CHAPTER 4

Brownian motion

Brownian motion is one of the most important and interesting stochastic processes. The history of the Brownian motion began in 1827 when the botanist Robert Brown looked through a microscope at small particles (pollen grains) suspended in water. He noted that the particles were moving chaotically. The mechanism causing this chaotic motion can be explained as follows. The particle collides with water molecules. Any collisions results in a displacement of the particle in some direction. The number of collisions is large, but the impact of any collision is small. To compute the total displacement of the particle caused by all collisions we have to add a very large number of very small random variables (impacts of individual collisions), like in the central limit theorem.

A similar situation appears when we try to model a price of an asset. The price, considered as a function of time, is subject to random changes due to the influence of some random events. If we assume that any random event has a very small impact on the price and that the number of events is very large, we are in the same situation when modelling the Brownian particle. This is why the Brownian motion is one of the main building blocks for stochastic processes used in financial mathematics.

In this chapter we will define a stochastic process $\{B(t)\colon t \geq 0\}$ (called the *Brownian motion* or the *Wiener process*) which is a mathematical model for the experiment described above.

4.1. Discrete approximation to the Brownian motion

Let us now try to model the motion of a small pollen grain particle in a fluid mathematically. First of all, we will model the motion of the particle in *one* dimension (that is, on the real line), because to model the motion in three dimensions we can model the three coordinates of the particle separately. So, we want to model a particle which moves on the real line due to random impacts which can shift the particle to the left or to the right. Assume without restriction of generality that at time 0 the particle starts at position 0. Denote by N the parameter describing the number of collisions of the particle with water molecules per unit time. This parameter should be very large. Assume that any collision causes a displacement of the particle by a distance $\delta > 0$ (which should be very small) either to the left or to the right, both possibilities having the same probability 1/2. A sample path of such particle (the coordinate of the particle as a function of time) is shown on Figure 1, left. Note that in this model we ignore the inertia of the particle. That is, the impacts are assumed to change the position of the particle, but we don't try to model the speed of the particle. This approach is justified if the fluid has large viscosity.

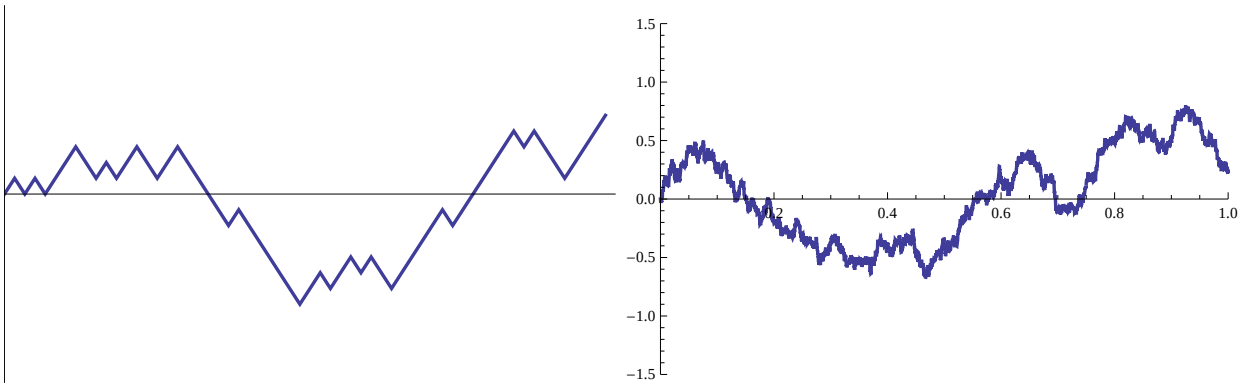

FIGURE 1. Left: A sample path of the process $B_{N,\delta}$. Right: A sample path of the Brownian motion

A more precise description of the model is as follows. Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with

$$\mathbb{P}[\xi_i = +1] = \mathbb{P}[\xi_i = -1] = \frac{1}{2}.$$

Define a stochastic process $\{B_{N,\delta}(t) \colon t \geq 0\}$ describing the position of the particle at time t as follows. The position of the particle at time $t = \frac{k}{N}$, where $k \in \mathbb{N}_0$, is given by the sum of the first k impacts:

$$B_{N,\delta}\left(\frac{k}{N}\right) = \delta \cdot (\xi_1 + \ldots + \xi_k).$$

For $t \in (\frac{k}{N}, \frac{k+1}{N})$ we can define $B_{N,\delta}(t)$ by linear interpolation, as in Figure 1.

It is clear from the definition that the process $\{B_{N,\delta}(t) \colon t \geq 0\}$ has the following two properties:

(1) $B_{n,\delta}(0) = 0$.
(2) For every integer numbers $0 \leq k_1 \leq k_2 \leq \ldots \leq k_n$, the increments

$$B_{N,\delta}\left(\frac{k_1}{N}\right), B_{N,\delta}\left(\frac{k_2}{N}\right) - B_{N,\delta}\left(\frac{k_1}{N}\right), \ldots, B_{N,\delta}\left(\frac{k_n}{N}\right) - B_{N,\delta}\left(\frac{k_{n-1}}{N}\right)$$

are independent.

Let us now determine the approximate distribution of these increments. First of all, let us look at the position of the particle at time 1:

$$B_{N,\delta}(1) = \delta \cdot (\xi_1 + \ldots + \xi_N).$$

This position is a random variable and its expectation and variance are given by

$$\mathbb{E}B_{N,\delta}(1) = 0, \quad \operatorname{Var} B_{N,\delta}(1) = \delta^2 N.$$

Now, we want to see what happens in the scaling limit as $N \to \infty$ (meaning that the number of collisions of particle with water molecules is very large) and, at the same time, $\delta \to 0$ (meaning that the displacement caused by any collision is very small); see Figure 1, right. It is natural to require that $\operatorname{Var} B_{N,\delta}(1)$ should stay constant (independent of N and δ) because

otherwise we will not obtain any meaningful limit. We will choose this constant to be equal to 1 which leads to the requirement

$$\delta = \frac{1}{\sqrt{N}}.$$

If this relation holds, then by the central limit theorem we obtain that

$$B_{N,\delta}(1) = \frac{\xi_1 + \ldots + \xi_N}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathrm{N}(0,1).$$

Similarly, for more general increments one obtains the following property:

$$B_{N,\delta}(t+h) - B_{N,\delta}(t) \xrightarrow[N\to\infty]{d} \mathrm{N}(0,h).$$

So, in the limit, the increments of our process should have the normal distribution.

4.2. Definition of the Brownian motion

The considerations of the preceding section make the following definition natural.

DEFINITION 4.2.1. A stochastic process $B = \{B(t)\colon t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *Brownian motion* or *Wiener process* if

(1) $B(0) = 0$.
(2) B has *independent increments*, that is for all $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$ the random variables
$$B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$$
are independent.
(3) B has *normal increments*, that is for all $t \geq 0$ and $h > 0$,
$$B(t+h) - B(t) \sim \mathrm{N}(0,h).$$
(4) B has *continuous sample paths*, that is for all $\omega \in \Omega$, the function $t \mapsto B(t;\omega)$ is continuous in t.

First of all, one has to ask whether a process satisfying these four requirements exists. This question is non-trivial and will be positively answered in Section 4.5 below. Here we sketch an idea of a possible approach to proving existence. The first three properties in the definition of the Brownian motion deal with the finite dimensional distributions of the process B only. It can be shown using Kolmogorov's existence theorem that a process with finite-dimensional distributions satisfying coonditions 1, 2, 3 exists. To be able to apply Kolmogorov's existence theorem one has to verify that the family of finite-dimensional distributions given by conditions 1, 2, 3 is consistent, that is that these conditions do not contradict each other. Essentially, this verification boils down to the following argument. If we know that for some $0 \leq t_1 \leq t_2 \leq t_3$ the increments

$$B(t_2) - B(t_1) \sim \mathrm{N}(0, t_2 - t_1) \text{ and } B(t_3) - B(t_2) \sim \mathrm{N}(0, t_3 - t_2)$$

are independent, then by the convolution property of the normal distribution, we must have

$$B(t_3) - B(t_1) = (B(t_3) - B(t_2)) + (B(t_2) - B(t_1)) \sim \mathrm{N}(0, (t_3 - t_2) + (t_2 - t_1)) = \mathrm{N}(0, t_3 - t_1).$$

Since this is in agreement with condition 3, there seems to be no contradiction between the conditions 1, 2, 3. Thus, we can apply Kolmogorov's existence theorem to construct a process satisfying conditions 1, 2, 3. However, Kolmogorov's theorem does not guarantee that the

resulting process satisfies condition 4, so that an additional modification of the construction is needed to make condition 4 satisfied. This is why we choose a different approach to prove the existence of a process satisfying conditions 1, 2, 3, 4; see Section 4.5.

The following example shows that it is not possible to drop condition 4 from the definition of the Brownian motion.

EXAMPLE 4.2.2. Assume that we have a process $\{B(t)\colon t \geq 0\}$ satisfying conditions 1, 2, 3, 4. We will show how, by modifying B, we can construct a process $\tilde{B}$ which satisfies properties 1, 2, 3, but violates property 4. this proves that property 4 is not a corollary of properties 1, 2, 3. Take a random variable $U \sim \mathrm{U}[0,1]$ independent of the process B. Define a new process $\{\tilde{B}(t)\colon t \geq 0\}$ by

$$\tilde{B}(t) = \begin{cases} B(t), & \text{if } t \neq U, \\ 0, & \text{if } t = U. \end{cases}$$

This process has the same finite-dimensional distributions as B. Indeed, the vectors

$$(B(t_1), \ldots, B(t_n)) \text{ and } (\tilde{B}(t_1), \ldots, \tilde{B}(t_n))$$

are equal unless $U \in \{t_1, \ldots, t_n\}$, but this event has probability 0. So, both random vectors are a.s. equal and hence, have the same distribution. This implies that the process $\{\tilde{B}(t)\colon t \geq 0\}$ also satisfies conditions 1, 2, 3. However, it does not satisfy condition 4 because the probability that its sample path is continuous is 0. Namely, we have

$$\lim_{t \to U, t \neq U} \tilde{B}(t) = \lim_{t \to U, t \neq U} B(t) = B(U).$$

This limit is a.s. different from $\tilde{B}(U) = 0$ because

$$\mathbb{P}[B(U) = 0] = \int_0^1 \mathbb{P}[B(u) = 0] du = \int_0^1 0 du = 0.$$

Thus, the probability that the sample path of $\tilde{B}$ has a discontinuity at U is 1.

4.3. Multivariate Gaussian distributions and Gaussian processes

It follows from the definition of the Brownian motion that its one-dimensional distributions are Gaussian, namely

$$B(t) \sim \mathrm{N}(0, t).$$

What about the multidimensional distributions of the Brownian motion? It turns out that these distributions are so-called multivariate Gaussian distributions. The aim of this section is to define the multivariate Gaussian distributions.

By definition, a random variable X has a (univariate) Gaussian distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ (notation: $X \sim \mathrm{N}(\mu, \sigma^2)$) if the density of X has the form

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad t \in \mathbb{R}.$$

It is convenient to extend this definition to the case $\mu \in \mathbb{R}$, $\sigma^2 = 0$ by declaring $X \sim \mathrm{N}(\mu, 0)$ if $X = \mu$ almost surely. The characteristic function of a Gaussian random variable $X \sim$

$\mathrm{N}(\mu, \sigma^2)$ has the form

$$\varphi_X(s) = e^{is\mu - \frac{1}{2}\sigma^2 s^2}, \quad s \in \mathbb{R}.$$

The random variable X is called standard Gaussian if it is Gaussian with $\mu = 0$ and $\sigma^2 = 1$, that is if the density of X is given by

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad t \in \mathbb{R}.$$

We will now extend the definition of the Gaussian distribution from random variables to random vectors. Let us start with the definition of a standard Gaussian random vector.

DEFINITION 4.3.1. Fix dimension $d \in \mathbb{N}$. A random vector $X = (X_1, \ldots, X_d)^T$ is called *d-dimensional standard Gaussian* if

(1) $X_1, \ldots, X_d \sim \mathrm{N}(0, 1)$ are standard Gaussian random variables and
(2) $X_1, \ldots, X_d$ are independent random variables.

By independence, the joint density of a d-dimensional standard Gaussian vector X is given by

$$f_{X_1,\ldots,X_d}(t_1, \ldots, t_d) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}(t_1^2 + \ldots + t_d^2)} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle t, t \rangle},$$

where $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$; see Figure 2.

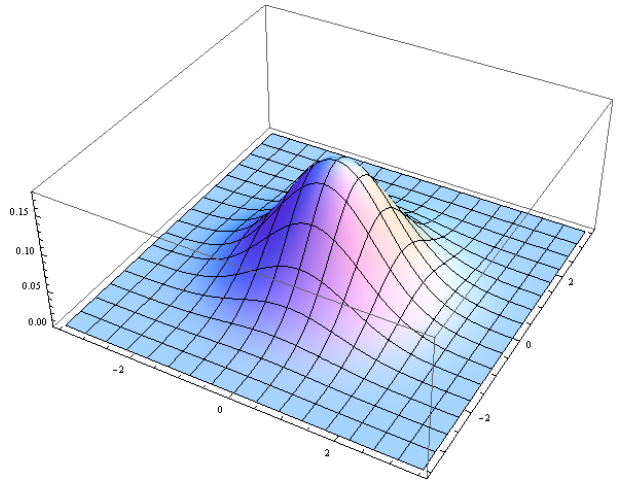

FIGURE 2. The 2-dimensional standard Gaussian density.

The expectation vector of X is equal to zero (because all components X_i have zero mean by definition). The covariance matrix of X is the $d \times d$-identity matrix (because the variance of any component X_i is 1 and different components are independent and hence uncorrelated):

$$\mathbb{E}X = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathrm{Cov}\, X = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}.$$

The next lemma states that the standard Gaussian distribution remains unchanged under rotations of the space around the origin.

LEMMA 4.3.2. *If X is d-dimensional standard Gaussian random vector and A an orthogonal $d \times d$-matrix, then the random vector AX is also standard Gaussian.*

PROOF. Recall that the orthogonality of the matrix A means that $AA^T = A^TA = \mathrm{Id}$. It follows that $\det A = \pm 1$ and in particular, A is invertible. By the transformation formula, the density of the random vector AX is

$$f_{AX}(t) = f_X(A^{-1}t)|\det(A^{-1})| = f_X(A^{-1}t) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle A^{-1}t, A^{-1}t\rangle} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle t,t\rangle} = f_X(t),$$

where we used that $\langle A^{-1}t, A^{-1}t\rangle = \langle (A^{-1})^T A^{-1}t, t\rangle = \langle (AA^T)^{-1}t, t\rangle = \langle t, t\rangle$. □

Then next lemma will be used in the construction of the Brownian motion in Section 4.5.

LEMMA 4.3.3. *Let X_1 and X_2 be independent Gaussian random variables with mean 0 and $\operatorname{Var} X_1 = \operatorname{Var} X_2 = \sigma^2$. Then, the random variables*

$$Y_1 = \frac{X_1 + X_2}{\sqrt{2}} \text{ and } Y_2 = \frac{X_1 - X_2}{\sqrt{2}}$$

are also independent and Gaussian with mean zero and variance σ^2.

PROOF. By definition, the random vector $(X_1/\sigma, X_2/\sigma)^T$ is 2-dimensional standard Gaussian. By Lemma 4.3.2, we obtain that the random vector

$$\begin{pmatrix} \frac{Y_1}{\sigma} \\ \frac{Y_2}{\sigma} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{X_1}{\sigma} \\ \frac{X_2}{\sigma} \end{pmatrix}$$

is also two-dimensional standard Gaussian, because the matrix in the above equality is orthogonal. It follows that the random vector $(Y_1/\sigma, Y_2/\sigma)^T$ is also 2-dimensional standard Gaussian. Hence, the random variables Y_1/σ and Y_2/σ are independent and standard Gaussian. □

Now we are going to define the general (non-standard) multivariate Gaussian distribution. Essentially, we declare a random vector to be multivariate Gaussian if this random vector can be represented as an affine transform of some standard Gaussian random vector.

DEFINITION 4.3.4. A random vector $Y = (Y_1, \ldots, Y_d)^T$ is called *d-dimensional Gaussian* if there is some $m \in \mathbb{N}$, some m-dimensional *standard* Gaussian vector $X = (X_1, \ldots, X_m)^T$, some $d \times m$-matrix A and some $\mu \in \mathbb{R}^d$ so that

$$Y \stackrel{d}{=} AX + \mu.$$

EXERCISE 4.3.5. Show that the expectation and the covariance matrix of Y are given by

$$\mathbb{E}Y = \mu, \quad \operatorname{Cov} Y = AA^T.$$

NOTATION 4.3.6. We usually denote the covariance matrix by $\Sigma := \operatorname{Cov} Y = AA^T$ (not by Σ^2), and write $Y \sim \mathrm{N}_d(\mu, \Sigma)$. Note that the parameter μ takes values in $\mathbb{R}^d$, whereas the covariance matrix Σ can be any symmetric, positive semidefinite matrix.

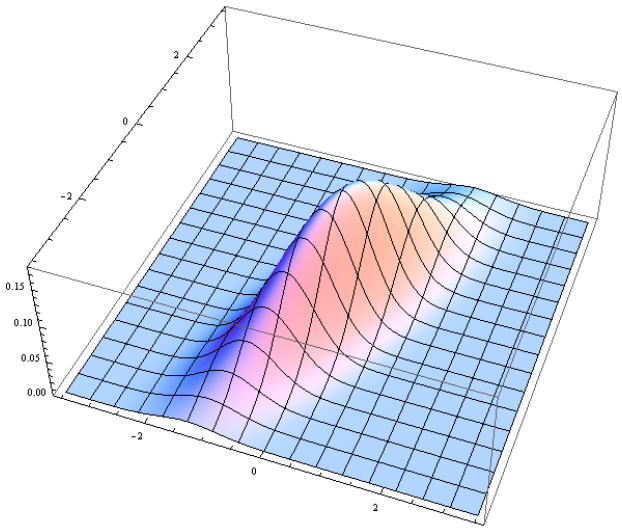

FIGURE 3. A two-dimensional (non-standard) Gaussian density

Any affine transformation of a Gaussian vector is again a Gaussian vector:

LEMMA 4.3.7. *If $Y \sim \mathrm{N}_d(\mu, \Sigma)$ is a d-dimensional Gaussian vector, A' is a $d' \times d$-matrix and $\mu' \in \mathbb{R}^{d'}$, then*

$$A'Y + \mu' \sim \mathrm{N}_{d'}(A'\mu + \mu', A'\Sigma A'^T)$$

PROOF. By definition, we can represent Y in the form $Y = AX + \mu$, where $AA^T = \Sigma$ and the vector X is m-dimensional standard Gaussian. The d'-dimensional random vector

$$A'Y + \mu' = A'(AX + \mu) + \mu' = (A'A)X + (A'\mu + \mu')$$

is also an affine transform of X and hence, multivariate Gaussian. The parameters of $A'Y + \mu'$ are given by

$$\mathbb{E}[A'Y + \mu'] = A'\mu + \mu', \quad \mathrm{Cov}(A'Y + \mu') = (A'A)(A'A)^T = A'AA^TA'^T = A'\Sigma A'^T.$$

□

REMARK 4.3.8. In particular, any component Y_i of a Gaussian random vector $(Y_1, \ldots, Y_d)^T$ is a Gaussian random variable. The converse is not true: If $Y_1, \ldots, Y_d$ are Gaussian random variables, then it's in general not true that $(Y_1, \ldots, Y_d)^T$ is a Gaussian random vector. However, if we additionally require that $Y_1, \ldots, Y_d$ should be independent, the statement becomes true.

LEMMA 4.3.9. *Let $Y_1, \ldots, Y_d$ be independent Gaussian random variables. Then, $(Y_1, \ldots, Y_d)^T$ is a Gaussian random vector.*

PROOF. Let $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$. Then, we can write $Y_i = \sigma_i X_i + \mu_i$, where X_i are standard normal and independent. So, the random vector $(Y_1, \ldots, Y_d)^T$ is an affine transformation of some standard Gaussian random vector $(X_1, \ldots, X_d)^T$ and hence, itself d-dimensional Gaussian. □

LEMMA 4.3.10. *The characteristic function of a d-dimensional Gaussian random vector* $Y \sim \mathrm{N}_d(\mu, \Sigma)$ *is given by*

$$\varphi_Y(t) := \mathbb{E}e^{i\langle t, Y\rangle} = e^{i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle}, \quad t \in \mathbb{R}^d.$$

PROOF. Fix $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. The mapping $y \mapsto \langle t, y\rangle$ is a linear map from $\mathbb{R}^d$ to $\mathbb{R}$ whose matrix is given by $(t_1, \ldots, t_d)$. By Lemma 4.3.7, the random variable $Z := \langle t, Y\rangle$ is Gaussian with expectation $\langle \mu, t\rangle$ and variance $\langle t, \Sigma t\rangle$. We have

$$\varphi_Y(t) = \mathbb{E}e^{i\langle t, Y\rangle} = \mathbb{E}e^{iZ} = \varphi_Z(1) = e^{i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle}.$$

where in the last step we used the known formula for the characteristic function of the Gaussian random variable Z. □

EXERCISE 4.3.11. Let $X_1, X_2, \ldots$ be a sequence of d-dimensional Gaussian vectors whose expectations μ_n converge to μ and covariance matrices Σ_n converge to Σ. Show that X_n converges in distribution to $\mathrm{N}_d(\mu, \Sigma)$.

What is the density of a multivariate Gaussian distribution $\mathrm{N}_d(\mu, \Sigma)$? First of all, this density does not always exist, as the following example shows.

EXAMPLE 4.3.12. Let us construct an example of a two-dimensional Gaussian random vector which has no density. Let X be a standard normal random variable. The two-dimensional vector $(X, X)^T$ is Gaussian because it can be represented as a linear transformation AX, where

$$A : x \mapsto \begin{pmatrix} x \\ x \end{pmatrix}.$$

However, the random vector $(X, X)^T$ has no density (with respect to the two-dimensional Lebesgue measure) because X takes values in the line $\{(x, x) : x \in \mathbb{R}\}$ which has Lebesgue measure 0. Note that the covariance matrix of $(X, X)^T$ is equal to

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

This matrix is degenerate, meaning that its determinant is 0.

The next lemma gives a formula for the density of the multivariate Gaussian distribution in the case when Σ a non-degenerate matrix.

LEMMA 4.3.13. *The density of a d-dimensional Gaussian random vector* $Y \sim \mathrm{N}_d(\mu, \Sigma)$, *where* Σ *is a non-degenerate matrix, is given by*

$$f_Y(t) = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle t-\mu, \Sigma^{-1}(t-\mu)\rangle}.$$

If the matrix Σ *is degenerate, then* Y *has no density with respect to the Lebesgue measure on* $\mathbb{R}^d$.

PROOF. Since the matrix Σ is positive semidefinite, we can write $\Sigma = \Sigma^{1/2} \cdot \Sigma^{1/2}$ for some symmetric matrix $\Sigma^{1/2}$. We have the representation

$$Y \stackrel{d}{=} \Sigma^{1/2} X + \mu,$$

where X is a standard Gaussian vector on $\mathbb{R}^d$. Consider the transformation

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto \Sigma^{1/2}x + \mu.$$

Then, $T(X) \stackrel{d}{=} Y$.

1. If Σ is degenerate, then the image of T is a subspace of $\mathbb{R}^d$ having dimension strictly smaller than d. It follows that the image of T has Lebesgue measure 0. So, Y takes values in a subset of $\mathbb{R}^d$ which has Lebesgue measure 0. It follows that Y has no density.

2. If we assume that $\det \Sigma \neq 0$, we have the inverse transformation

$$T^{-1}(y) = \Sigma^{-1/2}(y - \mu).$$

The density of X is

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle x,x\rangle}, \quad x \in \mathbb{R}^d.$$

Now we can compute the density of Y by using the transformation of density theorem:

$$f_Y(y) = f_X(T^{-1}(y))|\det T^{-1}| = \frac{1}{(\sqrt{2\pi})^d\sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle \Sigma^{-1/2}(y-\mu),\Sigma^{-1/2}(y-\mu)\rangle}, \quad y \in \mathbb{R}^d.$$

Using the symmetry of the matrix $\Sigma^{1/2}$, we obtain

$$f_Y(y) = \frac{1}{(\sqrt{2\pi})^d\sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle (y-\mu),\Sigma^{-1}(y-\mu)\rangle}, \quad y \in \mathbb{R}^d.$$

which is the required formula. □

For general random vectors it is known that the independence of components implies their uncorrelatedness, but the converse is, generally speaking, not true. It is an important property of the multivariate Gaussian distribution that for this distribution, the independence and the uncorrelatedness of the components are *equivalent*.

THEOREM 4.3.14. *Let $Y = (Y_1, \ldots, Y_d)^T$ be a random vector with multivariate Gaussian distribution. Then, the following properties are equivalent:*

(1) *The random variables $Y_1, \ldots, Y_d$ are independent.*
(2) $\mathrm{Cov}(Y_i, Y_j) = 0$ *for all $i \neq j$.*

PROOF. It is known that (1) implies (2) even without the multivariate Gaussian assumption. We prove that (2) implies (1). Assume that $\mathrm{Cov}(Y_i, Y_j) = 0$ for all $i \neq j$. The components Y_k are Gaussian, say $Y_k \sim \mathrm{N}(\mu_k, \sigma_k^2)$. By the uncorrelatedness, the covariance matrix of Y is a diagonal matrix, whereas the expectation vector of Y may be, in general, arbitrary:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_d^2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix}.$$

The characteristic function of Y is given by

$$\varphi_{Y_1,\ldots,Y_d}(t_1, \ldots, t_d) = e^{i\langle \mu,t\rangle - \frac{1}{2}\langle t,\Sigma t\rangle} = e^{i\sum_{k=1}^d \mu_k t_k - \frac{1}{2}\sum_{k=1}^d \sigma_k^2 t_k^2} = \prod_{k=1}^d e^{i\mu_k t_k - \frac{1}{2}\sigma_k^2 t_k^2} = \prod_{k=1}^d \varphi_{Y_k}(t_k).$$

This implies that $Y_1, \dots, Y_d$ are independent. $\square$

Recall that two random vectors $X = (X_1, \dots, X_n)^T$ and $Y = (Y_1, \dots, Y_m)^T$ defined on a common probability space are called independent if for every Borel sets $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ we have

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B].$$

EXERCISE 4.3.15. Let $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ be a Gaussian random vector. Show that the random vectors $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_m)$ are independent if and only if

$$\mathrm{Cov}(X_i, Y_j) = 0$$

for all $i = 1, \dots, n$ and $j = 1, \dots, m$.

4.4. Brownian motion as a Gaussian process

A stochastic process is called Gaussian if its finite-dimensional distributions are multivariate Gaussian. More precisely:

DEFINITION 4.4.1. A stochastic process $\{X(t) \colon t \in T\}$ is called Gaussian if for every $n \in \mathbb{N}$ and every $t_1, \dots, t_n \in T$, the random vector $(X(t_1), \dots, X(t_n))^T$ is n-dimensional Gaussian.

EXAMPLE 4.4.2. Let us show that the Brownian motion is a Gaussian process. Take some $0 \le t_1 \le t_2 \le \ldots \le t_n$. We show that the vector $(B(t_1), \dots, B(t_n))$ is Gaussian. Consider the random variables

$$\Delta_i = B(t_i) - B(t_{i-1}).$$

By the definition of the Brownian motion, these random variables are independent and each has Gaussian distribution. It follows from Lemma 4.3.9 that the random vector $(\Delta_1, \dots, \Delta_n)$ is n-dimensional Gaussian. We can represent $(B(t_1), \dots, B(t_n))$ as a linear transform of $(\Delta_1, \dots, \Delta_n)$:

$$B(t_i) = \Delta_1 + \ldots + \Delta_i.$$

It follows from Lemma 4.3.7 that the vector $(B(t_1), \dots, B(t_n))$ is also n-dimensional Gaussian.

REMARK 4.4.3. The finite dimensional distributions of a Gaussian process are uniquely determined by the expectation function $\mu(t) = \mathbb{E}X(t)$ and the covariance function

$$\Gamma(t_1, t_2) = \mathrm{Cov}(X(t_1), X(t_2)).$$

EXAMPLE 4.4.4. If B is a Brownian motion, then

$$\mathbb{E}B(t) = 0, \quad \Gamma(t_1, t_2) = \min(t_1, t_2).$$

Conversely, we have the following characterization of the Brownian motion.

THEOREM 4.4.5. *A stochastic process $\{B(t) \colon t \ge 0\}$ is a Brownian motion if and only if*

(1) *B is Gaussian;*
(2) *$\mathbb{E}B(t) = 0$ for all $t \ge 0$;*
(3) *$\mathrm{Cov}(B(t_1), B(t_2)) = \min(t_1, t_2)$ for all $t_1, t_2 \ge 0$;*
(4) *B has continuous sample paths.*

PROOF. It is an exercise to show that the above four conditions are equivalent to the conditions from the definition of the Brownian motion. □

The next theorem is called the *weak Markov property* of the Brownian motion.

THEOREM 4.4.6. *Let $\{B(t)\colon t \geq 0\}$ be a Brownian motion. Fix some $u \geq 0$. Then:*

(1) *The process $B_u(s) = B(u+s) - B(u), s \geq 0$, is also a Brownian motion.*
(2) *The processes $\{B(t)\colon 0 \leq t \leq u\}$ and $\{B_u(s)\colon s \geq 0\}$ are independent.*

PROOF. We will verify conditions of Theorem 4.4.5. The process B_u is Gaussian. Indeed, for every $s_1, \ldots, s_n$, the random vector $(B_u(s_1), \ldots, B_u(s_n))$ can be written as a linear transformation of the Gaussian random vector $(B(u+t_1), \ldots, B(u+t_n), B(u))$. Also, the process B_u has continuous sample paths because B does so by definition of the Brownian motion. In order to show that B_u is a Brownian motion, we compute the expectation and the covariance function of B_u. The expectation is given by

$$\mathbb{E}B_u(s) = \mathbb{E}(B(u+s) - B(u)) = 0.$$

The covariance function is given by

$$\begin{aligned}\operatorname{Cov}(B_u(s_1), B_u(s_2)) &= \operatorname{Cov}(B(u+s_1) - B(u), B(u+s_2) - B(u)) \\ &= \min(u+s_1, u+s_2) - u - u + u \\ &= \min(s_1, s_2).\end{aligned}$$

So, B_u is a Brownian motion.

We prove that the processes $\{B(t)\colon 0 \leq t \leq u\}$ and $\{B_u(s)\colon s \geq 0\}$ are independent. First of all, we need to define what does this mean.

DEFINITION 4.4.7. Two stochastic process $\{X(t)\colon t \in T\}$ and $\{Y(s)\colon s \in S\}$ defined on the same probability space are called independent if for all $t_1, \ldots, t_n \in T$ and $s_1, \ldots, s_m \in S$ the vector $(X(t_1), \ldots, X(t_n))$ is independent of $(Y(s_1), \ldots, Y(s_m))$.

To show that the processes $\{B(t)\colon 0 \leq t \leq u\}$ and $\{B_u(s)\colon s \geq 0\}$ are independent, it suffices to show that there is no correlation between these two processes. Take some $0 \leq t \leq u$ and $s \geq 0$. Then,

$$\operatorname{Cov}(B(t), B_u(s)) = \operatorname{Cov}(B(t), B(u+s) - B(u)) = t_i - t_i = 0.$$

This proves the independence. □

The next theorem states the *self-similarity property* of the Brownian motion.

THEOREM 4.4.8. *Let $\{B(t)\colon t \geq 0\}$ be a Brownian motion and let $a > 0$. Then, the process*

$$\left\{\frac{B(at)}{\sqrt{a}}\colon t \geq 0\right\}$$

is again a Brownian motion.

PROOF. Exercise.

4.5. Lévy's construction of the Brownian motion

THEOREM 4.5.1. *The Brownian motion exists. Concretely: It is possible to construct a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a stochastic process* $\{B(t)\colon t \geq 0\}$ *on this probability space such that*

(1) $B(0) = 0$.
(2) B *has independent increments.*
(3) $B(t+h) - B(t) \sim \mathrm{N}(0, h)$ *for all* $t, h \geq 0$.
(4) *For every* $\omega \in \Omega$ *the function* $t \mapsto B(t;\omega)$ *is continuous in* t.

PROOF. First we will show how to construct the Brownian motion for $t \in [0,1]$.

STEP 1: Construction on the set of dyadic rationals. Consider the sets

$$D_n = \left\{ \frac{k}{2^n} \colon k = 0, 1, \ldots, 2^n \right\}, \quad n \in \mathbb{N}_0.$$

The first few sets are given by

$$D_0 = \{0, 1\}, \quad D_1 = \left\{0, \frac{1}{2}, 1\right\}, \quad D_2 = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}, \quad \ldots$$

Note also that $D_0 \subset D_1 \subset \ldots$. Let D be the set of dyadic rationals in $[0,1]$:

$$D = \bigcup_{n=0}^{\infty} D_n.$$

By Kolmogorov's existence theorem, we can construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ carrying a collection $\{Z_t \colon t \in D\}$ of independent standard normal random variables indexed by D.

For every $n \in \mathbb{N}_0$ we will construct a family of random variables $\{B(d) \colon d \in D_n\}$ such that

(1) For all $r < s < t$ in D_n the random variables $B(t) - B(s) \sim \mathrm{N}(0, t-s)$ and $B(s) - B(r) \sim \mathrm{N}(0, s-r)$ are independent.
(2) The processes $\{B(d) \colon d \in D_n\}$ and $\{Z_t \colon t \in D \backslash D_n\}$ are independent.

We use induction over n.

Case $n = 0$: For $n = 0$ we define $B(0) = 0, B(1) = Z_1 \sim \mathrm{N}(0,1)$.

Passing from $n-1$ *to* n: Assume we have constructed $\{B(d) \colon d \in D_{n-1}\}$ for which Properties (1) and (2) hold. We construct $\{B(d) \colon d \in D_n\}$ as follows. For $d \in D_n \backslash D_{n-1}$ define

$$B(d) = \frac{B(d - 2^{-n}) + B(d + 2^{-n})}{2} + \frac{Z_d}{2^{\frac{n+1}{2}}}.$$

Property (2) holds because for $d \in D_n$, the random variable $B(d)$ is defined in terms of the random variables $\{Z_t \colon t \in D_n\}$ only.

We prove Property (1). Define random variables:

$$X_1 := \frac{B(d + 2^{-n}) - B(d - 2^{-n})}{2} \sim \mathrm{N}(0, 2^{-n-1}),$$
$$X_2 := \frac{Z_d}{2^{\frac{n+1}{2}}} \sim \mathrm{N}(0, 2^{-n-1}).$$

The random variables X_1 and X_2 are independent by Property (2) (which, by the induction assumption, holds for $n-1$ instead of n). By Lemma 4.3.3, the random variables

$$X_1 + X_2 = B(d) - B(d - 2^{-n}) \sim \mathrm{N}(0, 2^{-n}),$$
$$X_1 - X_2 = B(d - 2^{-n}) - B(d) \sim \mathrm{N}(0, 2^{-n})$$

are independent.

The above shows that any two "neighboring" increments of the form $B(d) - B(d - 2^{-n})$, $B(d-2^{-n})-B(d)$, where $d \in D_n \backslash D_{n-1}$, are independent. In fact, we show that all increments

$$B(d) - B(d - 2^{-n}), \quad d \in D_n \backslash \{0\},$$

are independent. This implies Property (2). The vector formed by these increments is Gaussian since it is a linear transform of the standard Gaussian vector $\{Z_t \colon t \in D_n\}$. Consider two intervals of the form

$$I_1 = [d_1 - 2^{-n}, d_1], \quad I_2 = [d_2 - 2^{-n}, d_2], \quad d_1, d_2 \in D_n \backslash \{0\}, \quad d_1 < d_2.$$

They are separated by some $d \in D_j$, where we choose j to be minimal with this property. We prove that the increments of B over these intervals are independent. We have considered the case if $j = n$ above. Therefore, let $j < n$. The intervals I_1 and I_2 are contained in $K_1 = [d - 2^{-j}, d]$ and $K_2 = [d + 2^{-j}, d]$ since otherwise, we could replace d by $d \pm 2^{-j}$ which has smaller j. By the induction assumption, the increments of B over the intervals K_1 and K_2 are independent. The increments over the intervals I_1 and I_2 are defined using the increments over K_1 and K_2 and some disjoint subsets of the family $\{Z_t \colon t \in D_n\}$. Hence, the increments over I_1 and I_2 are independent.

This completes the construction of $\{B(t) \colon t \in D\}$.

Step 2: Extending the construction to $[0, 1]$. Define a sequence $F_0, F_1, \ldots$ of random functions on the interval $[0, 1]$ as follows. Let $F_0(t) = Z_1 t$, for $t \in [0, 1]$. Further, define

$$F_n(t) = \begin{cases} 0, & t \in D_{n-1}, \\ 2^{-\frac{n+1}{2}} Z_t, & t \in D_n \backslash D_{n-1}, \end{cases}$$

and let $F_n(t)$ be defined by linear interpolation between the points from D_n.

For $d \in D_n$ we defined in Step 1

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d).$$

We prove that there is a measurable set $\Omega_1 \subset \Omega$ with $\mathbb{P}[\Omega_1] = 1$ such that for all $\omega \in \Omega_1$ there exists $N = N(\omega) \in \mathbb{N}$ such that for all $n > N$,

$$\sup_{t \in [0,1]} |F_n(t)| \le 3\sqrt{n} 2^{-n/2}. \tag{4.5.1}$$

Let us prove (4.5.1). Let $c > 1$. Then, for large enough n,

$$\mathbb{P}[|Z_d| > c\sqrt{n}] = 2\mathbb{P}[Z_d > c\sqrt{n}] \le 2e^{-c^2 n/2}. \tag{4.5.2}$$

Here, we used the asymptotics

$$\mathbb{P}[Z_d > x] \sim \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \quad x \to \infty,$$

which can be proven using the L'Hôspital rule. We have, using (4.5.2),

$$\sum_{n=0}^{\infty} \mathbb{P}[\exists d \in D_n \colon |Z_d| \geq c\sqrt{n}] \leq \sum_{n=0}^{\infty} \sum_{d \in D_n} \mathbb{P}[|Z_d| \geq c\sqrt{n}] \leq \sum_{n=0}^{\infty} (2^n + 1) \cdot 2e^{-c^2 n/2} < \infty,$$

where the last step holds if $c > \sqrt{2 \log 2}$, for example, if $c = 3$. By the Borel-Cantelli lemma, we obtain that (4.5.1) holds.

It follows from (4.5.1) that for all $\omega \in \Omega_1$ the series $\sum_{n=0}^{\infty} F_n(t;\omega)$ converges uniformly over $t \in [0,1]$. The sum of the series is denoted by $B(t;\omega)$. Since the sum of a uniformly convergent series of continuous functions is continuous, we have that for all $\omega \in \Omega_1$ the function $t \mapsto B(t;\omega)$ is continuous.

Step 3: We show that the process $\{B(t) \colon t \in [0,1]\}$ constructed in Step 2 has independent and normal increments. Take some $0 \leq t_1 \leq \ldots \leq t_n \leq 1$. Since the set D is dense in $[0,1]$ we can find $t_{1,n} \leq \ldots \leq t_{n,k} \in D$ so that $\lim_{k\to\infty} t_{i,k} = t_i$. By the continuity of B we have

$$\Delta_{i+1} := B(t_{i+1}) - B(t_i) = \lim_{k\to\infty} (B(t_{i+1,k}) - B(t_{i,k})) = \lim_{k\to\infty} \Delta_{i+1,k},$$

where $\Delta_{i+1,k} := B(t_{i+1,k}) - B(t_{i,k})$. The vector $(\Delta_{1,k}, \ldots, \Delta_{n,k})$ is Gaussian by the construction from Step 1, with mean 0. Again, by the construction of Step 1, we have

$$\operatorname{Cov}(\Delta_{i+1,k}, \Delta_{j,k}) = (t_{i+1,k} - t_{i,k}) \mathbb{1}_{i=j} \to (t_{i+1} - t_i) \mathbb{1}_{i=j}, \quad k \to \infty.$$

It follows (see Exercise 4.3.11), the random vector $(\Delta_1, \ldots, \Delta_n)$ is also Gaussian, with mean 0 and with covariance matrix

$$\operatorname{Cov}(\Delta_{i+1}, \Delta_j) = (t_{i+1} - t_i) \mathbb{1}_{i=j}.$$

In particular, the components of this vector are independent and the variance of Δ_{i+1} is $t_{i+1} - t_i$. This proves that $\{B \colon t \in [0,1]\}$ has independent increments and that $B(t+h) - B(t) \sim \mathrm{N}(0,h)$.

Step 4: We extend the construction to all $t \geq 0$. Take independent copies $B_0 = B, B_1, B_2, \ldots$ of the process $\{B(t) \colon t \in [0,1]\}$ constructed in Steps 1–3 and glue them together. Concretely, for $t \in [n, n+1]$ define

$$B(t) = B_n(t-n) + \sum_{i=0}^{n-1} B_i(1).$$

The process $\{B(t) \colon t \geq 0\}$ defined in this way is Gaussian and has continuous sample paths. It is not difficult to check that its covariance function coincides with the covariance function of the Brownian motion. So, the process $\{B(t) \colon t \geq 0\}$ is the Brownian motion. □

4.6. Non-differentiability of Brownian motions paths

Theorem 4.6.1 (Paley, Wiener, Zygmund). *Let $\{B(t) \colon t \geq 0\}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, with probability 1, the function $t \to B(t)$ is nowhere*

differentiable. Concretely: There is a measurable set $\Omega' \subset \Omega$ with $\mathbb{P}[\Omega'] = 1$ such that for all $\omega \in \Omega'$ and for all $t_0 \geq 0$ the function $t \to B(t;\omega)$ has no derivate at t_0.

REMARK 4.6.2. We will prove even more. For a function $f : \mathbb{R} \to \mathbb{R}$ define

$$D^+ f(t) = \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \text{ (upper right derivative)},$$

$$D^- f(t) = \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \text{ (lower right derivative)}.$$

If $D^+ f(t) = D^- f(t)$ is finite, then we say that f is differentiable from the right. In a similar way one can define the upper left derivative and the lower left derivative. Consider the set

$$A := \{\omega \in \Omega : \exists t_0 \in [0,1] \text{ such that } -\infty < D^- B(t_0;\omega) \leq D^+ B(t_0;\omega) < +\infty\}.$$

We would like to show that $\mathbb{P}[A] = 0$, that is for almost every sample path of the Brownian motion and for every $t_0 \geq 0$ we have $D^+ B(t_0) = +\infty$, or $D^- B(t_0) = -\infty$, or both. However, it is not immediately clear whether the set A is measurable. Therefore, we will prove a somewhat weaker statement: There is a measurable set A' with $\mathbb{P}[A'] = 0$ such that $A \subset A'$.

PROOF. We have $A \subset \cup_{M \in \mathbb{N}} A_M$, where

$$A_M = \left\{ \omega \in \Omega : \exists t_0 \in [0,1] \text{ such that } \sup_{h \in [0,1]} \left| \frac{B(t_0+h) - B(t_0)}{h} \right| \leq M \right\}.$$

Fix some $M \in \mathbb{N}$. We show that $\mathbb{P}[A_M] = 0$. Take some $n \in \mathbb{N}, n \geq 3$. Any $t_0 \in [0,1]$ must be in some interval $t_0 \in [\frac{k-1}{2^n}, \frac{k}{2^n}], k = 1, \ldots, 2^n$. If the event A_M occurs and $t_0 \in [\frac{k-1}{2^n}, \frac{k}{2^n}]$, then the following three events also occur:

(1) $F^{(1)}_{n,k} : |B(\frac{k+1}{2^n}) - B(\frac{k}{2^n})| \leq |B(\frac{k+1}{2^n}) - B(t_0)| + |B(t_0) - B(\frac{k}{2^n})| \leq \frac{3M}{2^n}$.
(2) $F^{(2)}_{n,k} : |B(\frac{k+2}{2^n}) - B(\frac{k+1}{2^n})| \leq |B(\frac{k+2}{2^n}) - B(t_0)| + |B(t_0) - B(\frac{k+1}{2^n})| \leq \frac{5M}{2^n}$.
(3) $F^{(3)}_{n,k} : |B(\frac{k+3}{2^n}) - B(\frac{k+2}{2^n})| \leq |B(\frac{k+3}{2^n}) - B(t_0)| + |B(t_0) - B(\frac{k+2}{2^n})| \leq \frac{7M}{2^n}$.

Consider the event $F_{n,k} = F^{(1)}_{n,k} \cap F^{(2)}_{n,k} \cap F^{(3)}_{n,k}$. Then, for every $n \geq 3$ we have

$$A_M \subset \bigcup_{k=1}^{2^n} F_{n,k}.$$

We well estimate the probabilities $\mathbb{P}[F^{(1)}_{n,k}]$, $\mathbb{P}[F^{(2)}_{n,k}]$, $\mathbb{P}[F^{(3)}_{n,k}]$. For example, for $\mathbb{P}[F^{(3)}_{n,k}]$ we have

$$\mathbb{P}[F^{(3)}_{n,k}] = \mathbb{P}\left[\left| B\left(\frac{k+3}{2^n}\right) - B\left(\frac{k+2}{2^n}\right)\right| \leq \frac{7M}{2^n}\right] = \mathbb{P}\left[\frac{|N|}{\sqrt{2^n}} < \frac{7M}{2^n}\right] = \mathbb{P}\left[|N| \leq \frac{7M}{\sqrt{2^n}}\right],$$

where N is a standard normal random variable. Denoting by $f_N(t)$ its density (which is smaller than $1/\sqrt{2\pi} < 1/2$), we have

$$\mathbb{P}[F^{(3)}_{n,k}] = \int_{-\frac{7M}{2^{n/2}}}^{\frac{7M}{2^{n/2}}} f_N(t)dt \leq \frac{7M}{2^{n/2}}.$$

Similarily, one shows that

$$\mathbb{P}[F^{(1)}_{n,k}] \leq \frac{7M}{2^n}, \quad \mathbb{P}[F^{(2)}_{n,k}] \leq \frac{7M}{2^{n/2}}.$$

Since the events $F_{n,k}^{(1)}$, $F_{n,k}^{(2)}$, $F_{n,k}^{(3)}$ are independent (by the independence of increments of the Brownian motion), we have

$$\mathbb{P}[F_{n,k}] = \mathbb{P}[F_{n,k}^{(1)}] \cdot \mathbb{P}[F_{n,k}^{(2)}] \cdot \mathbb{P}[F_{n,k}^{(3)}] \leq \frac{(7M)^3}{2^{3n/2}}.$$

It follows that

$$\mathbb{P}[A_M] \leq \mathbb{P}[\cup_{k=1}^{2^n} F_{n,k}] \leq 2^n \frac{(7M)^3}{2^{3n/2}} = \frac{(7M)^3}{2^{n/2}}.$$

Since this holds for every $n \geq 3$, we have $\mathbb{P}[A_M] = 0$ and hence, the set $A' := \cup_{M \in \mathbb{N}} A_M$ has probability 0. We can now take $\Omega' = \Omega \backslash A'$. $\square$