribution if their laws converges weakly to the distribution in the above sense.

A similar theorem holds for $\mathbb{R}^d$-valued random variables, and we will later encounter an infinite-dimensional version.

The *strong law of large numbers* says that the average of an i. i. d. sequence of integrable random variables converges almost surely to the expected value of the random variable. It is very helpful if we want to approximate the expected value of a given random variable $X \in L^1(\Omega, \mathcal{F}, P)$. Given a sequence X_n of independent copies of X, i. e. the $(X_n)_{n\in\mathbb{N}}$ are independent and all of them have the same law as X. Consequently, $X_n \in L^1(\Omega, \mathcal{F}, P)$ for each n.

Theorem 1.16. *Let $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$, denote the sequence of partial sums of (X_n). Then*

$$\lim_{n\to\infty} \frac{S_n}{n} = E(X) \quad a.\ s.$$

For proofs and more general formulations of the central limit theorem and the strong law of large numbers see, for example, Bauer[1] or Billingsley[3].

1.4. **Gaussian Random Variables.** In this subsection we study properties of multi-dimensional Gaussian random variables. Recall that the one-dimensional Gaussian measure with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$ has the density $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(x-m)^2}{2\sigma^2}\big)$.

Definition 1.17. Given an $\mathbb{R}^d$-valued random variable X. Define the vector $m = E(X) = (E(X_1), \dots, E(X_d))$ and the *covariance matrix* $\Sigma = (\sigma_{ij})_{i,j=1,\dots,d}$ by $\sigma_{ij} = E\big((X_i - E(X_i))(X_j - E(X_j))\big)$, $i, j = 1 \dots, d$. X follows a *Gaussian distribution* with mean m and covariance matrix Σ, symbolically $X \sim \mathcal{N}(m, \Sigma)$, if all the one-dimensional projections have (one-dimensional) Gaussian distributions:

$$\langle \lambda\,, X\rangle \sim \mathcal{N}\big(\langle \lambda\,, m\rangle\,, \langle \lambda\,, \Sigma\lambda\rangle\big), \quad \forall \lambda \in \mathbb{R}^d.$$

Remark 1.18. Note that the covariance matrix is a symmetric, non-negative definite matrix. For us, Dirac distributions are Gaussian distributions (with variance 0). This is convenient for notational purposes and also because with this convention, the L^2-limit of a sequence of Gaussian random variables is again a Gaussian random variable.

Furthermore, note that for an $\mathbb{R}^d$-valued Gaussian random variable X and an $m\times d$-matrix A, the random variable AX is nn $\mathbb{R}^m$-valued Gaussian random variable.

Proposition 1.19. *Given a d-dimensional Gaussian random variable X with pairwise uncorrelated components, i. e. with a diagonal covariance matrix. Then the components $X_1, \dots, X_d$ of X are independent.*

Proof. The proof is rather simple, since for a diagonal covariance matrix the distribution function of the vector X is the product of the distribution functions of its components, which implies independence of the components. □

Remark 1.20. Proposition 1.19 says that uncorrelated random variables, which are *jointly* Gaussian, are independent. Of course, uncorrelatedness usually does not imply independence, which makes it so very hard to determine independence in practice. As the following example shows, this is not even true for Gaussian random variables which are not jointly Gaussian.

Example 1.21. Let $X \sim \mathcal{N}(0, 1)$ and for each $c \geq 0$ define the random variable Y_c by

$$Y_c(\omega) = \begin{cases} X(\omega), & |X(\omega)| \leq c \\ -X(\omega), & |X(\omega)| > c \end{cases}.$$

Then, for each non-negative c, $Y_c \sim \mathcal{N}(0, 1)$. We calculate the correlation between X and Y_c. For $c = 0$, $Y_0 = -X$ and thus $E(XY_0) = -1$. In general,

$$E(XY_c) = E\big(X^2 \mathbf{1}_{\{|X|\leq c\}}\big) - E\big(X^2 \mathbf{1}_{\{|X|>c\}}\big).$$

For $c \to \infty$, the second term on the right hand side converges to 0 by dominated convergence. Thus, $\lim_{c\to\infty} E(XY_c) = 1$, which implies – by continuity in c – existence of a c_0 such that X and Y_{c_0} are uncorrelated. (Numerically, $c_0 \approx 0.54$.) But for each c,

$$P(X > c, Y_c > c) = 0 \neq P(X > c)P(Y_c > c),$$

implying that X and Y_c are never independent. It is easy to see (directly) that the vector (X, Y_c) is not Gaussian.

2. Stochastic Processes

Definition 2.1. Given a probability space $(\Omega, \mathcal{F}, P)$ and an index set I, a *stochastic process* is a collection $\{S_i, i \in I\}$ of random variables $S_i : \Omega \to \mathbb{R}$ (or $\mathbb{R}^d$, we will not discuss more general processes), $i \in I$.

We are mainly interested in stochastic processes in continuous time, i. e. with $I = [0, T]$ or $I = [0, \infty[$ and then we write $(S_t)_{t\in[0,T]}$ or $(S_t)_{t\in[0,\infty[}$, respectively.

2.1. Kolmogorov's Extension Theorem.

In modeling, one often postulates distributional properties of the finite dimensional projections $(S_{i_1}, \dots, S_{i_n})$, $i_1, \dots, i_n \in I$, of a stochastic process and asks oneself whether it is possible to construct a process satisfying these properties (it might happen the some postulates contradict each other, even though it is not obvious from the beginning; we will see a few examples of this kind.)

More precisely, given the finite dimensional marginals, i. e. the distributions of all finite-dimensional projections of a process, is it possible to find a probability space such that one can define a stochastic process thereon having the required marginals? It is clear, that some consistency relations must hold: the distribution of (S_{i_1}, S_{i_2}) already determines both the distributions of S_{i_1} and S_{i_2} and (S_{i_2}, S_{i_1}). It turns out, that this consistency requirement is already enough to guarantee the existence of a stochastic process.

Theorem 2.2. *Given an index set I and a family $\{\mu_{i_1,\dots,i_n} \mid n \in \mathbb{N}, i_1, \dots, i_n \in I\}$, where $\mu_{i_1,\dots,i_n}$ is a probability measures on $\mathbb{R}^n$ (or $\mathbb{R}^{dn}$), satisfying the consistency relations*

$$\mu_{i_1,\dots,i_n}(A_{i_1} \times \cdots \times A_{i_n}) = \mu_{i_{\sigma(1)},\dots,i_{\sigma(n)}}(A_{i_{\sigma(1)}} \times \cdots \times A_{i_{\sigma(n)}})$$

for all $n \in \mathbb{N}$, $i_1, \dots, i_n \in I$, $A_{i_1}, \dots, A_{i_n} \in \mathcal{B}(\mathbb{R})$ and all permutations σ of n numbers, and

$$\mu_{i_1,\dots,i_n}(A_{i_1} \times \cdots \times A_{i_n}) = \mu_{i_1,\dots,i_n,j_1,\dots,j_m}(A_{i_1} \times \cdots \times A_{i_n} \times \mathbb{R} \times \cdots \times \mathbb{R})$$

for all $n, m \in \mathbb{N}$, $i_1, \dots, i_n, j_1, \dots, j_m \in I$, $A_{i_1}, \dots, A_{i_n} \in \mathcal{B}(\mathbb{R})$ (or the respective conditions for $\mathbb{R}^d$ instead of $\mathbb{R}$). Then there is a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $(S_i)_{i\in I}$ thereon such that

$$\mu_{i_1,\dots,i_n}(A_{i_1} \times \cdots \times A_{i_n}) = P(S_{i_1} \in A_{i_1}, \dots, S_{i_n} \in A_{i_n})$$

for all $n \in \mathbb{N}$, $i_1, \dots, i_n \in I$, $A_{i_1}, \dots, A_{i_n} \in \mathcal{B}(\mathbb{R})$. In other words, $\mu_{i_1,\dots,i_n} = (S_{i_1}, \dots, S_{i_n})_* P$.

Remark 2.3. The proof of Kolmogorov's extension theorem shows that we can choose the probability space $\Omega = \mathbb{R}^I$ with the σ-field generated by the 1-dimensional projections, i. e. $\mathcal{F} = \sigma(\{\pi_i, i \in I\}) = \mathcal{B}(\mathbb{R})^{\otimes I}$, where $\pi_i : \mathbb{R}^I \to \mathbb{R}$, $(x_j)_{j\in I} \mapsto x_i$, and $S_i = \pi_i$, $i \in I$. In the case of $I = [0, \infty[$, we get the space of all functions $[0, \infty[\to \mathbb{R}$ as our probability space. (Note that it is not surprising that one might need a large enough probability space.) This representation of the stochastic process – with $\Omega = \mathbb{R}^I$, $\mathcal{F} = \mathcal{B}(\mathbb{R})^{\otimes I}$, $S_i(\omega) = \omega_i$ and P such that S has the prescribed marginals – is called *canonical representation* of the process.

The proof is quite technical. We refer to Bauer[1] and Billingsley[3] for a measure theoretic proof and to Teichmann[10] for a functional analytic proof.

Remark 2.4. It is possible to interpret the newly constructed probability space $(\Omega, \mathcal{F}, P)$ as the projective limit of the family of probability spaces $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_{i_1,\dots,i_n})$ or the respective spaces in the d-dimensional case.

2.2. **Equality of Stochastic processes.** Apart from equality of stochastic processes in terms of finite-dimensional marginals, we will discuss two important notions of equality. In this subsection, we consider stochastic processes defined on some *complete* probability space $(\Omega, \mathcal{F}, P)$ indexed by a general index set I.

Definition 2.5. Two stochastic processes S and U are *indistinguishable* if

$$P\Big(\{\omega \in \Omega \mid \forall i \in I : S_i(\omega) = U_i(\omega)\}\Big) = 1.$$

This is one occasion where it is convenient to assume completeness of the probability space: otherwise, we could not be sure that the set above is even measurable, even if we knew that the inner measure of the set was 1.

Indistinguishability is a very strong notion, and, indeed, it is often too strong, as the following example shows.

Example 2.6. Let $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}([0,1]), dx)$ and define the following stochastic processes (indexed by $I = [0,1]$) thereon: $S_i(\omega) = 0$ for all $\omega \in [0,1]$ and

$$U_i(\omega) = \begin{cases} 0 & \omega \neq i \\ 1 & \omega = i \end{cases},$$

$i, \omega \in [0,1]$. Then, even though for each i, $U_i \neq 1$ in only one ω, we have $P(\{\forall i \in [0,1] : S_i = U_i\}) = 0$, there is not a single ω such that $S_i(\omega) = U_i(\omega)$ for all i. Nevertheless, the finite-dimensional marginals of these two processes are the same. This already shows that the finite-dimensional marginals do not uniquely determine a process. (This is rather obvious since the finite-dimensional marginals do not determine the probability space at all, provided it is rich enough.)

Definition 2.7. Two stochastic processes S and U are *versions* or *modifications* of each other if

$$\forall i \in I : \quad P\Big(\{\omega \in \Omega \mid S_i(\omega) = U_i(\omega)\}\Big) = 1.$$

Naturally, Definition 2.7 is much weaker than Definition 2.5. The two definitions are equivalent if the index set is at most countably infinite. Note that two processes which are versions of each others have the same finite-dimensional distributions.

2.3. **Regularity of Stochastic Processes.** In this subsection, we consider a stochastic process $(S_t)_{t \in [0,\infty[}$ in continuous time with values in $\mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. It is easy to generalize the notions to processes with values in polish spaces.

Definition 2.8. S is called *measurable* if the mapping $S : [0,\infty[\times \Omega \to \mathbb{R}$, $(t, \omega) \mapsto S_t(\omega)$ is measurable with respect to the product σ-field $\mathcal{B}([0,\infty[) \otimes \mathcal{F}$.

Despite being fairly natural, the notion of measurability of processes does not play a big rôle in the theory.

Definition 2.9. A stochastic process S is called *separable* if there is a countable, dense subset $D \subset [0,\infty[$ and a measurable set $N \in \mathcal{F}$ with $P(N) = 0$ such that $\forall t \in [0,\infty[$ and any sequence $(t_n)_{n \in \mathbb{N}}$ with $t_n \in D$, $n \in \mathbb{N}$, and $\lim_{n\to\infty} t_n = t$ we have:

$$\exists \lim_{n\to\infty} S_{t_n}(\omega) = S_t(\omega), \quad \forall \omega \in N^c.$$

Definition 2.10. S is called *continuous* if there is a measurable set $N \in \mathcal{F}$ with $P(N) = 0$ such that $\forall \omega \in N^c$ the function $t \mapsto S_t(\omega)$ is continuous.

Remark 2.11. For a fixed ω, the function $t \mapsto S_t(\omega)$ is called a *trajectory* or a *path* of the process.

Clearly, any continuous (separable) process allows for a modification such that continuity (separability) holds for all sample paths, i. e. $\forall\omega$. Any continuous process is also separable. It can be shown that each real-valued stochastic process has a separable modification (it might happen that the separable modification takes values in $\overline{\mathbb{R}}$), see Billingsley[3], Theorem 38.1, but there are real-valued stochastic processes which do not have measurable modifications, see Stoyanov[9].

Kolmogorov established a powerful method to determine whether a given process has a continuous modification. Recall that a function $f : [0,\infty[\to \mathbb{R}$ is called *locally Hölder continuous* of order γ if $\forall t_0 \in [0,\infty[$ we find a neighborhood U of t_0 and a constant $C > 0$ such that

$$\sup_{s,t\in U, s\neq t} \frac{|f(s)-f(t)|^\gamma}{|s-t|} \leq C.$$

Theorem 2.12 (Kolmogorov-Čentsov)**.** *Given a stochastic process S such that there are positive constants α, β, c with*

$$E\Big(|S_s - S_t|^\alpha\Big) \leq c\,|s-t|^{1+\beta}, \quad \forall s,t \in [0,\infty[.$$

Then there is a modification S' of S such that all trajectories of S' are locally Hölder continuous of order γ for all $0 < \gamma < \frac{\beta}{\alpha}$. In particular, S' is continuous.

Again, the proof is rather technical and therefore omitted, e. g. see Bauer[1] or Teichmann[10].

Finally, we want to be able to define an integral with respect to stochastic processes. For this reason, it is natural to discuss processes with trajectories of finite variation. Given $f : [0,\infty[\to \mathbb{R}$, the *variation* of f on $[0,t]$ is defined by

$$V(f)_t = \sup\Big\{\sum_{i=1}^{n-1} |f(t_{i+1}) - f(t_i)| \;\Big|\; n \in \mathbb{N},\, 0 = t_0 < \cdots < t_n = t\Big\}. \tag{2.1}$$

The *total variation* of f is the function $t \mapsto V(f)_t$ and we say that f has *finite variation* if $V(f)_t < \infty$, $\forall t \in [0,\infty[$. Note that a function f of finite variation may have unbounded variation in the sense that $\lim_{t\to\infty} V(f)_t = \infty$.

Definition 2.13. A process S has *finite variation* if there is a measurable set $N \in \mathcal{F}$ with $P(N) = 0$ such that $\forall\omega \in N^c$ the trajectories $t \mapsto S_t(\omega)$ are functions of finite variation.

2.4. **Stochastic Processes as Measures on Path Space.** Kolmogorov's extension theorem, Theorem 2.2, provides a construction of a probability measure P on the measurable space $(\mathbb{R}^{[0,\infty[}, \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[})$ such that the stochastic process given by the evaluation functionals $S_t(\omega) = \pi_t(\omega) = \omega(t)$ has the prescribed finite-dimensional marginals.

Now we try to take another point of view: regard a stochastic process $(S_t)_{t\in[0,\infty[}$ – defined on a general probability space $(\Omega,\mathcal{F},P)$ – as a random variable taking values in a suitable space of curves, i. e. consider the random variable

$$\omega \mapsto S_\cdot(\omega) = (t \mapsto S_t(\omega)). \tag{2.2}$$

Of course, $S_\cdot$ is always a function, thus $\forall\omega \in \Omega : S_\cdot(\omega) \in \mathbb{R}^{[0,\infty[}$, and we even have $S_\cdot : (\Omega,\mathcal{F}) \to (\mathbb{R}^{[0,\infty[}, \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[})$ is a measurable function. Indeed, measurability of $S_\cdot$ with respect to $(\mathcal{F}, \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[})$ is equivalent to measurability of S_t with respect to $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ for all $t \in [0,\infty[$ – recall that $\mathcal{B}(\mathbb{R})^{\otimes[0,\infty[}$ is nothing else but the σ-field generated by all the projections – and the latter is satisfied by definition of a stochastic process. Therefore, we can define the law $(S_\cdot)_*P$ of $S_\cdot$, which is a probability measure on $(\mathbb{R}^{[0,\infty[}, \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[})$, in fact, the law of $S_\cdot$ is just the

probability measure given in the canonical representation of the process, i. e. the probability measure given by Kolmogorov's extension theorem!

Remark 2.14. In some sense, one wants to use the canonical representation of a stochastic process (as a measure on its path-space), since this contains all necessary information and it is often easier to work with a more concrete space than with the general, abstract probability space. Sometimes, one also uses the additional structure of path space, like the vector space structure.

While the canonical representation of a process is appealing because of its generality, it has severe restrictions: some very natural, interesting subsets of $\mathbb{R}^{[0,\infty[}$ turn out to be non-measurable.

Lemma 2.15. *$\forall A \in \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[}\ \exists I \subset [0,\infty[$ with $|I| \leq \aleph_0$ such that $\forall x \in \mathbb{R}^{[0,\infty[}$ and $\forall y \in A$ we have:*

$$\forall t \in [0,\infty[\colon x(t) = y(t) \Longrightarrow x \in A.$$

Less formal, Lemma 2.15 says that a $\mathcal{B}(\mathbb{R})^{\otimes[0,\infty[}$-measurable set is already determined by countably many time points. For a proof of the lemma see Bauer[1] or Billingsley[3].

Corollary 2.16. *Let $C([0,\infty[) \subset \mathbb{R}^{[0,\infty[}$ denote the set of all continuous functions $f : [0,\infty[\to \mathbb{R}$. Then $C([0,\infty[) \notin \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[}$. In other words: a statement like "the process S almost surely has continuous paths" does not make sense.*

Proof. Assume that $C([0,\infty[)$ is measurable and let y be any continuous function. Let I be the countable set from Lemma 2.15. We can certainly find a non-continuous function $x : [0,\infty[\to \mathbb{R}$ such that $\forall t \in I : x(t) = y(t)$. By assumption, we may conclude that $x \in C([0,\infty[)$, i. e. x is continuous, a contradiction. □

Remark 2.17. If we work with the canonical representation, a statement like "$P\big(C([0,\infty[)\big) = 0.7$" is not well-defined. In the case of continuous processes, we could, of course, pass to the completion. Nevertheless, it is unsatisfactory that sets like the set of continuous functions or the set of separable functions are not measurable.

If we are given a continuous stochastic process S it just does not make sense to work with the full canonical probability space: after all, we know that the (outer) probability of all non-continuous paths is 0, i. e. the process is concentrated on a very small subspace. There are basically two possible approaches to define S as a random variable on the space of continuous functions, both of which lead to the same result. Recall that $C([0,\infty[)$ is a polish space when endowed with the topology of uniform convergence on compact subsets.

Lemma 2.18. *The trace σ-field $C([0,\infty[) \cap \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[} = \{C([0,\infty[) \cap A \mid A \in \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[}\}$ coincides with the Borel σ-field on $C([0,\infty[)$:*

$$C([0,\infty[) \cap \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[} = \mathcal{B}\big(C([0,\infty[)\big).$$

For the proof see Bauer[1]. Now we can either restrict P to $C([0,\infty[)$ by considering the trace probability $P'\big(A \cap C([0,\infty[)\big) = P(A)$ for $A \cap C([0,\infty[)$ or we again consider the map $S_\cdot : \Omega \to C([0,\infty[)$, this time as a random variable taking values in the measurable space $\big(C([0,\infty[), \mathcal{B}(C([0,\infty[))\big)$, and define P' as the law of $S_\cdot$, i. e. $P' = (S_\cdot)_* P$.

Summarizing this subsection, when working with a continuous stochastic process, we can always assume that $\Omega = C([0,\infty[)$ (or $C([0,T])$) endowed with its Borel σ-field.

2.5. **Martingales.** The concepts of martingales is one of the most fundamental concepts in the theory of stochastic processes. Unfortunately, lack of time and space does not allow us to go into details, so this subsection is nothing more but a collection of a few definitions. We start with a probability space $(\Omega, \mathcal{F}, P)$.

Definition 2.19. A *filtration* is an increasing family $(\mathcal{F}_t)_{t\in[0,\infty[}$ (or $t \in [0,T]$, or $t \in \mathbb{N}, \dots$) of sub-σ-fields of $\mathcal{F}$, i. e. $\forall t : \mathcal{F}_t \subset \mathcal{F}$ and for $s < t$ we have $\mathcal{F}_s \subset \mathcal{F}_t$. The quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ is called a *filtered probability space*.
A stochastic process $(S_t)_{t\in[0,\infty[}$ is called *adapted* to the filtration $(\mathcal{F}_t)_{t\in[0,\infty[}$ if S_t is $\mathcal{F}_t$-measurable for all $t \in [0,\infty[$.

Definition 2.20. Given a filtered probability space as in Definition 2.19. A *martingale* is an adapted stochastic process S such that for all $t \in [0,\infty[$ we have $S_t \in L^1(\Omega, \mathcal{F}, P)$ and

$$E(S_t|\mathcal{F}_s) = S_s \text{ a. s. }, \quad \forall s < t.$$

Note that the martingale property depends on the filtration and on the probability measure. The problems whether a martingale remains a martingale or whether a specific stochastic process becomes a martingale under a change of the filtration or a change of the probability measure are interesting and often very difficult.

Remark 2.21. If the index t actually represents time, a filtration is often interpreted as a model of the flow of information: $\mathcal{F}_t$ is thought to represent all the information available at time t – e. g. available to all the traders on the market. Then, a martingale could be interpreted as a *fair game*: the best guess – in the least square sense – of the future value S_T at the present time t is the current value S_t.

Definition 2.22. A *stopping time* is a random variable $\tau : \Omega \to [0,\infty]$ such that $\{\omega \mid \tau(\omega) \le t\} \in \mathcal{F}_t$, $\forall t \in [0,\infty[$. Note that a stopping time may have the value $+\infty$.

In stochastic analysis, one often assumes the following condition on the underlying filtered probability space.

Definition 2.23. A filtered probability space is said to satisfy the *usual conditions* if (1) $(\Omega, \mathcal{F}, P)$ is a complete probability space and $\forall t$ the σ-field $\mathcal{F}_t$ contains all the sets of $\mathcal{F}$ with probability 0 and
(2) The filtration is right-continuous, i. e. $\forall t$

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s.$$

Definition 2.24. A *locale martingale* is a stochastic process S such that there is a sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times with $\tau_n \to \infty$ for $n \to \infty$ such that the stopped process

$$S_t^{\tau_n} = S_{t\wedge\tau_n} = S_{\min(t,\tau_n)}$$

is a martingale for each $n \in \mathbb{N}$. The sequence $(\tau_n)_{n\in\mathbb{N}}$ is called *localizing sequence* for the locale martingale S.

3. Brownian Motion

Brownian motion is named after the British botanist Robert Brown, who first observed the shivery motion of pollen particles in liquid. As a mathematical model, it was first introduced by Thorvald Thiele in 1880. Independently, Louis Bachelier used Brownian motion as a model for stock markets already in his PhD thesis in 1900. 5 years later, Albert Einstein introduced (the mathematical) Brownian motion to physics. Brownian motion is also known as Wiener process after Norbert

Wiener who – together with Paul Lévy – made important contributions to the analysis of Brownian motion in the 1920s.

From now on, we always work with a complete probability space $(\Omega, \mathcal{F}, P)$. Whenever we use a filtered probability space, we assume the usual conditions to be in force, c. f. Definition 2.23.

Definition 3.1. An $\mathbb{R}$-valued stochastic process $B = (B_t)_{t\in[0,\infty[}$ is called *Wiener process* if
(i) B is a *Gaussian process*, i. e. $\forall n \in \mathbb{N}$, $\forall (t_1, \dots, t_n) \in [0,\infty[^n$ the random vector $(B_{t_1}, \dots, B_{t_n})$ is an n-dimensional Gaussian random variable and
(ii) the mean value function satisfies $E(B_t) = 0$, $t \in [0,\infty[$, and the covariance function is given by $E(B_t B_s) = t \wedge s = \min(t, s)$, $s, t \in [0,\infty[$.

Definition 3.1 specifies all the finite-dimensional marginals since a Gaussian distribution is uniquely determined by its first and second moments. Thus, we could immediately construct a Wiener process by appealing to Kolmogorov's extension theorem – the consistency requirements are satisfied. We will, however, give a more constructive, less abstract construction later on.

Proposition 3.2. *(i) The Wiener process has* stationary and independent increments*, i. e.* $B_{t+h} - B_t \sim \mathcal{N}(0, h)$ $\forall t$ $(h > 0)$ *and for* $0 \le s < t \le u < v$ *we have* $B_t - B_s \perp\!\!\!\perp B_v - B_u$.
(ii) The Wiener process has a continuous modification.

Proof. By Gaussianity, $(B_t, B_{t+h}) \sim \mathcal{N}(0, \Sigma)$ with $\Sigma = \left(\begin{smallmatrix} t & t \\ t & t+h \end{smallmatrix}\right)$, implying that $B_{t+h} - B_t$ is again Gaussian and we are only left to check the parameters: $E(B_{t+h} - B_t) = 0$ and $E\big((B_{t+h} - B_t)^2\big) = (t + h) - 2t + t = h$. Thus, $B_{t+h} - B_t \sim \mathcal{N}(0, h)$.

By a similar reasoning we see that $(B_t - B_s, B_v - B_u)$ is 2-dimensional Gaussian random variable with covariance $E\big((B_t - B_s)(B_v - B_u)\big) = t - t - s + s = 0$. As a consequence of Proposition 1.19, we get $B_t - B_s \perp\!\!\!\perp B_v - B_u$.

By induction (or using the Laplace transform), one can prove that

$$E\big((B_t - B_s)^{2k}\big) = \frac{(2k)!}{k!2^k}(t - s)^k$$

for $s < t$ and $k \in \mathbb{N}$. Now apply the Kolmogorov-Čentsov theorem, Theorem 2.12, with $\alpha = 2k$, $\beta = k - 1$. This gives existence of Hölder-continuous paths of order $0 < \gamma < \frac{k-1}{2k}$ and for $k \to \infty$ existence of a Hölder-continuous modification of order $0 < \gamma < \frac{1}{2}$. In particular, there is a continuous modification. □

Henceforth, we will always work with a continuous version of the Wiener process, and we add a third requirement in Definition 3.1:

$$(iii) \quad \forall \omega \in \Omega : \ t \mapsto W_t(\omega) \text{ is continuous.}$$

In stochastic analysis, one needs the richer structure of a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty[}, P\big)$ (satisfying the usual conditions).

Definition 3.3. A *Brownian motion* is an adapted process $B = (B_t)_{t\in[0,\infty[}$ satisfying:
(1) B is a Gaussian process with $E(B_t) = 0$ and $E(B_s B_t) = s \wedge t$ for $s, t \in [0,\infty[$,
(2) all sample paths of B are continuous, i. e. $\forall \omega \in \Omega : \ t \mapsto B_t(\omega)$ is continuous,
(3) $B_t - B_s \perp\!\!\!\perp \mathcal{F}_s$ for all $s, t \in [0,\infty[$, $s < t$.

Remark 3.4. The main difference between a Wiener process and a Brownian motion is adaptedness of Brownian motion and the fact that the Brownian increment $B_t - B_s$ is independent of all the information $\mathcal{F}_s$ at time s. Note that our nomenclature – distinguishing between Wiener processes and Brownian motion – is by no means canonical: for most authors, these two names are synonyms. We want, however,

to stress the aforementioned difference: in stochastic analysis, one needs to work with Brownian motion, not with Wiener processes. An alternative way to stress this difference would be to call a Wiener process an *intrinsic* Brownian motion, since it is determined only by "inner" properties – i. e. by the relations between the distributions at different points in time – whereas Brownian motion has to account for "extern" factors, namely the given filtration.

More precisely, let B be a Wiener process (with continuous paths). Then we may define its *natural filtration* by $\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t)$, the σ-field generated by all the random variables B_s with $s \leq t$. One can show that the completion – with respect to the null sets of $\mathcal{F}$ – satisfies the usual conditions and B is a Brownian motion with respect to this (special) filtered probability space. In this sense, each Wiener process is a Brownian motion. This, of course, is not true for a fixed, given filtration, which might also take other sources of information into account.

To summarize, each Brownian motion is a Wiener process, but the converse is not true. We only have that for each Wiener process there is a filtration such that the Wiener process is a Brownian motion with respect to that filtration.

From now on, we always use the natural filtration of a Wiener process, so from now on we use both names synonymously.

Remark 3.5. It is a rather surprising fact that we do not really need to require Gaussianity. Given a stochastic process S with stationary and independent increments, $S_0 = 0$, $E(S_t) = 0$, $E(S_s S_t) = s \wedge t$ and with continuous sample paths, then S is a Gaussian process and, hence, a Wiener process.

As discussed before, we can interpret the Brownian motion as a probability measure P on the measurable space $\big(C([0,\infty[), \mathcal{B}(C([0,\infty[))\big)$. The probability space $\big(C([0,\infty[), \mathcal{B}(C([0,\infty[)), P\big)$ is called *Wiener space*.

3.1. **Construction of Brownian Motion.** We have already remarked that Kolmogorov's extension theorem guarantees the existence of Brownian motion. Here, we give two alternative, more concrete constructions. As a prerequisite, note that for any distribution we can find a sequence of independent random variables all of which having the given distribution. This fact, which we have already used without much thought, is actually a special case of Kolmogorov's extension theorem: the case of a countable index set, where all the finite-dimensional marginals are given by the product measure of the distribution of the 1-dimensional marginals and where all 1-dimensional marginals are the same.

The first approach approximates Brownian motion by random walks. For simplicity, we restrict ourselves to the construction of a Brownian motion defined on $[0,1]$. General Brownian motion can then be obtained by "sticking together" a sequence of independent Brownian motions on $[0,1]$.

Theorem 3.6 (Donsker's Theorem)**.** *Given an i. i. d. sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with $E(X_n) = 0$ and $E(X_n^2) = 1$, e. g. take $P(X_n = -1) = P(X_n = 1) = \frac{1}{2}$. Now fix $n \in \mathbb{N}$ and define the process $B^{(n)}$ on $[0,1]$ by*

$$B_t^{(n)}(\omega) = \frac{1}{\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sqrt{n}} X_{[nt]+1}(\omega),$$

where $S_m = \sum_{i=1}^m X_i$ and $[x]$ denotes the largest integer smaller than x. In other words, $B^{(n)}$ is a piecewise linear interpolation of the random walk $\frac{1}{\sqrt{n}} S_m$, $m = 0, \ldots, n$. Denote by $B_*^{(n)} P$ the corresponding distribution on $C([0,1])$ and denote by $B_* P$ the distribution of the Wiener process on $[0,1]$. Then

$$B_*^{(n)} P \rightharpoonup B_* P, \quad n \to \infty,$$

again in the sense of weak convergence.

Proof. For a proof of Donsker's Theorem see Billingsley[2]. □

Remark 3.7. Donsker's Theorem is also known as *functional central limit theorem*. Indeed, it resembles the central limit theorem quite a lot, with the law of Brownian motion playing the rôle of the standard normal law.

The following, more abstract approach is appealing since it already prepares the construction of the Itô integral. This time, let $(X_n)_{n\in\mathbb{N}}$ denote an i. i. d. sequence of $\mathcal{N}(0,1)$-distributed random variables. Furthermore, let $(e_n)_{n\in\mathbb{N}}$ denote some ONB of the Hilbert space $L^2\big([0,\infty[, \mathcal{B}([0,\infty[), dt\big)$. We define a map

$$\eta : L^2\big([0,\infty[, \mathcal{B}([0,\infty[), dt\big) \to L^2\big(\Omega, \mathcal{F}, P\big), \quad \eta(e_n) = X_n,\ n \in \mathbb{N}. \tag{3.1}$$

It is not difficult to check that η is well-defined as a linear, continuous map and even gives an isometry in the sense that

$$\forall f, g \in L^2\big([0,\infty[\big) :\ \langle f\,, g\rangle_{L^2([0,\infty[)} = \langle \eta(f)\,, \eta(g)\rangle_{L^2(\Omega)}\,.$$

The image $\eta\big(L^2([0,\infty[)\big)$ of η is a Gaussian subspace of $L^2(\Omega)$, i. e. for $n \in \mathbb{N}$ and $f_1, \dots, f_n \in L^2([0,\infty[)$ the vector $(\eta(f_1), \dots, \eta(f_n))$ is jointly Gaussian with mean value 0. Note that $\mathbf{1}_{[0,t]} \in L^2([0,\infty[)$, $\forall t \in [0,\infty[$, and define $B_t = \eta(\mathbf{1}_{[0,t]})$. The process $(B_t)_{t\in[0,\infty[}$ is Gaussian with

$$E(B_t) = 0$$

and

$$E(B_t B_s) = \big\langle \eta(\mathbf{1}_{[0,t]})\,, \eta(\mathbf{1}_{[0,s]})\big\rangle_{L^2(\Omega)} = \int_0^\infty \mathbf{1}_{[0,t]}(u)\mathbf{1}_{[0,s]}(u)du = t \wedge s$$

for $t, s \in [0,\infty[$. Thus, B defines a Brownian motion. Note that this construction does not immediately give a continuous process: we need to pass to a continuous version.

3.2. **Regularity of Brownian Motion.** Apart from Hölder continuity, the paths of Brownian motion exhibit rather irregular behaviour. We start with a heuristic argument from Billingsley[3].

Given a Brownian motion B and a real number $c > 0$, define the process B' by

$$B'_t(\omega) = \frac{1}{c} B_{c^2 t}(\omega), \quad t \in [0,\infty[.$$

B' is again a Brownian motion. Note that time is contracted with a factor c^2, whereas scale is contracted by c, a relation which is often written as

$$dB_t \approx \sqrt{dt}. \tag{3.2}$$

It seems reasonable that – for c large enough – B should somewhere on $[0,c]$ have a chord with slope exceeding 1 with probability close to 1. But this means that B' has, with the same high probability, a chord with a slope exceeding c on the time-interval $[0, \frac{1}{c}]$. Consequently, Brownian motion should have chords with arbitrarily large slope on every arbitrarily small interval!

We collect some of the path properties of Brownian motion in the next theorem. The proofs are often rather technical and therefore omitted, with the exception of the first property, which will be proved later on.

Theorem 3.8. *(1) The paths of Brownian motion have a. s. infinite variation on each bounded interval. In particular,* $P\big(V(B)_t = +\infty\big) = 1$ *for each* $t \in [0,\infty[$.
(2) The paths of Brownian motion are a. s. nowhere *differentiable.*
(3) Given $\varepsilon > 0$. *B a. s. crosses* 0 *infinitely often on the interval* $[0,\varepsilon]$.

(4) Let $Z(\omega) = \{t \in [0,\infty[\mid B_t(\omega) = 0\}$. *$Z$ is a. s. an unbounded, nowhere dense, perfect set with Lebesgue measure* 0.

Theorem 3.8, (2) means that the pathological case of a continuous function which is nowhere continuous is the generic, usual case of a continuous function, at least if measured by the Wiener measure. Recall that a set is perfect if it is closed and has no isolated points. A set A is nowhere dense if for each open interval I there is an open interval $J \subset I$ such that $J \cap A = \emptyset$.

Given a stochastic process S. For any partition $\Delta = \{0 = t_0 < \cdots < t_n = t\}$ of $[0,t]$ let

$$T_t^{\Delta}(S) = \sum_i (S_{t_{i+1}} - S_{t_i})^2$$

and

$$|\Delta| = \max_{i=0,\ldots,n-1}(t_{i+1} - t_i).$$

Definition 3.9. S is of *finite quadratic variation* if there is a finite process $\langle S\,, S\rangle$ such that for all $t \in [0,\infty[$ and any sequence Δ_n of partitions of $[0,t]$ such that $|\Delta_n| \to 0$ we have

$$\lim_{n\to\infty} T_t^{\Delta_n}(S) = \langle S\,, S\rangle_t\,. \tag{3.3}$$

The process $\langle S\,, S\rangle$ is called the *quadratic variation process* associated with S. The limit should actually be understood as a limit in probability, uniform in t, but for our purposes it is enough to consider limits in $L^2(\Omega)$, separately for each t.

Theorem 3.10. *The Wiener process B has finite quadratic variation*

$$\langle B\,, B\rangle_t = t, \quad \forall t \in [0,\infty[.$$

Moreover, along uniform partitions of the form $\Delta_n = \{0, \frac{1}{2^n}t, \ldots, \frac{2^n-1}{2^n}t, t\}$ *the convergence in (3.3) also holds in the sense of a. s. convergence.*

Proof. We only prove the first part. Given any partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$ and call it Δ. Then

$$\begin{aligned}\big\|T_t^{\Delta}(B) - t\big\|^2_{L^2(\Omega)} &= E\Big[\Big(\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2 - t\Big)^2\Big] \\ &= E\Big[\Big(\sum_{i=0}^{n-1}\big((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)\big)\Big)^2\Big]\end{aligned}$$

Multiplying out the last line gives

$$\big\|T_t^{\Delta}(B) - t\big\|^2_{L^2(\Omega)} = \sum_{i=0}^{n-1} E\Big[\Big((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)\Big)^2\Big] + \tag{3.4}$$

$$+\,2\sum_{0\le i<j\le n-1} E\Big[\Big((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)\Big)\Big((B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j)\Big)\Big].$$

By independence of the increments of Brownian motion, we get

$$\sum_{0\le i<j\le n-1} E\Big[\Big((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)\Big)\Big((B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j)\Big)\Big] =$$

$$= \sum_{0\le i<j\le n-1} E\Big[(B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)\Big] E\Big[(B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j)\Big] = 0.$$

For each $i = 0, \dots, n-1$, the corresponding part of the first term gives, using $E\big((B_{t_{i+1}} - B_{t_i})^4\big) = 3(t_{i+1} - t_i)^2$ and $E\big((B_{t_{i+1}} - B_{t_i})^2\big) = t_{i+1} - t_i$,

$$\begin{aligned} E\Big[\Big((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)\Big)^2\Big] &= \\ &= E\Big[(B_{t_{i+1}} - B_{t_i})^4 - 2(t_{i+1} - t_i)(B_{t_{i+1}} - B_{t_i})^2 + (t_{i+1} - t_i)^2\Big] \\ &= 2(t_{i+1} - t_i)^2. \end{aligned}$$

Inserting these intermediate results into (3.4), we get

$$\big\|T_t^\Delta(B) - t\big\|^2_{L^2(\Omega)} = 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2 \le 2t\,|\Delta| \to 0, \quad \text{for } |\Delta| \to 0. \tag{3.5}$$ □

As promised, we now prove that paths of Brownian motion are a. s. of infinite variation.

Corollary 3.11. *The Wiener process a. s. is of infinite variation on every interval* $[s,t]$, $s < t$.

Proof. Since $\langle B\,,B\rangle_t - \langle B\,,B\rangle_s = t - s$, where convergence of (3.3) is understood in $L^2(\Omega)$, there is a sequence $\Delta_m = \{t_i^m\}$ of partitions of the interval $[s,t]$ such that

$$\lim_{m\to\infty} T^{\Delta_m}(B) = t - s, \quad a.\ s. \tag{3.6}$$

Along this sequence of partitions we may estimate

$$\begin{aligned} \sum_i (B_{t_{i+1}^m} - B_{t_i^m})^2 &\le \max_i \left|B_{t_{i+1}^m} - B_{t_i^m}\right| \times \sum_i \left|B_{t_{i+1}^m} - B_{t_i^m}\right| \\ &\le \max_i \left|B_{t_{i+1}^m} - B_{t_i^m}\right| (V(B)_t - V(B)_s). \end{aligned}$$

The left hand side converges to the positive number $t-s$, whereas on the right hand side $\max_i \left|B_{t_{i+1}^m} - B_{t_i^m}\right| \to 0$ for $m \to \infty$ by uniform continuity of B on $[s,t]$. This is only possible, if the variation of B on $[s,t]$ is a. s. infinite, i. e. $V(B)_t - V(B)_s = \infty$. □

Corollary 3.11 shows that it is not possible to define integrals with respect to Brownian motion in a pathwise sense using Lebesgue-Stieltjes integrals.

3.3. **The Law of Iterated Logarithms.** The law of iterated logarithms gives the long- and short-time asymptotics of the paths of Brownian motion. Note that for a Brownian motion B, the process B' defined by $B'_t = tB_{1/t}$ for $t > 0$ and $B'_0 = 0$ is a Brownian motion. This allows us to infer the long-time behaviour from the short-time behaviour and vice-versa.

The strong law of large numbers implies that

$$\lim_{t\to\infty} \frac{B_t}{t} = 0, \quad \text{a. s.}$$

The law of iterated logarithms gives the precise rate of growth, at least in some sense.

Theorem 3.12. *For a Brownian motion B on $[0,\infty[$, the following equations hold with respect to almost sure convergence.*

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = +1, \qquad \liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = -1$$

$$\limsup_{t\to 0+} \frac{B_t}{\sqrt{2t\log\log(1/t)}} = +1, \qquad \liminf_{t\to 0+} \frac{B_t}{\sqrt{2t\log\log(1/t)}} = -1.$$

Proof. For a proof see Bauer[1] or Revuz-Yor[8]. □

3.4. **Markov Processes.** Before we continue our discussion of Brownian motion, we need to define an important class of processes, the Markov processes. Start with a stochastic process $S = (S_t)_{t\in[0,\infty[}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ which we endow with the natural filtration of S, i. e. with the filtration $\mathcal{F}_t = \sigma(S_s,\, 0 \le s \le t)$. As usual, we restrict ourselves to $\mathbb{R}$-valued processes and remark that a generalization is straightforward.

Definition 3.13. S is a *Markov-process* if and only if for all bounded, measurable functions f

$$E\big(f(S_u)\big|\mathcal{F}_t\big) = E\big(f(S_u)\big|S_t\big), \quad \forall u \ge t \in [0,\infty[.$$

In other words: the conditional distribution of S_u with respect to $\mathcal{F}_t$ is the same as the conditional distribution with respect to S_t.

This means that the process is without memory in the sense that complete knowledge of the process up to time t does not give us more information on the process at some future time $u > t$ than only knowledge of the process at time t, the process has forgotten the past (before t). Alternatively, S is a Markov process if and only if $\mathcal{F}_t$ and $\sigma(S_u,\, u \ge t)$ are conditionally independent given S_t, $\forall t \in [0,\infty[$.

For $x \in \mathbb{R}$, $A \in \mathcal{B}(\mathbb{R})$ and $t < u$ define

$$P_{t,u}(x, A) = E\big(S_u \in A\big|S_t = x\big). \tag{3.7}$$

$P_{t,u}$ is a *Markov kernel*, i. e. for fixed A the function $x \mapsto P_{t,u}(x, A)$ is Borel-measurable, and for fixed x the function $A \mapsto P_{t,u}(x, A)$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We interpret $P_{t,u}(x, A)$ as the probability for the process to be within the set A at time u provided that the process starts at point x at time t.

Furthermore, $P_{t,u}$ is a *transition function*, i. e. it satisfies the *Chapman-Kolmogorov equation*

$$P_{t,u}(x, A) = \int_{\mathbb{R}} P_{v,u}(y, A) P_{t,v}(x, dy), \tag{3.8}$$

for any $t < v < u$.

For the following considerations we need to go back to the canonical representation of the stochastic process S. We choose $\Omega = R^{[0.\infty[}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})^{\otimes[0,\infty[}$ and we may assume that S_t is the evaluation functional on $\mathbb{R}^{[0,\infty[}$, i. e. $S_t(\omega) = \omega(t)$. As before, we choose the natural filtration $(\mathcal{F}_t)_{t\ge0}$. Recall that we identify a stochastic process with a probability measure on $\mathbb{R}^{[0,\infty[}$. Before, we used the probability measure P to define the transition functions $P_{t,u}$. Now we want to go the other way round: given the transition functions, can we construct a Markov process with these transition functions?

Theorem 3.14. *Given family $(P_{t,u})_{0\le t<u<\infty}$ of transition functions, i. e. of Markov kernels satisfying the Chapman-Kolmogorov equation (3.8), and a probability measure μ on $\mathbb{R}$. Then there is a unique probability measure P^μ on the path space $(\Omega, \mathcal{F})$ such that S is a Markov process with transition function $(P_{t,u})_{0\le t<u<\infty}$ and initial distribution μ, i. e.*

$$(S_0)_* P^\mu(A) = P^\mu(S_0 \in A) = \mu(A), \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

In particular, we write P^x for the probability measure on path space such that $P^x(S_0 = x) = 1$ for $x \in \mathbb{R}$.

From now on we assume that the Markov process is *time homogeneous*, i. e. that $P_{t,u}$ only depends on the difference $u - t$ for all $t < u$. With $P_t = P_{0,t}$ we get $P_{t,u} = P_{u-t}$. For a bounded measurable function f define

$$P_t f(x) = \int_{\mathbb{R}} f(y) P_t(x, dy) = E\big(f(S_t)\big|S_0 = x\big), \quad x \in \mathbb{R}, \tag{3.9}$$

with the convention that

$$E\big(f(S_t)\big|S_0 = x\big) = \int_\Omega f(S_t(\omega))P^x(d\omega). \tag{3.10}$$

The Chapman-Kolmogorov equation (3.8) implies that $P_tP_sf = P_{t+s}f$, thus $(P_t)_{t\geq 0}$ defines a semi-group, the *transition semi-group*.

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is a C_0-function – a continuous function vanishing at infinity – if f is continuous and $\lim_{|x|\to\infty} f(x) = 0$. We denote the set of all C_0-functions by $C_0(\mathbb{R})$.

Definition 3.15. A Markov process S is called *Feller process* if
(1) its transition semi-group maps C_0-functions to C_0-functions, symbolically $P_t(C_0(\mathbb{R})) \subset C_0(\mathbb{R})$, and
(2) for each C_0-function f and each $x \in \mathbb{R}$ we have $P_tf(x) \to f(x)$ as $t \to 0+$.

The transition semi-group of a Feller process defines a contraction semi-group on the space $C_0(\mathbb{R})$ in the sense of functional analysis.

Definition 3.16. The *infinitesimal generator* of a Feller process S is the infinitesimal generator A of the corresponding transition semi-group, i. e. the operator $A : \mathcal{D}_A \subset C_0(\mathbb{R}) \to C_0(\mathbb{R})$ defined by

$$Af(x) = \lim_{t\to 0+} \frac{P_tf(x) - f(x)}{t} = \lim_{t\to 0+} \frac{E\big(f(S_t)\big|S_0 = x\big) - f(x)}{t}.$$

Dynkin's formula for the Feller process S says that for each function $f \in \mathcal{D}_A$, the process $(M_t^f)_{t\in[0,\infty[}$ defined by

$$M_t^f = f(S_t) - f(S_0) - \int_0^t Af(S_s)ds$$

is a martingale. Note that the converse is true, too: Given a Feller process S and some function $f \in C_0(\mathbb{R})$. If there is a C_0-function g such that

$$f(S_t) - f(S_0) - \int_0^t g(S_s)ds$$

defines a martingale, then $f \in \mathcal{D}_A$ and $Af = g$. In particular,

$$P_tf(x) = f(x) + \int_0^t P_s(Af)(x)ds.$$

We skip the discussion of Markov processes here, although there is still a lot to say, and return to Brownian motion. The reader is referred to Revuz-Yor[8] for more information on this topic.

3.5. **Brownian Motion as a Markov Process.** We begin by proving that Brownian motion is a Markov process.

Theorem 3.17. *Brownian motion B is a Markov process.*

Proof. Since a distribution is uniquely characterized by its moment generating function, it is enough to show that the moment generating functions – i. e. Laplace transforms – of $P(B_{t+s} \in \cdot|\mathcal{F}_t)$ and $P(B_{t+s} \in \cdot|B_t)$ coincide for $t, s > 0$. For $u \in \mathbb{R}$ calculate

$$\begin{aligned}
E\big(\exp(uB_{t+s})\big|\mathcal{F}_t\big) &= \exp(uB_t)E\big(\exp(u(B_{t+s} - B_t))\big|\mathcal{F}_t\big)\\
&= \exp(uB_t)E\big(\exp(u(B_{t+s} - B_t))\big)\\
&= \exp(uB_t)\exp(u^2s/2)\\
&= \exp(uB_t)E\big(\exp(u(B_{t+s} - B_t))\big|B_t\big)\\
&= E\big(\exp(uB_{t+s})\big|B_t\big).
\end{aligned}$$

□

In fact, we even have for any fixed stopping time τ and measurable bounded function f and $t > 0$

$$E\big(f(B_{\tau+t})\big|\mathcal{F}_\tau\big) = E\big(f(B_{\tau+t})\big|B_\tau\big), \tag{3.11}$$

a fact known as *strong Markov property.* (Recall that the Markov property is (3.11) for deterministic times τ only.)

Corollary 3.18. *Fix $t > 0$ and define a stochastic process B' by*

$$B'_u = B_{t+u} - B_t, \quad u \geq 0.$$

B' is a Brownian motion independent of $\mathcal{F}_t$. The same holds true when t is replaced by some stopping time τ.

Next we want to compute the infinitesimal generator of the Brownian motion. For this we need to understand the conditional distribution $P(B_t \in \cdot|B_0 = x)$, which is only well-defined for $x = 0$. Somehow, we need to harmonize the point of view regarding Markov processes with the point of view regarding Brownian motion.

Recall that we can always understand Brownian motion as a probability measure P on $(C([0,\infty[), \mathcal{B}(C([0,\infty[)))$. Then, $B_t(\omega) = \omega(t)$, $\omega \in C([0,\infty[)$, is nothing else but the evaluation functional at t. For each real number x, define a probability measure P^x on $\Omega = C([0,\infty[)$ by $P^x(A) = P(A - x)$ for $A \in \mathcal{B}(C([0,\infty[))$, i. e. P^x is the law of the process $x + B_t$. The transition semi-group of Brownian motion is now defined by

$$P_t f(x) = E\big(f(x + B_t)\big) = \int_{C([0,\infty[)} f(B_t(\omega))P^x(d\omega). \tag{3.12}$$

Note that P^x is exactly the P^x given by Theorem 3.14 for the initial distribution $B_0 \equiv x$ a. surely.

It turns out that the infinitesimal generator of Brownian motion is $\frac{1}{2}\Delta$, one half times the Laplace operator, which is a reformulation of the fact that the heat kernel is a Gaussian density (modulo the factor $\frac{1}{2}$). More precisely, for a C_0-function f, we have

$$P_t f(x) = E\big(f(x + B_t)\big) = \int_{\mathbb{R}} f(y)\frac{1}{\sqrt{2\pi t}}\exp\big(-\frac{(y-x)^2}{2t}\big)dy \tag{3.13}$$

and it is well-known that the generator of the semi-group (3.13) is $\frac{1}{2}\Delta$. Put differently, we have learned that we can write the solution to the heat equation $\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\Delta u(t,x)$ with $u(0,x) = f(x)$ by

$$u(t,x) = E\big(f(x + B_t)\big). \tag{3.14}$$

We will come back to this later, in a more general setting.

3.6. **Reflection Principle.**

Theorem 3.19 (Reflection principle)**.** *Let τ be a stopping time and define a stochastic process B' by*

$$B'_t(\omega) = \begin{cases} B_t(\omega), & t \leq \tau(\omega) \\ 2B_{\tau(\omega)}(\omega) - B_t(\omega), & t > \tau(\omega) \end{cases}.$$

Then B' is a Brownian motion, too.

For a proof of the reflection principle see Revuz-Yor[8]. The name comes from the observation that B' follows the same paths as B until time τ, then it follows the reflected path along the horizontal line passing through (τ, B_τ), since

$$B'_t - B_\tau = -(B_t - B_\tau).$$

For example, let τ be the first time B takes the value 1, i. e.

$$\tau(\omega) = \inf\{t \in [0, \infty[\mid B_t(\omega) = 1\}$$

with the convention that $\inf \emptyset = \infty$. Note that $P(\tau < \infty) = 1$. Then $B'_t = B_t$ until $B_t = 1$ for the first time. Afterwards, $B'_t = 2 - B_t$.

3.7. **Multidimensional Brownian Motion.** By a d-dimensional Brownian motion we understand a stochastic process $B = (B_t)_{t\in[0,\infty[} = \big((B^1_t, \dots, B^d_t)\big)_{t\in[0,\infty[}$ with values in $\mathbb{R}^d$ such that each of components $(B^1_t)_{t\in[0,\infty[}, \dots, (B^d_t)_{t\in[0,\infty[}$ is a one-dimensional Brownian motion and the Brownian motion $B^1, \dots, B^d$ are independent.

We remark that multi-dimensional Brownian motion does not share all the properties of one-dimensional Brownian motion. For example, one- and two-dimensional Brownian motions are *recurrent*: given an open set $O \subset \mathbb{R}$ (or $O \subset \mathbb{R}^2$) and a one-dimensional (two-dimensional) Brownian motion B, then B a. s. hits O infinitely often. On the other hand for $d \geq 3$, Brownian motion is *transient*: given an open ball $O \subset \mathbb{R}^d$ and a d-dimensional Brownian motion B starting outside O, then there is a positive probability that B never hits O.

All the properties mentioned so far remain, however, true for d-dimensional Brownian motion with minor modifications, e. g. the quadratic variation is $d \times t$ instead of t and the infinitesimal generator is d-dimensional Laplacian times $\frac{1}{2}$.

4. Stochastic Integration

In this section, we introduce the stochastic integral with respect to Brownian motion. For simplicity, we restrict ourselves to the L^2-theory. The strategy closely resembles the strategy for defining the Lebesgue integral for general measures: we first identify a class of processes, for which there is a natural choice for the integral, then we extend it to the closure of this class of simple integrands using limits.

We only define the stochastic integral of real-valued stochastic processes with respect to one-dimensional Brownian motion and remark that the extension to the multi-dimensional case is straightforward.

The standing assumption in this section is that we are given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ and a Brownian motion B with respect to the filtration. We assume the usual conditions to be in force.

4.1. Stochastic Integral for Simple Processes.

Definition 4.1. A stochastic process $H = (H_t)_{t\in[0,\infty[}$ is called *simple predictable process* if it is of the form

$$H_t = \sum_{i=0}^{n-1} H_i \mathbf{1}_{]t_i, t_{i+1}]}(t) \tag{4.1}$$

for some $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, and $H_i \in L^2(\Omega, \mathcal{F}, P)$ with H_i being $\mathcal{F}_{t_i}$-measurable, $i = 0, \dots, n-1$. We denote the space of all simple predictable processes by $\mathcal{E}$.

We define the Itô-integral on $\mathcal{E}$ using Riemannian sums.

Definition 4.2 (Itô integral for simple predictable processes)**.** Let $H \in \mathcal{E}$ with the representation (4.1). Then we define the Itô integral of H with respect to B by

$$I(H) = \sum_{i=0}^{n-1} H_i (B_{t_{i+1}} - B_{t_i}). \tag{4.2}$$

The time dependent Itô integral of H with respect to B evaluated at time t is defined by

$$I(H)_t = \sum_{i=0}^{n-1} H_i(B_{t_{i+1}\wedge t} - B_{t_i\wedge t}) = I(H\mathbf{1}_{[0,t]}). \tag{4.3}$$

We also introduce the notations

$$I(H) = \int_0^\infty H_s dB_s = H \cdot B, \quad I(H)_t = \int_0^t H_s dB_s = (H \cdot B)_t.$$

We collect some properties of the stochastic integral for simple predictable processes.

Lemma 4.3. *For all $H \in \mathrm{E}$ the process $(I(H)_t)_{t\in[0,\infty[}$ is a continuous martingale, which is linear in H. We have the following formulas:*

$$E\Big(\int_0^\infty H_s dB_s\Big) = 0 \tag{4.4}$$

and

$$E\Big[\Big(\int_0^\infty H_s dB_s\Big)^2\Big] = \int_0^\infty E\big(H_s^2\big) ds. \tag{4.5}$$

Proof. We start with the proof of equation (4.4). Let $H \in \mathcal{E}$ have the representation (4.1). Then,

$$\begin{aligned} E\Big(\int_0^\infty H_s dB_s\Big) &= E\Big(\sum_{i=0}^{n-1} H_i(B_{t_{i+1}} - B_{t_i})\Big) \\ &= \sum_{i=0}^{n-1} E\Big(E(H_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_{t_i})\Big) \\ &= \sum_{i=0}^{n-1} E\Big(H_i E((B_{t_{i+1}} - B_{t_i})|\mathcal{F}_{t_i})\Big) = 0, \end{aligned}$$

where we used $\mathcal{F}_{t_i}$-measurability of H_i and $(B_{t_{i+1}} - B_{t_i}) \perp\!\!\!\perp \mathcal{F}_{t_i}$ in the last step. For the proof of equation (4.5) we proceed in a similar way.

$$\begin{aligned} E\Big[\Big(\int_0^\infty H_s dB_s\Big)^2\Big] &= E\Big[\Big(\sum_{i=0}^{n-1} H_i(B_{t_{i+1}} - B_{t_i})\Big)^2\Big] \\ &= \sum_{i=0}^{n-1} E\Big(H_i^2(B_{t_{i+1}} - B_{t_i})^2\Big) \\ &\quad + 2\sum_{i<j} E\Big(H_i H_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\Big) \\ &= \sum_{i=0}^{n-1} E\Big(H_i^2 E\big((B_{t_{i+1}} - B_{t_i})^2\big|\mathcal{F}_{t_i}\big)\Big) \\ &\quad + 2\sum_{i<j} E\Big(H_i H_j(B_{t_{i+1}} - B_{t_i}) E\big(B_{t_{j+1}} - B_{t_j}\big|\mathcal{F}_{t_j}\big)\Big) \\ &= \sum_{i=0}^{n-1} E(H_i^2)(t_{i+1} - t_i) = \int_0^\infty E(H_s^2) ds, \end{aligned}$$

where we again used independence of $B_{t_{j+1}} - B_{t_j}$ of $\mathcal{F}_{t_j}$ and that H_i, H_j, $B_{t_{i+1}}$ and B_{t_i} are $\mathcal{F}_{t_j}$-measurable for $j > i$.

Continuity and linearity of the stochastic integral are obvious, the martingale property follows from $E(B_{t_{i+1}\wedge t} - B_{t_i\wedge t}|\mathcal{F}_s) = B_{t_{i+1}\wedge s} - B_{t_i\wedge s}$ for $0 \leq t, s < \infty$. $\square$

Remark 4.4. Note that the formulas corresponding to (4.4) and (4.5) for finite time t instead of infinite time follow immediately by replacing H by $H\mathbf{1}_{[0,t]}$.

We can summarize Lemma 4.3 by saying that the Itô-integral

$$I : \mathcal{E} \subset L^2\big([0,\infty[\times\Omega, \mathcal{B}([0,\infty[) \otimes \mathcal{F}, dt \otimes P\big) \to L^2\big(\Omega, \mathcal{F}, P\big)$$

is an isometry. Note that we use the interpretation of a stochastic process H as a map $(t,\omega) \mapsto H_t(\omega)$ here. The extension of the stochastic integral to the norm-closure of $\mathcal{E}$ is then just abstract nonsense, provided we have identified the norm-closure.

4.2. **Extension of the Stochastic Integral.**

Definition 4.5. The *predictable σ-field* $\mathcal{P}$ is the σ-field on $[0,\infty[\times\Omega$ generated by all adapted, left-continuous processes, i. e. by all processes S satisfying

- $\forall t \in [0,\infty[$: $S_t : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable,
- $\forall \omega \in \Omega$: $S_\cdot(\omega) : [0,\infty[\to \mathbb{R}$ is left-continuous (actually, we only need left-continuity on a set of probability 1).

A stochastic process S is called *predictable*, if it is measurable with respect to the predictable σ-field $\mathcal{P}$.

Definition 4.6. A process S defined on $[0,\infty[\times\Omega$ is *progressively measurable*, if $\forall t \in [0,\infty[$, the restriction of S to $[0,t] \times \Omega$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable. The *progressive σ-field* is the σ-field on $[0,\infty[\times\Omega$ generated by all progressively measurable processes. We denote the progressive σ-field by Prog.

Remark 4.7. A process $(S_n)_{n\in\mathbb{N}}$ is called predictable if S_n is $\mathcal{F}_{n-1}$-measurable for each n, and we immediately understand the rationale for the name: we know the value of S at time n already only given the information $\mathcal{F}_{n-1}$ already available at time $n-1$. In the continuous case, the value of a left continuous process S at time t satisfies $S_t = \lim_{\epsilon\to 0+} S_{t-\epsilon}$, so the value is known given the information available immediately before time t, the information known an infinitesimal time befor t, so to say.

In finance, the predictable processes ar often also called *strategies*.

Remark 4.8. In our case, for the purpose of defining the stochastic integral with respect to Brownian motion, we can work with the progressive σ-field. In general, this is not possible and one can only define the stochastic integral for predictable processes. Note that each progressively measurable process and each predictable process is adapted. We have the chain of inclusions

$$\mathcal{P} \subset \text{Prog} \subset \mathcal{B}([0,\infty[) \otimes \mathcal{F}.$$

Example 4.9. Each simple predictable process is adapted and left-continuous, therefore each simple predictable process is predictable, symbolically

$$\mathcal{E} \subset L^2\big([0,\infty[\times\Omega, \mathcal{P}, dt \otimes P\big).$$

Theorem 4.10. $\overline{\mathcal{E}} = L^2\big([0,\infty[\times\Omega, \mathcal{P}, dt \otimes P\big)$*, i. e. the simple predictable processes are dense in the space of predictable, square integrable processes.*

This means, we are able integrate all predictable, square-integrable processes with respect to Brownian motion. For a proof see Teichmann[10]. We use the same notation as introduced in Definition 4.2. The following two theorems follow easily by abstract nonsense or by approximation of predictable processes by simple predictable processes.

Theorem 4.11. *For all $H \in L^2(\mathcal{P})$ the Itô integral $t \mapsto (H \cdot B)_t$ is a continuous L^2-martingale.*

Theorem 4.12 (Itô's Lemma)**.** *For each $H \in L^2([0,\infty[\times\Omega, \mathcal{P}dt \otimes P)$ we have*

$$E\Big(\int_0^\infty H_s dB_s\Big) = 0$$
$$E\Big[\Big(\int_0^\infty H_s dB_s\Big)^2\Big] = \int_0^\infty E\big(H_s^2\big)ds.$$

Given a process $H \in L^2([0,\infty[\times\Omega, \mathcal{P}, dt\otimes P)$ with continuous trajectories. Then we can calculate the Itô integral of H as follows

$$\int_0^t H_s dB_s = \lim_{N\to\infty} \sum_{i=0}^{2^N-1} H_{\frac{ti}{2^N}}(B_{\frac{t(i+1)}{2^N}} - B_{\frac{ti}{2^N}}), \tag{4.6}$$

where the limit is understood in $L^2(\Omega)$.

Remark 4.13. The first possible extension of the stochastic integral is to predictable processes H satisfying

$$P\Big(\int_0^\infty H_s^2 ds < \infty\Big) = 1$$

instead of the stronger square-integrability condition. The stochastic integral for processes like this is well-defined. Obviously, Itô's lemma is no longer valid, and the integrated process

$$t \mapsto \int_0^t H_s dB_s$$

is a local martingale instead of a true martingale.

4.3. **Itô's Formula.** Itô's formula is the change of variables formula for stochastic integration theory. We start with a reminder on the change of variables formula for functions of finite variation. Let $x : \mathbb{R} \to \mathbb{R}$ be a function of finite variation and let $f : \mathbb{R} \to \mathbb{R}$ be a C^1-function. Then the change of variables formula reads

$$df(x(t)) = f'(x(t))dx(t)$$

or in another formulation

$$f(x(t)) - f(x(0)) = \int_0^t f'(x(s))dx(s). \tag{4.7}$$

Example 4.14. The integral of a finite-variation function x with $x(0) = 0$ with respect to itself satisfies

$$\int_0^t x(s)dx(s) = \frac{x(t)^2}{2},$$

since

$$x(t)^2 = \int_0^t 2x(s)\dot{x}(s)ds = 2\int_0^t x(s)dx(s).$$

Let us do the same calculation for Brownian motion. Fix a partition $0 = t_0 < \cdots < t_n = t$ and calculate

$$\begin{aligned}\sum_{i=0}^{n-1} B_{t_i}(B_{t_{i+1}} - B_{t_i}) &= \frac{1}{2}\sum_{i=0}^{n-1}\Big(B_{t_{i+1}}^2 - B_{t_i}^2 - (B_{t_{i+1}} - B_{t_i})^2\Big)\\ &= \frac{1}{2}\sum_{i=0}^{n-1}(B_{t_{i+1}}^2 - B_{t_i}^2) - \frac{1}{2}\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2.\end{aligned}$$

Note that $\sum_{i=0}^{n-1}(B_{t_{i+1}}^2 - B_{t_i}^2)$ is a telescopic sum equal to B_t^2. $\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2$ converges to t in $L^2(\Omega)$ for the fineness of the partition going to 0, as we have

seen while computing the quadratic variation of Brownian motion. Again for the fineness of the partition going to 0, the left hand side converges in $L^2(\Omega)$ to the integral of Brownian motion with respect to itself. So we have shown that

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t). \tag{4.8}$$

How do we get the extra t-term in equation (4.8) as compared to the result for finite-variation functions? Let us first recall why the change of variables formula (4.7) holds in the case of finite variation. Fix time t and a small time-increment $\Delta t > 0$ and assume additional regularity on f and x. By Taylor's formula,

$$\Delta f(x(t)) = f'(x(t))\Delta x(t) + \frac{1}{2}f''(x(t))(\Delta x(t))^2 + o\big((\Delta x(t))^2\big), \tag{4.9}$$

where $\Delta f(x(t)) = f(x(t+\Delta t)) - f(x(t))$ and $\Delta x(t) = x(t+\Delta t) - x(t)$. Again by Taylor's formula we get

$$\Delta x(t) = \dot{x}(t)\Delta t + \ddot{x}(t)(\Delta t)^2 + o\big((\Delta t)^2\big). \tag{4.10}$$

Equation (4.10) shows that $(\Delta x(t))^2 = o(\Delta t)$, so we may neglect the second order term in equation (4.9) because it is very small as compared to the time increment. Formally, this gives the change of variables formula for finite variation functions.

If we try the same reasoning for Brownian motion B instead of x, we recall that ΔB_t is of the order $\sqrt{\Delta t}$, see (3.2). This means that the second order term in (4.9) is of the order of Δt and cannot be neglected. All the higher order terms in Taylor's formula can, however, still be neglected as compared to the time increment. This formal reasoning yields the change of variables formula

$$f(B_t) = f(0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds. \tag{4.11}$$

This is a first special case of *Itô's formula*. Before formulating the general result, we apply formula (4.11) to our above example.

Example 4.15. Let $f(x) = x^2$. Then formula (4.11) applied to the process B_t reads

$$B_t^2 = 2\int_0^t B_s dB_s + \int_0^t ds,$$

which immediately gives the correct result $\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t)$.

Theorem 4.16 (Itô's Formula, one dimensional case)**.** *Let $u, v \in L^2\big([0,\infty[\times\Omega, \mathcal{P}, dt\otimes P\big)$ be two predictable, square-integrable processes, let X_0 be an $\mathcal{F}_0$-measurable, square-integrable random variable. Define the stochastic process X by*

$$X_t = X_0 + \int_0^t u_s ds + \int_0^t v_s dB_s. \tag{4.12}$$

For a function $f : [0,\infty[\times\mathbb{R} \to \mathbb{R}$ which is C^1 in time and C^2 in the space variable, we get

$$f(t, X_t) = f(0, X_0) + \int_0^t f_t(s, X_s)ds + \int_0^t f_x(s, X_s)u_s ds + \\ + \int_0^t f_x(s, X_s)v_s dB_s + \frac{1}{2}\int_0^t f_{xx}(s, X_s)v_s^2 ds,$$

where f_t denotes the derivative of f with respect to the time variable and f_x, f_{xx} denote the first and second derivative of f with respect to the space variable, respectively.

We remark that the formal argument presented above can be made precise: approximate f by C^3-functions in space and u and v by simple predictable processes and write the the Taylor-expansion of order 3 for $f(t, X_t)$. Then one can show that the second order term converges to the one in Itô's formula and that the rest term vanishes in the limit, see Protter[7] for a detailed proof in full generality.

Remark 4.17. Processes of the form (4.12) are also called *Itô processes*. There is an often used shorthand notation for Itô's formula: introduce the notation

$$dX_t = u_t dt + v_t dB_t,$$

then Itô's formula reads

$$df(t, X_t) = f_t(t, X_t) + f_x(t, X_t)dX_t + \frac{1}{2} f_{xx}(t, X_t)(dX_t)^2, \tag{4.13}$$

with the formal multiplication rules "$(dt)^2 = dtdB_t = 0$" and "$(dB_t)^2 = dt$" applied for the calculation of $(dX_t)^2$.

Theorem 4.18 (Itô's Formula, multi-dimensional case)**.** *Let B denote d-dimensional Brownian motion, let u be an n-dimensional predictable, square integrable process and let v be an $\mathbb{R}^{n\times d}$-dimensional predictable, square integrable process (using any norms on $\mathbb{R}^n$ and $\mathbb{R}^{n\times d}$. Furthermore, let X_0 be an n-dimensional, $\mathcal{F}_0$-measurable square integrable random variable and define*

$$X_t = X_0 + \int_0^t u_s ds + \int_0^t v_s dB_s, \tag{4.14}$$

an n-dimensional process ($v_s dB_s$ is to be understood as a matrix-vector multiplication). For a given function $f : [0, \infty[\times\mathbb{R}^n \to \mathbb{R}$ assumed to be C^1 in the time variable and C^2 in the space variables, we get

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)dt + \sum_{i=1}^n \frac{\partial f}{\partial x^i}(t, X_t)dX_t^i + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j}(t, X_t)dX_t^i dX_t^j,$$

where $x = (x^1, \dots, x^n)$, and we use the formal multiplication rules "$(dt)^2 = dtdB_t^i = 0$", $i = 1 \dots, d$, and "$dB_t^i dB_t^j = \delta_{ij}$", $i, j = 1 \dots, d$, with δ_{ij} being the Kronecker δ.

5. Stochastic Differential Equations

In this section we briefly study a special class of Itô process given by so-called stochastic differential equations (SDEs).

5.1. **Existence of Solutions.** We fix a finite time horizon $T > 0$. Let B be a d-dimensional Brownian motion, $a : \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ measurable functions. Sometimes we will call a the drift and σ the volatility. By a *stochastic differential equation* we understand an equation

$$X_t^x = x + \int_0^t a(X_s^x)ds + \int_0^t \sigma(X_s^x)dB_s, \quad t \in [0, T], \tag{5.1}$$

where $x \in \mathbb{R}^n$ is the initial value of the SDE. We also introduce the short-hand notation

$$dX_t^x = a(X_t^x)dt + \sigma(X_t^x)dB_t. \tag{5.2}$$

Of course, it is also possible to consider non-autonomous SDEs, where a and σ depend on x and t. Most of the theorems hold for this case, too, sometimes the conditions need to be slightly adjusted. It is even possible to introduce SDEs where the data are stochastic, i. e. $a = a(\omega, t, x)$, $\sigma = \sigma(\omega, t, x)$, but then the theory gets more difficult.

Definition 5.1. A stochastic process X^x is called *strong solution* to the SDE (5.1) if X^x is predictable, satisfies

$$\int_0^T E\big(|a(X_s^x)|\big)ds + \int_0^T E\big(|\sigma(X_s^x)|^2\big)ds < \infty$$

and

$$X_t^x = x + \int_0^t a(X_s^x)ds + \int_0^t \sigma(X_s^x)dB_s, \quad \forall t \in [0,T],$$

where the equality holds almost surely. Note that $|\cdot|$ denotes some norms on $\mathbb{R}^n$ and $\mathbb{R}^{n\times d}$.

Remark 5.2. There is also the notion of a *weak solution* to the SDE (5.1). A weak solution is a triple consisting of a filtered probability space satisfying the usual conditions, a Brownian motion B defined thereon and, finally, a process X^x defined thereon satisfying the SDE. Intuitively, the idea is that for a strong solution, the probability space and the Brownian motion are fixed and the solution is a function of the Brownian motion in a sense that can be made precise.

On the other hand, the concept of a weak solution is more flexible and allows, for example, also that the Brownian motion is a function of the solution process X, or more generally, both influence each other. Without proof, we remark that the SDE

$$X_t = \int_0^t \operatorname{sign}(X_s)dB_s \tag{5.3}$$

does not have a strong solution, but it has weak solutions. Note that each weak solution of (5.3) is itself a Brownian motion.

Theorem 5.3 (Existence and Uniqueness)**.** *Let the coefficients a and σ be Lipschitz with at most linear growth, i. e. there is a constant $C > 0$ such that*

$$|a(x) - a(y)| + |\sigma(x) - \sigma(y)| \le C\,|x - y|\,, \quad \forall x,\, y \in \mathbb{R}^n \tag{5.4}$$

and

$$|a(x)| + |\sigma(x)| \le C(1 + |x|), \quad \forall x \in \mathbb{R}^n. \tag{5.5}$$

Then there is a unique strong solution X^x of the SDE (5.1)*. This strong solution has continuous paths and satisfies*

$$E\Big(\sup_{0\le t\le T} |X_t^x|^2\Big) < D(1 + |x|^2) \tag{5.6}$$

for some constant D depending only on T and C – and the chosen norms.

For a proof we refer to Øksendal[6].

The conditions of Theorem 5.3 are the same as for the global existence and uniqueness theorem for ODEs. Note that the Lipschitz condition for the volatility can be slightly weakened: it is enough to assume a Hölder condition of order greater or equal $\frac{1}{2}$.

Example 5.4 (Geometric Brownian Motion)**.** Given constants $\mu, \sigma \in \mathbb{R}$, let $S = (S_t)_{t\in[0,T]}$ denote the solution of the SDE

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \tag{5.7}$$

with initial value $S_0 > 0$. It is possible to derive the exact, explicit formula

$$S_t = S_0 \exp\Big((\mu - \frac{1}{2}\sigma^2)t + \sigma B_t\Big) \tag{5.8}$$

using Itô's formula. In particular, for $\mu = 0$, the process

$$\mathcal{E}(\sigma B)_t = \exp\Big(\sigma B_t - \frac{1}{2}\sigma^2 t\Big)$$

satisfies the "exponential" SDE

$$d\mathcal{E}(\sigma B)_t = \sigma \mathcal{E}(\sigma B)_t dB_t.$$

For this reason, $\mathcal{E}(\sigma B)$ is called the "stochastic exponential" of σB. Note the difference between stochastic and ordinary exponential, which is again a consequence of the second order correction term in Itô's formula. The concept of the stochastic exponential can be defined, in exactly the same way, for more general processes such as continuous martingales and plays an important rôle in the theory.

We will come back to the geometric Brownian motion later, when we discuss a more extensive example from mathematical finance.

Although the theory of stochastic differential equations has many similarities with the theory of ordinary differential equations, there are several important differences, many of them caused by the second order term in the stochastic change-of-variables formula, i. e. Itô's formula. We have already exhibited one instance of this behaviour in the previous example, namely the form of the stochastic exponential.

Example 5.5. Let us study the solution to the two-dimensional SDE

$$dX_t = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X_t dB_t,$$

where B denotes a one-dimensional Brownian motion. Note that the driving matrix is a rotation, so we would expect – in analogy to the deterministic case – the solution to stay on the circle around the origin where its starting point is located. let us fix $X_0 = (1,0)^T$. Then we can check that

$$X_t = e^{\frac{t}{2}} \begin{pmatrix} \cos B_t \\ \sin B_t \end{pmatrix},$$

the solution looks like an outwards winding spiral, c. f. Figure 1. The reason for this is that the high variation of the Brownian increments makes the process leave the circle all the time, it drives the solution outwards. If we want the solution to stay on the circle, we have to add an inward-pointing drift. Indeed, the solution to the SDE

$$dY_t = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} Y_t dB_t - \frac{1}{2} Y_t dt$$

with initial value $Y_0 = (1,0)^T$ is given by $Y_t = (\cos B_t, \sin B_t)^T$ and stays on the circle.

5.2. **Markov Property and Infinitesimal Generators.** In the course of this subsection, let us assume we are given functions a and σ satisfying the conditions of Theorem 5.3 and let X^x denote the solution to the SDE (5.1) with initial value $x \in \mathbb{R}^n$.

Theorem 5.6. *The strong solution $X = (X_t^x)_{t\in[0,T], x\in\mathbb{R}^n}$ satisfies the Markov property, i. e. for $0 < s < t \leq T$ and for any bounded, measurable function $f : \mathbb{R}^n \to \mathbb{R}$ we have*

$$E\big(f(X_t^x)\big|\mathcal{F}_s\big) = E\big(f(X_{t-s}^y)\big)\big|_{y=X_s^x},$$

where $(\mathcal{F}_t)_{t\in[0,T]}$, as usual, denotes the natural filtration of Brownian motion.

Remark 5.7. The notation in Theorem 5.6 means the following: Let $h(y) = E(f(X_{t-s}^y))$, which is clearly a (deterministic) function in y. Then form the composition $h \circ X_s^x : \Omega \to \mathbb{R}$.

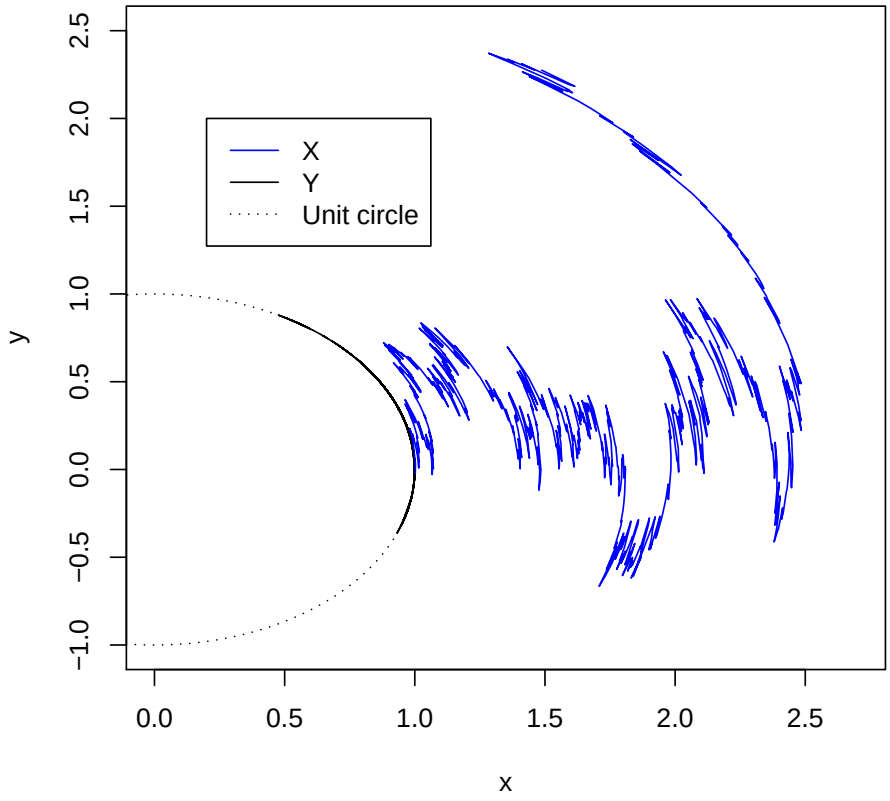

FIGURE 1. Sample path of the processes X and Y of Example 5.5

Proof. We give the proof as in Øksendal[6]. The solution to the SDE (5.1) obviously satisfies

$$X_t^x = X_s^x + \int_s^t a(X_u^x)du + \int_s^t \sigma(X_u^x)dB_u, \tag{5.9}$$

$s < t$. We introduce the notation $X_t^{s,x}$, $s \in [0,T]$, $t \in [s,T]$, for the solution to the SDE (5.1) *started* at time s at position $X_s^{s,x} = x$, i. e. $X_t^{s,x}$ satisfies the equation

$$X_t^{s,x} = x + \int_s^t a(X_u^{s,x})du + \int_s^t \sigma(X_u^{s,x})dB_u, \quad s \le t \le T.$$

Then, (5.9) implies, by uniqueness of solutions, that

$$X_t^x = X_t^{s,X_s^x}.$$

For the sake of clarity, we write $X_t^{s,x}(\omega) = F(x,s,t,\omega)$. By independence of the increments of Brownian motion, we can immediately conclude that the random variable $\omega \mapsto F(x,s,t,\omega)$ is independent of $\mathcal{F}_s$, for any $t \ge s$. Fix s and t and consider the function $g : \mathbb{R}^n \times \Omega \to \mathbb{R}^n$, $g(x,\omega) = f(F(x,s,t,\omega))$. g is $(\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F})$-$\mathcal{B}(\mathbb{R}^n)$-measurable. Consequently – since product-σ-fields are generated by measurable

rectangles – we can approximate g by sums of the form

$$\sum_{k=1}^{m} \phi_k(x)\psi_k(\omega).$$

This allows us to calculate the conditional expectation of $g(X_s^x(\omega),\omega)$ given $\mathcal{F}_s$ as follows.

$$\begin{aligned}
E\big(g\big(X_s^x(\omega),\omega\big)\big|\mathcal{F}_s\big)(\omega) &= E\Big(\lim \sum_k \phi_k\big(X_s^x(\omega)\big)\psi_k(\omega)\Big|\mathcal{F}_s\Big)(\omega) \\
&= \lim \sum_k E\big(\phi_k\big(X_s^x(\omega)\big)\psi_k(\omega)\big|\mathcal{F}_s\big)(\omega) \\
&= \lim \sum_k \phi_k\big(X_s^x(\omega)\big)E\big(\psi_k(\omega)\big|\mathcal{F}_s\big)(\omega) \\
&= \lim \sum_k E\big(\phi_k(y)\psi_k(\omega)\big|\mathcal{F}_s\big)(\omega)\Big|_{y=X_s^x(\omega)} \\
&= E\big(g(y,\omega)\big|\mathcal{F}_s\big)\Big|_{y=X_s^x(\omega)} = E\big(g(y,\omega)\big)\Big|_{y=X_s^x(\omega)}.
\end{aligned}$$

Here we applied several properties of conditional expectations: first we used continuity and linearity, then we used $\mathcal{F}_s$-measurability of $\phi_k\big(X_s^x(\omega)\big)$ and finally we used independence of $g(y,\omega)$ and $\mathcal{F}_s$ for each fixed, deterministic y.

Re-expressing the last equation into our original variables, we have shown the theorem. □

Since X is a Markov process, it has an infinitesimal generator, which we denote by L. Recall that according to Dynkin's formula, a given function $f \in C_0(\mathbb{R}^n)$ is in the domain of L if there is a $g \in C_0(\mathbb{R}^n)$ such that the process

$$f(X_t^x) - f(x) - \int_0^t g(X_s^x)ds$$

is a martingale. In this case we also have $Lf = g$. Assume that f is even in $C_c^2(\mathbb{R}^n)$, the set of twice continuously differentiable functions with compact support. Then we may apply Itô's formula and get

$$\begin{aligned}
f(X_t^x) = f(x) &+ \sum_{i=1}^{n} \int_0^t \frac{\partial f}{\partial x^i}(X_s^x)a^i(X_s^x)ds \\
&+ \sum_{i=1}^{n}\sum_{j=1}^{d} \int_0^t \frac{\partial f}{\partial x^i}(X_s^x)\sigma^{ij}(X_s^x)dB_s^j \\
&+ \frac{1}{2}\int_0^t \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s^x)b^{ij}(X_s^x)ds,
\end{aligned} \tag{5.10}$$

where $b = \sigma\sigma^T : \mathbb{R}^n \to \mathbb{R}^{n\times n}$. Recall that the Itô integral of a square integrable, predictable process is a martingale. Therefore, we get the desired result by moving $f(x)$ and all the dt-integrals to the left hand side in equation (5.10): the remainder process on the right hand side then is a martingale. This proves the following theorem.

Theorem 5.8. *Let L be the infinitesimal generator of the SDE (5.1). Then $C_c^2(\mathbb{R}^n) \subset \mathcal{D}(L)$ and for $f \in C_c^2(\mathbb{R}^n)$ we have*

$$Lf(x) = \sum_{i=1}^{n} a^i(x)\frac{\partial f}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j=1}^{n} b^{ij}(x)\frac{\partial^2 f}{\partial x^i \partial x^j}(x),$$

where $a(x) = (a^1(x), \dots, a^n(x))$ *is the drift and* $b(x) = (b^{ij}(x))_{i,j=1,\dots,n}$ *is the matrix defined by* $b = \sigma\sigma^T$.

5.3. **Feynman Kac Formula.** Now we are finally able to clarify the connections between SDEs and parabolic PDEs. More precisely, a certain class of PDEs can be realized as a functional applied to the solution to a corresponding SDE. We call this functional a "stochastic representation" of the PDE. The stochastic representation can be an important tool both for theoretical analysis of the PDE and for numerical calculation of its solutions, as we shall see later on.

In this subsection, we assume the conditions of Theorem 5.3 to be in force for the functions a and σ. Furthermore, we note that we do not aim for the weakest assumptions possible in the theorem statements. Let $b = \sigma\sigma^T$ and let L be the second order differential operator defined in Theorem 5.8, i. e. L is the infinitesimal generator of the SDE (5.1). We start by giving a stochastic representation for the solution of the Cauchy problem associated to L, i. e. a solution to the problem of finding a function $u : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ satisfying the equation

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = Lu(t,x), \quad t > 0,\ x \in \mathbb{R}^n \\ u(0,x) = f(x), \quad x \in \mathbb{R}^n \end{cases}, \tag{5.11}$$

for a given function $f : \mathbb{R}^n \to \mathbb{R}$. Note that L is understood as a differential operator in x, not in t.

Theorem 5.9 (Kolmogorov's Backward Equation)**.** *Given a function* $f \in C_c^2(\mathbb{R}^n)$ *and denote the solution of* (5.1) *at time* t *with initial valued* x *by* X_t^x*, as usual. Let*

$$u(t,x) = E\big(f(X_t^x)\big), \quad t \in [0,T],\ x \in \mathbb{R}^n. \tag{5.12}$$

Then u *solves the Cauchy problem* (5.11)*. In particular,* $u(t,\cdot) \in \mathcal{D}(L)$ *for each* $t \in [0,T]$*.*

Conversely, given a bounded $C^{1,2}$*-function* v *solution to the Cauchy problem (5.11). Then* $v(t,x) = E\big(f(X_t^x)\big)$*.*

Proof. We first prove that u given by (5.12) solves the PDE (5.11). First note that u is differentiable in t. Indeed, by Dynkin's formula we have $f(X_{t+s}^x) - f(X_t^x) = \int_t^{t+s} Lf(X_u^x)du +$ martingale for $s > 0$ and t. By taking the expectation we get

$$\frac{u(t+s,x) - u(t,x)}{s} = \frac{1}{s}\int_t^{t+s} E(Lf(X_u^x))du,$$

and the limit for $s \to 0+$ exists. Fix t and write $g(x) = u(t,x)$. By Theorem 5.6 we have

$$g(X_s^x) = E\big(f(X_t^y)\big)\big|_{y=X_s^x} = E\big(f(X_{t+s}^x)\big|\mathcal{F}_s\big).$$

Consequently,

$$\begin{aligned} \frac{E\big(g(X_s^x)\big) - g(x)}{s} &= \frac{1}{s}E\Big(E\big(f(X_{t+s}^x)\big|\mathcal{F}_s\big) - E\big(f(X_t^x)\big)\Big) \\ &= \frac{1}{s}\Big(E\big(f(X_{t+s}^x)\big) - E\big(f(X_t^x)\big)\Big) \\ &= \frac{u(t+s,x) - u(t,x)}{s} \xrightarrow{s\to 0+} \frac{\partial u}{\partial t}(t,x), \end{aligned}$$

since $t \mapsto u(t,x)$ is differentiable. This proves that

$$Lu(t,x) = \lim_{s\to 0+} \frac{E\big(g(X_s^x)\big) - g(x)}{s}$$

exists and is equal to $\frac{\partial u}{\partial t}(t,x)$, implying that u is a solution to the Cauchy problem (5.11).

Now we prove the converse statement. The function $v(t,x)$ satisfies the assumptions of Itô's formula. Apply Itô's formula to $v(T-t, X_t^x)$:

$$d_t v(T-t, X_t^x) = -\frac{\partial v}{\partial t}(T-t, X_t^x)dt + Lv(T-t, X_t^x)dt \\ + \sum_{i=1}^{n}\sum_{j=1}^{d} \frac{\partial v}{\partial x^i}(T-t, X_t^x)\sigma^{ij}(X_t^x)dB_t^j.$$

Note that $-\frac{\partial v}{\partial t} + Lv = 0$ because v satisfies the PDE (5.11). Therefore, the right hand side is a martingale with mean 0. Integrating from $t = 0$ to T and taking the expectation, we get – using $v(0, X_T^x) = f(X_T^x)$ –

$$E\big(f(X_T^x)\big) = v(T,x),$$

which shows that the stochastic representation (5.12) holds true for $t = T$. Note that the same argument as above can be carried for T replaced by $t \in]0,T[$, which concludes the proof. □

The *Feynman-Kac-formula* is a generalization of the stochastic representation (5.12).

Theorem 5.10 (Feynman-Kac Formula)**.** *Given $f \in C_c^2(\mathbb{R}^n)$ and $q \in C(\mathbb{R}^n)$ bounded from below. Define*

$$u(t,x) = E\Big(\exp\Big(-\int_0^t q\big(X_s^x\big)ds\Big)f\big(X_t^x\big)\Big), \quad t \in [0,T],\ x \in \mathbb{R}^n. \tag{5.13}$$

u satisfies the PDE

$$\frac{\partial u}{\partial t}(t,x) = Lu(t,x) - q(x)u(t,x) \tag{5.14}$$

with initial condition $u(0,x) = f(x)$, $t \in [0,T]$, $x \in \mathbb{R}^n$.

Conversely, given a bounded $C^{1,2}$-solution v to the PDE (5.14) with initial condition $v(0,x) = f(x)$. Then v satisfies the stochastic representation (5.13).

Remark 5.11. Note that we can also find stochastic representations of the corresponding Cauchy problems for $[0,T]$ replaced by $[0,\infty[$.

Remark 5.12. With a time change $t \to T - t$, (5.11) becomes

$$\frac{\partial u}{\partial t}(t,x) + Lu(t,x) = 0, \quad t \in [0,T],\ x \in \mathbb{R}^n \tag{5.15}$$

with terminal condition $u(T,x) = f(x)$, $x \in \mathbb{R}^n$. This problem has the stochastic representation

$$u(t,x) = E\big(f\big(X_T\big)\big|X_t = x\big) = E\big(f\big(X_T^{t,x}\big)\big). \tag{5.16}$$

In the non-autonomous case, i. e. $a = a(t,x)$, $\sigma = \sigma(t,x)$, we cannot switch between (5.11) and (5.15) as before. In this case, we call (5.15) Kolmogorov's backward equation. Note that in this case we have $L = L_t$, i. e. the operator L is time-dependent.

In the more general case of the Feynman-Kac formula, the time change results in the PDE

$$\frac{\partial u}{\partial t}(t,x) + Lu(t,x) + q(x)u(t,x) = 0, \quad t \in [0,T],\ x \in \mathbb{R}^n \tag{5.17}$$

with terminal condition $u(T,x) = f(x)$, $x \in \mathbb{R}^n$. The stochastic representation in this case reads

(5.18)

$$u(t,x) = E\Big(\Big(-\int_t^T q\big(X_s\big)ds\Big)f\big(X_T\big)\Big|X_t = x\Big) = E\Big(\Big(-\int_t^T q\big(X_s^{t,x}\big)ds\Big)f\big(X_T^{t,x}\big)\Big).$$

In the non-autonomous case, the Feynman-Kac formula is the stochastic representation (5.18) for the PDE (5.17).

In the next theorem, which we present without proof, we indicate how one can find stochastic representations for Dirichlet problems. Note that this theorem is only one out of a big theory. Let X_t^x, $t \in [0,\infty[$, denote the solution of the SDE (5.1) on $[0,\infty[$.

Theorem 5.13. *Let $D \subset \mathbb{R}^n$ be a* domain*, i. e. an open, connected set. For $x \in D$ let τ_D^x denote the first exit time for the process X^x with respect to D, i. e.*

$$\tau_D^x(\omega) = \inf\{t > 0 \mid X_t^x(\omega) \notin D\}.$$

Let ϕ be a bounded, continuous function on ∂D. Assume that the differential operator L is uniformly elliptic*, i. e. the eigenvalues of the matrices $(b^{ij}(x))_{i,j=1,\dots,n}$ are positive and bounded away from 0 for $x \in D$. Define a function*

$$u(x) = E\big(\phi\big(X^x_{\tau_D^x}\big)\big), \quad x \in D. \tag{5.19}$$

Note that u is well-defined since $X^x_{\tau_D^x} \in \partial D$ a. s. by continuity of the solution X_t^x in t.

Then $u \in C^{2+\alpha}(D)$ for all $0 \le \alpha < 1$ and it is a solution to the Dirichlet problem

$$\begin{cases} Lu(x) = 0, & x \in D \\ \lim\limits_{x \to y,\, x \in D} u(x) = \phi(y), & y \in D,\ y \text{ regular} \end{cases}, \tag{5.20}$$

where a point $y \in \partial D$ is called regular *if $P(\tau_D^y = 0) = 1$.*

Note that if L is uniformly elliptic on a neighborhood of $\overline{D}$, then each $y \in \partial D$ is regular.

Recall that Kolmogorov's backward equation is the heat equation for the infinitesimal generator L of X and there is a stochastic representation using the process X. Let L^* denote the (formal) adjoint operator of L, i. e.

$$L^* f(y) = \sum_{i,j=1}^n \frac{\partial^2}{\partial y^i \partial y^j}\big(b^{ij}(y) f(y)\big) - \sum_{i=1}^n \frac{\partial}{\partial y^i}\big(a^i(y) f(y)\big) \tag{5.21}$$

for $f \in C_c^2(\mathbb{R}^n)$. The heat equation for the second order differential operator L^* – which is called *Kolmogorov's forward equation* or *Fokker-Plank equation* – also has a stochastic representation.

Indeed, let us assume that the law of the random variable X_t^x has a density (with respect to Lebesgue measure) for each $t \in]0,T]$ and each $x \in \mathbb{R}^n$. For fixed t and x, we denote this density by $p_t(x,y)$, $y \in \mathbb{R}^n$, i. e.

$$E\big(f\big(X_t^x\big)\big) = \int_{\mathbb{R}^n} f(y) p_t(x,y)\, dy,$$

for bounded measurable functions f. Note that the representation for Kolmogorov's backward equation given in Theorem 5.9 implies that $p_t(x,y)$ is the *fundamental solution* of Kolmogorov's backward equation. Thus, the density P_t is often called *heat kernel*. This already suggests the following theorem.

Theorem 5.14 (Kolmogorov's Forward Equation)**.** *Assume that $y \mapsto p_t(x,y)$ is twice continuously differentiable for each $t \in]0,T]$ and each $x \in \mathbb{R}^n$ and C^1 in t. Then it satisfies the PDE*

$$\frac{\partial p_t}{\partial t}(x,y) = L^* p_t(x,y), \quad t \in]0,T],\ x,y \in \mathbb{R}^n, \tag{5.22}$$

where L^* acts on y.

Remark 5.15. If the operator L is elliptic in the sense of the theory of partial differential equations, then the conditions of Theorem 5.14 are satisfied, i. e. the process X has a smooth density.

Paul Malliavin proved that this is also true if the driving vector fields of the SDE (5.1) satisfy *Hörmander's condition*, i. e. the vector fields and their iterated Lie-brackets span $\mathbb{R}^n$ at each point $x \in \mathbb{R}^n$. In fact, Malliavin gave a probabilistic proof – and generalization – of Hörmander's Theorem on hypo-ellipticity of sum-of-squares operators.

Thus, Hörmander's condition is an algebraic sufficient condition for the assumptions of Theorem 5.14. If the conditions fail, Theorem 5.14 is still true in a disrtibutional sense.

5.4. **Numerics for Stochastic Differential Equations.** As before, denote the solution of (5.1) starting at $x \in \mathbb{R}^n$ by X_t^x, $t \in [0,T]$. In general, we will not be able to give an explicit formula for X_t^x. This is even more true here as in the case of ODEs. Therefore, we need approximation methods. We will discuss the most simple discretization method in more detail – en passant introducing two different notions of approximation for SDEs – and then we give hints for possible generalizations and higher order methods. The reader is referred to Kloeden-Platen[5] for detailed information on this topic.

Obviously, approximations to X_t^x should also be random variables. So our "discrete" approximations will usually only be discrete in time, but "continuous" – i. e. non-discrete – in ω.

The SDE-analogue of the Euler scheme for ODEs is again called *Euler* scheme or *Euler-Maruyama* scheme. For simplicity, we only write it down for equidistant meshes in the autonomous case but remark that one can use more general partitions of the time interval and that it is also applicable for non-autonomous SDEs. Fix a starting value $x \in \mathbb{R}^d$, the number of time steps N, i. e. we work with the uniform mesh $0 = t_0 < t_1 < \cdots < t_N = T$, and define $\Delta t_i = t_{i+1} - t_i$, $\Delta B_i = B_{t_{i+1}} - B_{t_i}$, $i = 0, \ldots, N-1$. Note that $\Delta t_i = \frac{T}{N}$ for each i in the equidistant case. We define the approximation to X^x by $\overline{X}_0^{(N)} = x$ and then recursively by

$$\overline{X}_{i+1}^{(N)} = \overline{X}_i^{(N)} + a(\overline{X}_i^{(N)})\Delta t_i + \sigma(\overline{X}_i^{(N)})\Delta B_i, \tag{5.23}$$

$i = 1, \ldots, N-1$.

Naturally, a numerical approximation should converge to the true solution if we let the number of time steps tend to infinity, symbolically

$$\lim_{N\to\infty} \overline{X}_N^{(N)} = X_T^x. \tag{5.24}$$

In which sense should (5.24) hold? Almost sure convergence seems to be problematic, because in order to check almost sure convergence, one would need to calculate the error for infinitely many ω's, which is virtually impossible. The following two concepts are more tractable, at least in a statistical sense, using the strong law of large numbers and the central limit theorem. Note that these definitions are not restricted to Euler schemes but are valid for any discrete approximation of an SDE.

Definition 5.16. A discrete approximation $\overline{X}_N^{(N)}$ of X^x at time T converges *strongly* to X^x if

$$\lim_{N\to\infty} E\big(\big|X_T^x - \overline{X}_N^{(N)}\big|\big) = 0.$$

Let $\delta_N = \max_{i=0,\ldots,N-1} \Delta t_i$ denote the maximal step size. The method is called to *converge strongly with order* $\gamma > 0$ if there is a positive constant C independent of δ such that

$$E\big(\big|X_T^x - \overline{X}_N^{(N)}\big|\big) \le C(\delta_N)^\gamma,$$

for each N.

Theorem 5.17. *Assume that the data satisfy the condition of the existence and uniqueness theorem, Theorem 5.3. Then the Euler-Maruyama approximation $\overline{X}_N^{(N)}$ converges strongly to X_T^x with order $\frac{1}{2}$.*

Proof. The proof is tedious but straightforward. The interested reader is referred to Kloeden-Platen[5], Theorem 10.2.2. □

Theorem 5.17 is rather disappointing since an order $\frac{1}{2}$ of convergence in a numerical method is quite slow. Note that strong convergence implies almost sure convergence along a subsequence of Ns.

If we want to solve the given SDE because we want to use the stochastic representation of the corresponding heat equation, i. e. Kolmogorov's backward equation, then, by Theorem 5.9, we actually need

$$\lim_{N\to\infty} E\big(f\big(\overline{X}_N^{(N)}\big)\big) = E\big(f\big(X_T^x\big)\big). \tag{5.25}$$

If the SDE is the model of a financial market and we want to calculate the price of a derivative, then, once again, we only need (5.25), as we shall see later. Note that strong convergence implies convergence in the sense of equation (5.25) for all Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}$, but probability theory has a weaker notion of convergence which is suitable for this purpose.

Definition 5.18. The discrete approximation $\overline{X}_N^{(N)}$ of X_T^x is said to *converge weakly* to X_T^x if

$$\lim_{N\to\infty} E\big(f\big(\overline{X}_N^{(N)}\big)\big) = E\big(f\big(X_T^x\big)\big), \quad \forall f \in C_p^l(\mathbb{R}^n),$$

where $C_p^l(\mathbb{R}^n)$ denotes the space of all l-times continuously differentiable functions $\mathbb{R}^n \to \mathbb{R}$ which together with their derivatives of order up to l have polynomial growth. $\overline{X}^{(N)}$ is said to converge to X_T^x with *weak order* $\gamma > 0$ if there is a positive constant C such that

$$\left|E\big(f\big(\overline{X}_N^{(N)}\big)\big) - E\big(f\big(X_T^x\big)\big)\right| \le C(\delta_N)^\gamma, \quad \forall f \in C_p^{2(\gamma+1)}$$

for each N large enough.

Usually, weak convergence is defined with respect to the space $C_b(\mathbb{R}^n)$ of bounded, continuous functions. The advantage of Definition 5.18 is that $C_p^l(\mathbb{R}^n)$ includes all polynomials, and thus one especially important class of functionals f. The definition is taken from Kloeden-Platen[5].

Theorem 5.19. *Let the data a and σ of the SDE be four times continuously differentiable. Then the uniform Euler-Maruyama approximation $\overline{X}_N^{(N)}$ converges weakly of order 1 to the solution X_T^x of the SDE.*

Proof. For the proof see Kloeden-Platen[5], Theorem 14.5.2, where one can also find versions of the theorem with less stringent differentiability assumptions. □

As we have already remarked, often one is not satisfied with an approximation $\overline{X}_N^{(N)}(\omega)$ for $X_T^x(\omega)$, $\omega \in \Omega$. Instead, one needs an approximation for $E(f(X_T^x))$ for some function f. Of course, we could take

$$E\big(f\big(\overline{X}_N^{(N)}\big)\big), \tag{5.26}$$

since strong convergence – and weak convergence, if f is regular enough – implies convergence of the sequence (5.26) to the desired number. In general, we are,

however, not able to explicitly calculate (5.26). For approximation we can use the strong law of large numbers, see Theorem 1.16, i. e. we approximate (5.26) by

$$\frac{1}{M}\sum_{m=1}^{M} f\big(\overline{X}_N^{(N,m)}\big), \tag{5.27}$$

where $(\overline{X}_N^{(N,m)})_{m\in\mathbb{N}}$ denotes a sequence of independent copies of $\overline{X}_N^{(N)}$. Here, N and M are fixed, large enough natural numbers. Note that this introduces another source of error: Indeed, additionally to the error from the approximation (5.26) we also have an error from the approximation of (5.26) by (5.27), i. e.

$$\begin{aligned}\text{Error} &= \Big|E\big(f\big(X_T^x\big)\big) - \frac{1}{M}\sum_{m=1}^{M} f\big(\overline{X}_N^{(N,m)}\big)\Big| \\ &\le \Big|E\big(f\big(X_T^x\big)\big) - E\big(f\big(\overline{X}_N^{(N)}\big)\big)\Big| + \Big|E\big(f\big(\overline{X}_N^{(N)}\big)\big) - \frac{1}{M}\sum_{m=1}^{M} f\big(\overline{X}_N^{(N,m)}\big)\Big| \\ &= \mathrm{E}_{\mathrm{disc}} + \mathrm{E}_{\mathrm{stat}},\end{aligned}$$

with $\mathrm{E}_{\mathrm{disc}}$ being interpreted as the error coming from the time discretization of the SDE and $\mathrm{E}_{\mathrm{stat}}$ as the statistical error coming from the application of the strong law of large numbers. Note that the statistical error is a random variable!

We can control the discretization error $\mathrm{E}_{\mathrm{disc}}$ using Theorem 5.17 or Theorem 5.19 – or rather the underlying error representations. For the control of the statistical error we can use the central limit theorem. Indeed, assume that the random variable $f(\overline{X}_N^{(N)})$ is square integrable, i. e.

$$\sigma_N(f) = \sqrt{E\Big(f\big(\overline{X}_N^{(N)}\big)^2\Big) - E\big(f\big(\overline{X}_N^{(N)}\big)\big)^2} < \infty.$$

Then the central limit theorem, c. f. Theorem 1.14, implies that

$$\sqrt{M}\,\frac{\frac{1}{M}\sum_{m=1}^{M} f\big(\overline{X}_N^{(N,m)}\big) - E\big(f\big(\overline{X}_N^{(N)}\big)\big)}{\sigma_N(f)} \rightharpoonup \mathcal{N}(0,1)$$

for $M\to\infty$. We continue the argument in a heuristic way. Now assume that M is already large enough such that the central limit theorem works, i. e. such that the left hand side of the above formula is already a centered Gaussian random variable for fixed M. This means that the statistical error $\mathrm{E}_{\mathrm{stat}}$ is the absolute value of an $\mathcal{N}(0,\sigma_N(f)^2/M)$-distributed random variable. Now we can control the statistical error as follows: assume we want to make sure that the statistical error satisfies $\mathrm{E}_{\mathrm{stat}}\le\epsilon$ with probability 0.9 for some fixed $\epsilon>0$. Note that for a $\mathcal{N}(0,1)$-distributed random variable Z we have

$$P(|Z|\le 1.65)\approx 0.9$$

and consequently

$$P\Big(|\mathrm{E}_{\mathrm{stat}}|\le 1.65\frac{\sigma_N(f)}{\sqrt{M}}\Big)\approx 0.9. \tag{5.28}$$

Now choose $M\ge\big(1.65\frac{\sigma_N(f)}{\epsilon}\big)^2$. Then the statistical error is with probability 0.9 smaller than ϵ. $\sigma_N(f)$ is not known and needs to be simulated as well.

Remark 5.20. The above method for approximation of an integral – such as the integral $E(f(\overline{X}_N^{(N)})$ – is called *Monte-Carlo simulation*. As we have seen, the error, in the statistical sense of equation (5.28), depends on the number M of samples like $\frac{1}{\sqrt{M}}$, so it is a method of order $\frac{1}{2}$, which is the main disadvantage of Monte-Carlo simulation. On the other hand, note that the dimension of the problem did

not enter into our argument at all. Indeed, the error of Monte-Carlo simulation is independent of the space dimension, which is a distinguished difference to most other methods and often makes Monte-Carlo simulation – and its variants – the only applicable method in high dimensions.

We summarize the *Euler-Monte-Carlo* method:

(1) Fix an error tolerance $\epsilon > 0$ and a probability $0 < \delta < 1$. Split the error tolerance into an error tolerance ϵ_1 for the discretization error and an error tolerance ϵ_2 for the statistical error. Calculate the $(1 - \frac{\delta}{2})$-quantile p_0 for the standard normal distribution, i. e. $P(|Z| \geq p_0) = \delta$ for $Z \sim \mathcal{N}(0, 1)$. Fix N and M.

(2) For m from 1 to M generate random numbers $\eta_1^m, \dots, \eta_N^m$ independent of each other and distributed according to $\mathcal{N}(0, \frac{T}{N} I_d)$. Calculate numbers $\overline{X}_N^{(N,m)}$ according to the rule $\overline{X}_0^{(N,m)} = x$ and

$$\overline{X}_{i+1}^{(N,m)} = \overline{X}_i^{(N,m)} + a\big(\overline{X}_i^{(N,m)}\big)\frac{T}{N} + \sigma\big(\overline{X}_i^{(N,m)}\big)\eta_i^m,$$

for $i = 0, \dots, N-1$.

(3) Calculate the estimates for the discretization error and for $\sigma_N(f)$ – denoted by $\hat{\sigma}_N(f)$ – using these numbers. If the estimate for the discretization error exceeds ϵ_1, increase the number N and go back to (2). If

$$M < \big(p_0 \frac{\hat{\sigma}_N(f)}{\epsilon_2}\big)^2,$$

i. e. if $P(\mathrm{E}_{\mathrm{stat}} \geq \epsilon_2) > \delta$, then increase M and go back to (2).

(4) Calculate and return the approximation value

$$\frac{1}{M}\sum_{m=1}^{M} f\big(\overline{X}_N^{(N,m)}\big).$$

It is also possible to construct higher order methods for the approximation of SDEs, in the sense of strong or weak convergence. In the Euler method, the Brownian motion is the only source of randomness: in each step, randomness only appears in form of the vector $\Delta B_i = (\Delta B_i^1, \dots, \Delta B_i^d)$ and it is easy to generate random numbers according to its distribution for the Monte-Carlo simulation. Already in the simplest higher order method, the *Milstein scheme*, a method of strong and weak order 1, the probabilistic structure is much more complicated. Indeed, in each step we need the Brownian increment and the increments of the iterated integrals of order two, i. e. the whole vector $J_i = (\Delta B_i^1, \dots, \Delta B_i^d, J_i^{11}, J_i^{12}, \dots, J_i^{dd})$, where

$$J_i^{j,l} = \int_{t_i}^{t_{i+1}} (B_s^j - B_{t_i}^j) dB_s^l, \quad j, l = 1, \dots, d.$$

We know that $J_i^{ll} = \frac{1}{2}((\Delta B_i^l)^2 - \Delta t_i)$, $l = 1, \dots, d$, but there is no such formula for the mixed terms ($j \neq l$). In dimension $d = 1$, generation of the random vector J_i is just as simple as generation of the random vector ΔB_i, therefore the Milstein-Monte-Carlo scheme is competitive to the Euler-Maruyama-Monte-Carlo scheme. In general, however, it is not easy to generate random numbers according to the distribution of J_i. In fact, one needs to sample the vector ΔB_i first and then one needs to sample the iterated integrals using Riemannian sums as in the definition of the Itô integral. This makes the Milstein scheme – and all higher order schemes – very expensive compared to the Euler scheme.

Remark 5.21. In some sense, the work from the Monte-Carlo simulation dominates the work from the discretization, and consequently the speed-up of higher order methods are not so dramatic as in the deterministic case. Indeed, assume we use a

method of order $p > 0$, i. e. for each realization ω the number of time steps – which we identify with the total work – is

$$N(\epsilon_1) = C_1 \epsilon_1^{-1/p},$$

for some positive constant C_1 and the time discretization tolerance ϵ_1 as before. In order to get a statistical error of less than ϵ_2 – at least with high probability – we need to compute

$$M(\epsilon_2) = C_2 \epsilon_2^{-2}$$

realizations, where C_2 is a positive constant. Now let $\epsilon > 0$ be our tolerance and, consequently, $\epsilon_1 = \lambda\epsilon$, $\epsilon_2 = (1-\lambda)\epsilon$ for some $0 < \lambda < 1$. The total work is

$$A(\lambda) = N(\epsilon_1)M(\epsilon_2) = C_1 C_2 \lambda^{-1/p}(1-\lambda)^{-2}\epsilon^{-1/p-2}. \tag{5.29}$$

This shows that the total work is proportional to $\epsilon^{-(2+\frac{1}{p})}$. At least for $p > 1$, the gain of using a higher order method becomes increasingly insignificant for the total work. A simple optimization shows that one should choose $\lambda = \frac{1}{2p+1}$. For example, if one uses a method of order one, than one third of the total tolerance should be reserved for the discretization error, the rest for the statistical error. A more serious analysis would require taking the dependence of the constants C_1 and C_2 on p into account.

5.5. **An Example: The Black-Scholes Model.** The Black-Scholes model is a model for a financial market with one stock and a constant interest rate. Its main advantage is that explicit pricing rules for several important options and hedging strategies are available. The model goes back to Fischer Black and Myron Scholes, using previous work by Paul Samuelson and Robert Merton. Scholes and Merton received the Nobel price for Economy in 1997, Black had died 2 years earlier while Samuelson had already got the Nobel prize earlier.

We are given a constant interest rate $r > 0$. We use continuous compounding, therefore the bank account – with r as interest rate – evolves such that one euro at time 0 becomes e^{rt} euros at time t. The stock is modeled as a geometric Brownian motion, see Example (5.7), i. e.

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \tag{5.30}$$

where μ and σ are constants called *drift* and *volatility*, respectively. B is a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$. The starting value is $S_0 > 0$. The explicit formula (5.8) shows that $S_t > 0$ for all t, as should be the case for the price of a stock.

In financial mathematics – especially in the case of constant or deterministic interest rates – one often studies the *discounted* variables, i. e. $\widetilde{S}_t = e^{-rt}S_t$. In some sense, this can be understood as a change of currency – the technical term is *change of numeraire* – in which the financial market is traded. Indeed, in our example we change the currency from euros to "discounted euros", which have the same value as euros today and change their values in the future in such a way that the interest rate – in discounted euros – is 0. The SDE for $\widetilde{S}$ is easily seen to be

$$d\widetilde{S}_t = (\mu - r)\widetilde{S}_t dt + \sigma \widetilde{S}_t dB_t. \tag{5.31}$$

This shows that it is no loss of generality to assume $r = 0$, which means starting with the discounted equation.

We want to calculate the price of a *European call option*, i. e. a contract giving the owner (of the option) at the prescribed time T the right to buy one stock for the prescribed price K. T is called *maturity* and K is the strike price. At time T, the owner takes a look at the quotes for the stock. If $S_T(\omega) > K$, then he or she exercises the option and buys the stock, since he only has to pay K euros instead

of $S_T(\omega)$ euros, and immediately sells it, giving her $S_T(\omega)$ euros. If the price of the stock is smaller than K, then exercising the option does not make sense. In any case, the owner makes a profit of $(S_T - K)_+$ euros at time T. This is called the payoff of the option.

The European put option is similar, instead of giving the right to buy the option for the strike price, it gives the right to sell it. The payoff of the European put option consequently is $(K - S_T)_+$.

Put and call options are the most basic derivatives on stocks. There are much more complicated ones depending on the average of the stock price in some period of time, its maximum or minimum prices and so on. A European option is an option which can only be exercised at one prescribed date. An option which can be exercised at any time prior to a prescribed date is called *American* option. Pricing of American options is much more complicated since it contains a (stochastic) optimization problem. Therefore, we will concentrate on the European call and put options.

It seems natural to determine the current price (at time 0) of a European call option with maturity T and strike price K by

$$C(0, S_0) = E(e^{-rT}(S_T - K)_+).$$

More general, $C(t, x)$ denotes the price of the option at time t subject to the condition that $S_t = x$ (note that this is a real number, not a random variable!) This is, however, not the correct formula for the price, because we may not use the measure P here – P is often called the *physical measure*. Instead, we need to use the so-called *risk-neutral measure* or *martingale measure*. If we used the physical measure P for pricing options and introduce the options as new assets in the market, then the market would contain *arbitrage opportunities*, i. e. it would be possible to construct a (self-financing) portfolio with value $V_t(\omega)$ at time t such that $P(V_0 = 0) = 1$ and $P(V_T \geq 0) = 1$ but $P(V_T > 0) > 0$. It is the most fundamental assumption of financial mathematics that there are no arbitrage opportunities: it is not possible to make riskless profit!

In the case of the Black-Scholes model, the SDE of the stock under the martingale probability measure Q is

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{5.32}$$

where W denotes a Brownian motion on $(\Omega, \mathcal{F}, Q)$. Note the surprising fact that equation (5.32) – and consequently the option prices – does not depend on the drift μ any longer. Instead, the risk-free interest rate plays the rôle of the stock. This gives an intuitive explanation for the term risk-neutral measure: under the physical measure, one typically has $\mu > r$, since the investor in the stock takes more risk than the investor on the bank account, thus he gets – on average – a higher profit. The difference $\mu - r$ is called *risk premium*. Under the risk-neutral measure Q, the investor is no longer risk avers, she only looks at the expected profit and disregards the risk. Therefore, the risk premium is 0 and the drift coincides with the interest rate.

Under Q, the stock price satisfies

$$S_t = S_0 \exp\Big((r - \frac{\sigma^2}{2})t + \sigma W_t\Big)$$

and the price of the European call is given by

$$C(0, S_0) = e^{-rT} E_Q\big((S_T - K)_+\big),$$

where E_Q denotes the integral with respect to Q. By elementary transformations we get the celebrated *Black-Scholes formula*

$$C(0, S_0) = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{5.33}$$

where $\Phi(x) = P(Z \leq x)$ for $Z \sim \mathcal{N}(0,1)$ denotes the distribution function of the standard normal measure and

$$d_1 = \frac{\log\left(\frac{S_0}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$
$$d_2 = \frac{\log\left(\frac{S_0}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}.$$

The value of the option at time t is found by replacing S_0 by the stock price S_t at time t and T by $T - t$.

On the other hand, the Feynman-Kac formula implies that $C = C(t, x)$ is solution to the PDE

$$\frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C}{\partial x^2}(t,x) + rx\frac{\partial C}{\partial x}(t,x) + \frac{\partial C}{\partial t}(t,x) - rC(t,x) = 0, \tag{5.34}$$

$t \in [0, T]$, $x > 0$. The boundary condition is given by $C(T, x) = (x - K)_+$, which also makes sense from the economic point of view. The PDE (5.34) is known as *Black-Scholes PDE*. Black and Scholes originally derived their PDE out of economical considerations and found their formula by solving the PDE using the Fourier transform method.

References

[1] Heinz Bauer, *Wahrscheinlichkeitstheorie*, de Gruyter, Berlin, 2002.

[2] Patrick Billingsley, *Convergence of Probability Measures*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1968.

[3] Patrick Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1995.

[4] Fima C. Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, London, 2001.

[5] Peter E. Kloeden and Eckhard Platen, *Numerical Solutions of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability 23, Springer-Verlag, Berlin, 1999.

[6] Bernt Øksendal, *Stochastic Differential Equations. An Introduction with Applications.*, Universitext, Springer-Verlag, Berlin, 2000.

[7] Philip E. Protter, *Stochastic Integration and Differential Equations*, Stochastic Modelling and Applied Probability 21, Springer-Verlag, Berlin, 2005.

[8] Daniel Revuz and Marc Yor, *Continuous Martingales and Brownian Motion*, Grundlehren der mathematischen Wissenschaften 293, Springer-Verlag, Berlin, 2005.

[9] Jordan Stoyanov, *Counterexamples in Probability*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Chichester, 1997.

[10] Josef Teichmann, *Stochastic Analysis with Applications to Financial and Actuarial Mathematics*, Lecture Notes, 2003.