1 Brownian motion

Brownian motion as a physical phenomenon was discovered by botanist Robert Brown as he observed a chaotic motion of particles suspended in water. The rigorous mathematical model of BM was introduced by Norbert Wiener, who also gave the first proof that BM (i.e. a stochastic process satisfying formal conditions) exists. In recognition of Wiener's contribution the BM is sometimes called the Wiener process.

In this section we discuss properties of the Brownian motion, which is the basic building block in stochastic analysis and its applications to financial mathematics and other fields. The geometric Brownian motion, often used to model the stock prices, is easily constructed from the BM.

1.1 Definition of BM and its finite-dimensional distributions

Preliminaries Events, random variables and stochastic processes will be defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of *elementary outcomes* and $\mathcal{F}$ is a σ-algebra[1] of *events* $A \subset \Omega$ to which probability $\mathbb{P}(A)$ can be assigned. We may think of ω as a complete description of a random experiment, and interpret the relation $\omega \in A$ as 'the event A occurs'.

The set Ω may be chosen finite if a random experiment has finitely many outcomes (e.g. tossing two dice); in this situation we usually take for $\mathcal{F}$ the power set 2^{Ω}. If the experiment is complex, like observation of stock prices in continuous time, we may choose for Ω a space of functions (then for measure-theoretic reasons the probability cannot be assigned to every subset of Ω). For most considerations, however, the nature of Ω is not important, provided Ω is rich enough to accomodate all random objects with desirable properties.

Unless otherwise specified, under a stochastic process we shall mean a random real-valued function $(X(t), t \geq 0)$ of a continuous time parameter t, where t takes values in some finite or infinite interval, like $[0, T]$ or $[0, \infty)$.

For every n and times $t_1 < \cdots < t_n$ the random variables $X(t_1), \ldots, X(t_n)$ have some joint probability distribution with distribution function

$$F_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) = \mathbb{P}(X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n).$$

These multivariate distributions are called *finite-dimensional distributions of the process* $(X(t), t \geq 0)$. If there exists a joint density $f_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n)$ we can also represent the joint distribution as

$$F_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{X(t_1),\ldots,X(t_n)}(y_1, \ldots, y_n) dy_1 \ldots dy_n.$$

More generally, for $D \subset \mathbb{R}^n$

$$\mathbb{P}((X_1, \ldots, X_n) \in D) = \int_D f_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

[1] This means that $\mathcal{F}$ is closed under set-theoretic operations $\cap, \cup, ^c$ performed on countably many sets $A_i \in \mathcal{F}$. A minimal system of axioms of σ-algebra is (i) $\Omega \in \mathcal{F}$; (ii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$; (iii) $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$ implies $\cup_i A_i \in \mathcal{F}$.

(n-dimensional integral).

Definition 1.1. A stochastic process $B = (B(t), t \geq 0)$ is called a (standard) *Brownian motion* if it has the following properties:

(i) $B(0) = 0$.

(ii) $B(t)$ is a continuous function of t.

(iii) For $0 \leq s < t$ the increment $B(t) - B(s)$ has normal distribution $\mathcal{N}(0, t-s)$.

(iv) For any finite set of times $0 = t_0 < t_1 < \cdots < t_n$ the increments

$$B(t_1) - B(t_0), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$$

are independent.

For every fixed t, $B(t)$ is not a number rather a random variable, so the full notation indicating the dependence on the elementary outcome $\omega \in \Omega$ would be $B(t, \omega)$. For fixed ω, the function $t \mapsto B(t, \omega)$ is often called a *path* or *trajectory* of the process. The continuity of paths is required in the sense 'almost surely' (abbreviated a.s.), meaning that property (ii) must hold for all ω with exception of an event of probability zero.

There are various explicit constructions of the BM via approximation by simpler processes. In particular, the BM can be obtained as a limit of scaled discrete-time random walks. The symmetric random walk starts at 0 and moves by steps ± 1 with equal probabilities at times $1, 2, \ldots$. If we modify the random walk by scaling the steps as $\pm 1/\sqrt{n}$ and let the moves occur at times i/n ($i = 1, 2, \ldots$), then for large n the increments will be approximately as in (iii), by independence of jumps and the central limit theorem. The latter suggests to think of BM as a continuous-time random walk of a particle which undergoes many small dislocations in every time interval.

Finite-dimensional distributions From Definition 1.1 we can easily derive the finite-dimensional distributions of BM, which is the joint distribution of random variables $B(t_1), B(t_2), \ldots, B(t_n)$ for arbitrary choice of n and the times $t_1 < \cdots < t_n$.

The one-dimensional distributions are obvious. By (i) and (ii) $B(t) = B(t) - B(0)$ has $\mathcal{N}(0, t)$ distribution with density

$$f_{B(t)}(x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right), \quad x \in \mathbb{R}.$$

To obtain the general finite-dimensional distributions of BM we need to recall some facts about the multivariate normal distribution. A n-dimensional random vector $\boldsymbol{V}$ (written as a row) is said to have the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Lambda)$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Λ, if the density of $\boldsymbol{V}$ is given by

$$f_{\boldsymbol{V}}(\boldsymbol{x}) = (2\pi \det \Lambda)^{-n/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})\Lambda^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^T\right), \quad \boldsymbol{x} \in \mathbb{R}^n.$$

A special property of the multivariate normal distribution is that the components of $\boldsymbol{V}$ are independent if and only if they are uncorrelated (then Λ is a diagonal matrix). If $\boldsymbol{V}$ has distribution $\mathcal{N}(\boldsymbol{\mu}, \Lambda)$, then for fixed vector $\boldsymbol{w}$ and fixed $n \times n$ matrix M the random vector $\boldsymbol{W} = \boldsymbol{w} + \boldsymbol{V}M$ has a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu} + \boldsymbol{w}M, M\Lambda M^T)$.

For $0 \le s < t$ the covariance of $B(s)$ and $B(t)$ is computed as

$$\begin{aligned}\mathrm{Cov}(B(s), B(t)) = \mathbb{E}[B(s)B(t)] = \mathbb{E}[B(s)(B(t) - B(s)) + B^2(s)] = \\ \mathbb{E}[B(s)]\mathbb{E}[(B(t) - B(s))] + \mathbb{E}[B^2(s)] = 0 + \mathrm{Var}[B(s)] = s,\end{aligned}$$

where (iii) and (iv) were used. For general s, t the covariance is $\mathrm{Cov}(B(s), B(t)) = s \wedge t$, where $s \wedge t$ is a shorthand notation for $\min(s, t)$.

Proposition 1.2. *For any $0 = t_0 < t_1 < \cdots < t_n$ the joint distribution of $B(t_1), B(t_2), \ldots, B(t_n)$ is multivariate normal with mean vector zero and the covariance matrix with coefficients $t_i \wedge t_j$ (for $1 \le i, j \le n$).*

Proof. We will use the linear transformation property of multivariate normal distributions.

First note that the joint distribution of the increments $B(t_1) - B(t_0), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ is a multivariate normal distribution, because each $B(t_i) - B(t_{i-1})$ has a one-dimensional normal distribution $\mathcal{N}(0, t_i - t_{i-1})$ and the increments are independent. Then observe that we can pass from the vector of increments to the vector $(B(t_1), B(t_2), \ldots, B(t_n))$ by a linear transformation

$$B(t_i) = [B(t_i) - B(t_{i-1})] + [B(t_{i-1}) - B(t_{i-2})] + \cdots + [B(t_1) - B(t_0)], \quad i = 1, \ldots, n,$$

hence the joint distribution of $(B(t_1), B(t_2), \ldots, B(t_n))$ is normal. □

Transition probabiliities We have seen that the vectors

$$(B(t_1), B(t_2), \ldots, B(t_n)) \text{ and } (B(t_1) - B(t_0), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1}))$$

(where $0 = t_0 < t_1 < \cdots < t_n$) contain the same information, for if we know one of them we can compute the other. This observation is often helpful for various computations related to the BM. In particular, we can calculate the *transition probabilities*:

$$\begin{aligned}\mathbb{P}(B(t_n) \le x_n | B(t_i) = x_i, 1 \le i \le n-1) = \\ \mathbb{P}(B(t_n) - B(t_{n-1}) \le x_n - x_{n-1} | B(t_i) - B(t_{i-1}) = x_i - x_{i-1}, 1 \le i \le n-1) = \\ \mathbb{P}(B(t_n) - B(t_{n-1}) \le x_n - x_{n-1}) = \\ \int_{-\infty}^{x_n - x_{n-1}} \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} \exp\left(-\frac{y^2}{2(t_n - t_{n-1})}\right) dy\end{aligned} \tag{1}$$

Let us think of given $s > 0$ as present time, and assume that the BM is in state $B(s) = x$. Changing in (1) notation as $t = t_n, s = t_{n-1}, x = x_{n-1}, y = x_n, k = n - 2$, the computation (1) shows that the conditional density of a future state $B(t)$, $t > s$, given any past states $B(t_1) = x_1, \ldots, B(t_k) = x_k$ at times $t_1 < \cdots < t_k < s$ and given the present state $B(s) = x$, does not depend on these past states:

$$f_{B(t)|B(s)=x}(y) = f_{B(t)|B(s)=x, B(t_1)=x_1, \ldots, B(t_k)=x_k}(y), \qquad s < t. \tag{2}$$

This conditional density $f_{B(t)|B(s)=x}(y)$ depends on t, s through the difference $\tau := t - s$, and is called the *transition density*

$$p(\tau, x, y) = \frac{1}{\sqrt{2\pi(\tau)}} \exp\left(-\frac{(x-y)^2}{2\tau}\right). \tag{3}$$

The probability that the BM moves in time τ from state x to some point in the interval $[a, b]$ is obtained by integration

$$\mathbb{P}(B(t+\tau) \in [a,b] | B(t) = x) = \int_a^b p(\tau, x, y) dy,$$

and this probability is not affected by conditioning on the states passed (i.e. values assumed) by the BM before time t.

Independence of the future $B(t)$ on the past states $B(t_1), \cdots, B(t_k)$, given the present state $B(s) = x$ (where $t_1 < \cdots < t_k < s < t$) (as formally expressed in (2)) is the pattern of *Markov property* of BM. We shall return to this property after a more rigorous definition of the 'past' of BM prior to some given time.

Markov property allows us to to write the joint probability density of $B(t_1), \ldots, B(t_n)$ in terms of the transition density (3) as the product

$$f_{B(t_1),\ldots,B(t_n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(t_i - t_{i-1}, x_{i-1}, x_i).$$

Note that $p(t, 0, x)$ is the density function of $B(t)$ (the normal $\mathcal{N}(0, t)$-density).

1.2 Filtration for Brownian motion

Elements of the basic σ-algebra $\mathcal{F}$ are events A, to which probability $\mathbb{P}(A)$ can be assigned. We may interpret $\mathcal{F}$ as all thinkable events A about which we can say, in principle, whether A occurs ($\omega \in A$) or not ($\omega \notin A$). To organise information about stages of a random experiment it is often important to consider sub σ-algebras $\mathcal{G} \subset \mathcal{F}$ which record some restricted or partial information. For instance, in the experiment of rolling two dice, the information about the first roll corresponds to a sub σ-algebra. The larger $\mathcal{G}$, the more information we have, i.e. the larger the class of events about which we can say that they occur or not.

In the theory of random processes we consider increasing families of σ-algebras (sub σ-algebras of $\mathcal{F}$), to capture the idea of accumulation of information as the time progresses.

A family $(\mathcal{F}_t, t \geq 0)$ of sub σ-algebras of $\mathcal{F}$ is called a *filtration* if $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s \leq t$. Intuitively, $\mathcal{F}_t$ is all information gained by the observer prior to time t, for instance the historical stock prices up to date.

Terminology: we say that event A is $\mathcal{F}_t$-measurable if $A \in \mathcal{F}_t$; and that a random variable ξ is $\mathcal{F}_t$-measurable if the event $\{\omega \in \Omega : \xi(\omega) \in [a, b]\}$ is $\mathcal{F}_t$-measurable for every interval $[a, b] \subset \mathbb{R}$.

Definition 1.3. We say that a stochastic process $X = (X(t), t \geq 0)$ is *adapted* to the filtration $(\mathcal{F}_t, t \geq 0)$ if the random variable $X(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$.

That is to say, the information available at time t is sufficient to evaluate $X(t)$.

With every stochastic process $X = (X(t), t \geq 0)$ one can always associate the *natural* filtration $(\mathcal{F}_t^X, t \geq 0)$, which records the past behaviour of X, but bears no extra information. This is the filtration in which $\mathcal{F}_t$ is the smallest σ-algebra containing the events $\{X(s) \in [a,b]\}$ for $s \leq t$ and $[a,b] \subset \mathbb{R}$. If $A \in \mathcal{F}_t^X$ this means that based on the observation of the piece of the path $(X(s), 0 \leq s \leq t)$ it is possible to find out whether A has occured or not.

A filtration is often included in the definition of a stochastic process (adapted to the filtration). If this is not specified, the the natural filtration is meant.

Definition 1.4. We say that $(\mathcal{F}_t, t \geq 0)$ is a filtration for the Brownian motion $B = (B_t, t \geq 0)$ if

(i) The BM is adapted, that is $B(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$.

(ii) The increment $B(t) - B(s)$ is independent of $\mathcal{F}_s$ for $0 \leq s < t$.

The natural filtration $(\mathcal{F}_t^B, t \geq 0)$ of the BM is an example. But sometimes richer filtrations are needed, e.g. when we consider another stochastic processes or random variables which are not just computable from the BM (e.g. independent of BM). The additional information (above that in $\mathcal{F}_t^B$) is not allowed to give clues about the future behaviour of the BM (property (ii)).

Example 1.5. Let $M(2) = \max\{B(t), t \leq 2\}$ be the largest value of BM on the time interval $[0,2]$. The event $\{M(2) > 5\}$ is $\mathcal{F}_3$-measurable, because once we know the path before $t = 3$, we can determine if $M(2) > 5$ or not. The event $\{M(2) > 5\}$ is not $\mathcal{F}_1$-measurable.

The process $X(t) = B(t/2)$ is adapted to $(\mathcal{F}_t^B, t \geq 0)$ (at every time t we know $(B(s), 0 \leq s \leq t)$ and $B(t/2)$ in particular), while $Y(t) = B(2t)$ not.

1.3 Martingale property

The standard BM starts with $B(0) = 0$ and on the average stays at 0, because $\mathbb{E}B(t) = 0$. Similarly, given $B(s) = x$ at future times $t > s$ BM will stay on the average at x, whichever the past. Such processes which tend to go neither up nor down are called martingales.

Definition 1.6. Let $X = (X(t), t \geq 0)$ be a stochastic process adapted to filtration $(\mathcal{F}_t, t \geq 0)$ and satisfying the integrability condition $\mathbb{E}|X(t)| < \infty$. We call X a *martingale* if for all $0 \leq s < t$

$$\mathbb{E}(X(t)|\mathcal{F}_s) = X(s) \quad \text{a.s.}$$

The conditional expectation $\mathbb{E}(X(t)|\mathcal{F}_s)$ is the mean value of $X(t)$ given all information available at time s. We always have $\mathbb{E}(X(s)|\mathcal{F}_s) = X(s)$ (which is a generalisation of the identity $\mathbb{E}(\xi|\xi) = \xi$ for a random variable ξ).

Theorem 1.7. *Brownian motion is a martingale.*

Proof. For $0 \le s < t$, by the independence of increments and adaptedness

$$\begin{aligned}\mathbb{E}(B(t)|\mathcal{F}_s) = \mathbb{E}((B(t) - B(s)) + B(s)|\mathcal{F}_s) = \\ \mathbb{E}(B(t) - B(s)|\mathcal{F}_s) + \mathbb{E}(B(s)|\mathcal{F}_s) = \\ \mathbb{E}(B(t) - B(s)) + B(s) = B(s).\end{aligned}$$

□

1.4 BM as a Markov process

For Markov process the future behaviour depends on the information accumulated by time s only through the state at time s. We encountered this property when discussing the transition probabilities of BM.

Definition 1.8. Let $(X(t), t \ge 0)$ be a stochastic process adapted to filtration $(\mathcal{F}_t, t \ge 0)$. We say that $(X(t), t \ge 0)$ is a *Markov process* if for all $0 \le s < t$

$$\mathbb{P}(X(t) \in [a, b]|\mathcal{F}_s) = \mathbb{P}(X(t) \in [a, b]|X(s)) \quad \text{a.s., for } [a, b] \subset \mathbb{R}. \tag{4}$$

If $(\mathcal{F}_t, t \ge 0) = (\mathcal{F}_t^X, t \ge 0)$ is the natural σ-algebra of X, the definition is equivalent to any of the following conditions

- in terms of the conditional distribution function of $X(t)$: for $t_1 < \cdots < t_k < s < t$

$$F_{X(t)|X(s)=x,X(t_1)=x_1,\ldots,X(t_k)=x_k}(y) = F_{X(t)|X(s)=x,X(t_1)=x_1,\ldots,X(t_k)=x_k}(y),$$

- in terms of the conditional density function of $X(t)$: for $t_1 < \cdots < t_k < s < t$

$$f_{X(t)|X(s)=x,X(t_1)=x_1,\ldots,X(t_k)=x_k}(y) = f_{X(t)|X(s)=x,X(t_1)=x_1,\ldots,X(t_k)=x_k}(y).$$

It will be convenient to re-formulate (4) as the equivalent condition

$$\mathbb{E}(f(X(t))|\mathcal{F}_s) = \mathbb{E}(f(X(t))|X(s)) \quad \text{a.s.} \tag{5}$$

which must hold for all functions $f(x)$ (for which expectations make sense). Note that (4) is a special case of (5), corresponding to the indicator function $f(x) = 1(x \in [a, b])$. Here and throughout $1(\cdots) = 1$ when $\cdots$ is true, and $1(\cdots) = 0$ when $\cdots$ is false.

To work with conditional expectations we need a lemma.

Lemma 1.9. *Let $\mathcal{G}$ be a σ-algebra and let X, Y be two random variables such that X is $\mathcal{G}$-measurable, and Y is independent of $\mathcal{G}$. Let $f(x, y)$ be a function and define $g(x) := \mathbb{E}f(x, Y)$. Then*

$$\mathbb{E}(f(X, Y)|\mathcal{G}) = g(X).$$

The intuitive idea here is that the information in $\mathcal{G}$ is sufficient to evaluate X, thus X can be treated as nonrandom constant when computing $\mathbb{E}(f(X, Y)|\mathcal{G})$. On the other hand, $\mathbb{E}(f(x, Y)|\mathcal{G}) = \mathbb{E}(f(x, Y)$ because Y is independent of $\mathcal{G}$.

Example 1.10. Consider the function $f(x,y) = xy$. We have $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G}) = X\,\mathbb{E}Y$, because X is $\mathcal{G}$-measurable and acts like a constant in $\mathbb{E}(XY|\mathcal{G})$, and by independence of Y from $\mathcal{G}$. On the other hand, $g(x) = \mathbb{E}(xY) = x\mathbb{E}Y$, whence subsititung X for dummy variable x yields $g(X) = X\mathbb{E}Y$.

Proof. We shall consider only the simplest case $\mathcal{G} = \sigma(X)$, of the σ-algebra generated by X. Recall that the conditional expectation $\mathbb{E}(\cdots|X)$ is a random variable (a function of X), determined as follows: we first calculate $\mathbb{E}(\cdots|X = x)$ thus obtaining a function of x, then substitute X in this function instead of the dummy variable x. For instance, $\mathbb{E}(e^X|X) = e^X$.

Following the recipe, using independence and the definition of g

$$\mathbb{E}(f(X,Y)|X = x) = \mathbb{E}(f(x,Y)|X = x) = \mathbb{E}f(x,Y) = g(x).$$

Substituting X for x yields $\mathbb{E}(f(X,Y)|X) = g(X)$, as claimed in the case $\mathcal{G} = \sigma(X)$. □

Theorem 1.11. *The Brownian motion with any filtration $(\mathcal{F}_t, t \geq 0)$ for the BM is a Markov process.*

Proof. We wish to check (5). For $t > s$ write

$$\mathbb{E}[f(B(t))|\mathcal{F}_s] = \mathbb{E}[f(B(s) + (B(t) - B(s)))|\mathcal{F}_s],$$

which is now of the form $\mathbb{E}[f(X + Y)|\mathcal{F}_s]$ with $X := B(s)$ $\mathcal{F}_s$-measurable and $Y := B(t) - B(s)$ independent of $\mathcal{F}_s$. To apply Lemma 1.9 introduce

$$g(x) := \mathbb{E}f(x + B(t) - B(s)) = \int_{-\infty}^{\infty} f(x+u)\frac{e^{-u^2/(2t-2s)}}{\sqrt{2\pi(t-s)}}du.$$

By the lemma

$$\mathbb{E}[f(B(t))|\mathcal{F}_s] = g(B(s)).$$

Conditioning this by $B(s)$ we obtain

$$\mathbb{E}[\mathbb{E}[f(B(t))|\mathcal{F}_s]|B(s)] = \mathbb{E}[g(B(s))|B(s)].$$

The left-hand side of the last formula is

$$\mathbb{E}[\mathbb{E}[f(B(t))|\mathcal{F}_s]|B(s)] = \mathbb{E}[f(B(t))|B(s)],$$

because $\sigma(B(s)) \subset \mathcal{F}_s$ ($B(s)$ is $\mathcal{F}_s$-measurable, so the σ-algebra generated by $B(s)$ is smaller than $\mathcal{F}_s$). The right-hand side is $\mathbb{E}[g(B(s))|B(s)] = g(B(s))$. It follows that $\mathbb{E}[f(B(t))|\mathcal{F}_s] = g(B(s)) = \mathbb{E}[f(B(t))|B(s)]$, which proves the Markov property in the form (5). □

1.5 Quadratic variation

The Brownian path has no jumps, but it has a highly erratic behaviour in what concerns the slope of the function at every point. It turns that the path is nowhere differentiable. Quadratic variation quantifies the irregular fluctuations of the path. We start with reviewing a related concept from the classical analysis.

Let f be a function on $[0,T]$. The *total* or *first-order variation* $V_T(f)$ is a measure of oscillation of the function. If the function is monotone (increasing or decreasing) it is just the size of the range $V_T(f) = |f(T)-f(0)|$. If the function is piecewise monotone, changing the direction of monotonicity at points t_i, we split $[0,T]$ at the turning points and sum up the variation over the resulting subintervals of monotonicity: $V_T(f) = \sum_i |f(t_{i+1} - f(t_i)|$. In general we consider partitions $\Delta = \{t_0, t_1, \dots, t_n\}$ of $[0,T]$ by some number of points $0 = t_0 < t_1 < \cdots < t_n = T$, and define the total variation as a limit (which may or may not exist)

$$V_T(f) = \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|,$$

as the 'mesh size' $|\Delta| = \max_{0\le i<n}(t_{i+1} - t_i)$ goes to 0.

If the function is differentiable then by the mean value theorem $f(t_{i+1}) - f(t_i) = f'(t_i^*)(t_{i+1} - t_i)$ for some intermediate point $t_i^* \in [t_i, t_{i+1}]$, and $\sum_{i=0}^{n-1} |f(t_{i+1} - f(t_i)|$ is a Riemann sum for the integral of the function $|f'(t)|$, therefore

$$V_T(f) = \int_0^T |f'(t)|dt.$$

It can be shown that for the Brownian motion $V_T(B) = \infty$. Here is a rough idea. Let us split $[0,T]$ in n equal pieces at points Ti/n, $0 \le i \le n$, thus obtaining a partition with $|\Delta| = T/n$. The expected value of the increment over a generic division interval is $\mathbb{E}|B(T(i+1)/n) - B(Ti/n)| = c/\sqrt{n}$, where $c = \sqrt{2T/\pi}$. Adding up over n intervals we obtain a quantity with expected value $nc/\sqrt{n} = c\sqrt{n}$ which goes to ∞ as $n \to \infty$.

This 'back of the envelope' calculation suggests to look at the sum of *squares* of the increments, to obtain a converging quantity as $|\Delta| \to 0$.

Definition 1.12. For function f defined on $[0,T]$ the *quadratic variation* of f up to time T is

$$\langle f\rangle(T) = \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|^2,$$

where $\Delta = \{t_0, t_1, \dots, t_n\}$ with $0 = t_0 < t_1 < \cdots < t_n = T$.

Suppose f is differentiable, with $\int_0^T |f'(t)|^2 dt < \infty$. Then using the mean value theo-

rem, we obtain

$$\langle f\rangle(T) = \lim_{|\Delta|\to 0}\sum_{i=0}^{n-1}|f(t_{i+1})-f(t_i)|^2 = \lim_{|\Delta|\to 0}\sum_{i=0}^{n-1}|f(t_i^*)|^2(t_{i+1}-t_i)^2 \le$$
$$\lim_{|\Delta|\to 0}|\Delta|\sum_{i=0}^{n-1}|f(t_i^*)|^2(t_{i+1}-t_i) = \lim_{|\Delta|\to 0}|\Delta|\cdot\lim_{|\Delta|\to 0}\sum_{i=0}^{n-1}|f(t_i^*)|^2(t_{i+1}-t_i) =$$
$$0\cdot\int_0^T|f'(t)|^2dt = 0.$$

In particular, $\langle f\rangle(T) = 0$ for every continuously differentiable function. If we understand, intuitively, $df(t)$ as increment of the function f over an 'infinitesimal' interval dt, and if we write $(dt)^2 = 0$ meaning that $(dt)^2$ is a quantity of the order smaller than dt, then the mnemonic rule to keep in mind for quadratic variation of smooth functions is $(df(t))^2 = 0$.

In contrast to smooth functions, the BM has nontrivial quadratic variation

$$\langle B\rangle(T) = T. \tag{6}$$

It should be stressed, however, that a care is needed in the interpretation of the formula. A delicate point is that the sum over partition intervals

$$Q_\Delta := \sum_{i=0}^{n-1}(B(t_{i+1})-B(t_i))^2$$

is not a fixed numerical value, rather a random variable, that is a function of $\omega\in\Omega$. Thus we need to be specific regarding the sense in which of Q_Δ approaches T as the mesh size Δ goes to 0. In the following theorem the *mean-square convergence* is asserted.

Theorem 1.13. *For Brownian motion B and $Q_\Delta := \sum_{i=0}^{n-1}(B(t_{i+1})-B(t_i))^2$*

$$\lim_{|\Delta|\to 0}\mathbb{E}(Q_\Delta - T)^2 = 0, \qquad T>0.$$

Proof. Note that Q_Δ is the sum of independent random variables $(B(t_{i+1})-B(t_i))^2$, with expected value

$$\mathbb{E}(B(t_{i+1})-B(t_i))^2 = \mathrm{Var}(B(t_{i+1})-B(t_i)) = t_{i+1}-t_i.$$

Summing over i yields

$$\mathbb{E}Q_\Delta = \sum_{i=0}^{n-1}\mathbb{E}(B(t_{i+1})-B(t_i))^2 = \sum_{i=1}^{n-1} = T.$$

To compute the variance write

$$\mathrm{Var}(B(t_{i+1})-B(t_i))^2 = \mathbb{E}[(B(t_{i+1})-B(t_i))^2-(t_{i+1}-t_i)]^2 =$$
$$\mathbb{E}(B(t_{i+1})-B(t_i))^4 - 2(t_{i+1}-t_i)\mathbb{E}(B(t_{i+1})-B(t_i))^2 + (t_{i+1}-t_i)^2.$$

The fourth moment of the standard normal distribution is 3, thus $\mathbb{E}B(1)^4 = 3$. On the other hand, $B(t) \stackrel{d}{=} \sqrt{t}B(1)$, where $\stackrel{d}{=}$ means 'has the same distribution'. Thus $\mathbb{E}[B(t)]^4 = \mathbb{E}[\sqrt{t}B(1)]^4 = 3t^2$. From this

$$\mathbb{E}(B(t_{i+1}) - B(t_i))^4 = 3(t_{i+1} - t_i)^2,$$

whence

$$\mathrm{Var}(B(t_{i+1}) - B(t_i))^2 = 3(t_{i+1} - t_i)^2 - 2(t_{i+1} - t_i)^2 + (t_{i+1} - t_i)^2 = 2(t_{i+1} - t_i)^2.$$

Therefore by independence

$$\mathrm{Var}(Q_\Delta) = \sum_{i=0}^{n-1} \mathrm{Var}(B(t_{i+1}) - B(t_i))^2 =$$

$$\sum_{i=0}^{n-1} 2(t_{i+1} - t_i)^2 \leq |\Delta| \sum_{i=0}^{n-1} 2(t_{i+1} - t_i) = 2|\Delta|T.$$

Letting $\Delta \to 0$ we have $\mathrm{Var}(Q_\Delta) \to 0$, and in the view of $\mathbb{E}Q_\Delta = T$ this exactly means that

$$\lim_{|\Delta| \to 0} \mathbb{E}(Q_\Delta - T)^2 = 0$$

□

The 'back of the envelope' computations help to better understand the result. Split $[0, T]$ in n equal pieces. The squared increment over the ith subinterval is a random variable

$$(B(T(i+1)/n) - B(Ti/n))^2 = TN_i^2/n,$$

where $N_1, \ldots, N_n$ are independent $\mathcal{N}(0,1)$-distributed with $\mathbb{E}N_i = 1$. Each of N_i^2/n deviate from its mean $1/n$, but the $\sum_i N_i^2/n$ converges to $\mathbb{E}N_i^2 = 1$ by the law of large numbers. The informal rule to remember is $(dB(t))^2 = dt$, which should not be taken literally, as it only makes sense when we sum up (i.e. integrate). In the same line, $dB(t)dt = 0$, meaning a quantity of the order smaller than dt.