0.1 Summary

(Lesson 7 of 07/11/19) Compiled: December 29, 2019

Summary of the previous lectures. We considered a more general stochastic process, a *Markov Process*, when the future only depends on the present. We wrote a Master Equation, and taking the continuum limit we get a second order partial differential equation, with two coefficients depending on the first two moments of the transition rate: f and D. We would want them to represent the *force* and *diffusion rate*, but we can't find their physical meaning. So we consider the Langenvin equation, reaching the desired physical meaning.
There, the increment depends on a *deterministic term* f and a *noise term*:

$$\mathrm{d}x(t) = f(x(t),t)\,\mathrm{d}t + \sqrt{2D(x(t),t)}\,\mathrm{d}B(t) \qquad f = \frac{F_{\text{ext}}}{\gamma}$$

If we discretize this equation, passing to finite differences, we get:

$$\Delta x(t) = f(x(t),t)\Delta t + \sqrt{2D(x(t),t)}\Delta B(t) \qquad \Delta B(t) \sim \frac{1}{\sqrt{2\pi\Delta t}}\exp\left(-\frac{\Delta B^2}{2\Delta t}\right)$$

This is needed because $\mathrm{d}x(t)\,/\,\mathrm{d}t$ is ill-defined (as we saw in the previous lecture). Note that $\Delta x(t) = x(t+\Delta t) - x(t)$.
We want to show that this kind equation leads to the same Fokker-Planck equation that we saw previously, and that was derived from the Master Equation. Then we would like to examine how much the stochastic amplitude (coefficient of $\mathrm{d}B(t)$) is related to *temperature*. In fact, we know already that f depends on F_{ext}, with $\boldsymbol{F}_{\text{ext}} = -\boldsymbol{\nabla} V$. We would like that, at constant temperature, the pdf of the stationary state will tend to the *Maxwell-Boltzmann distribution*:

$$\mathbb{P}(x,t) \xrightarrow[t\to\infty]{} \frac{1}{z}\exp\left(-\frac{V(x)}{k_B T}\right)$$

0.2 Stochastic integrals

We arrived at the Langevin equation:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,t) + \sqrt{2D(x,t)}\xi(t) \tag{1}$$

where $\xi(t)$ is a "rapidly varying, highly irregular function", i.e. such that for $t \neq t'$, $\xi(t)$ and $\xi(t')$ are statistically independent. As $\langle \xi(t)\rangle = 0$, this means that:

$$\langle \xi(t)\xi(t')\rangle = \delta(t-t')$$

Equation (1) does not make much sense, as $\dot{x}(t)$ does not exist anywhere. Even changing variables to $\mathrm{d}B$ (i.e. "multiplying" both sides by $\mathrm{d}t$) and integrating, we are left with the following equation:

$$x(t) = x(0) + \int_0^t f(x(\tau),\tau)\,\mathrm{d}\tau + \int_0^t \sqrt{2D(x(\tau),\tau)}\,\mathrm{d}B(\tau)$$

It is not clear how the last integral is defined, as it involves a *stochastic term* $\mathrm{d}B$.

So, before tackling the full problem, we take a step back and study the theory behind **stochastic calculus**. Let's introduce a *generic* integral of that kind:

$$S_t = \int_0^t G(\tau)\,\mathrm{d}B(\tau)$$

Intuitively, we could see this as an *infinite sum*, where each term $G(\tau)$ is weighted by the outcome of a random variable $B(\tau)$.
So, to compute it, an idea is to first introduce a *time discretization* $\{t_j\}_{j=0,\dots,n}$, with $t_n = t$, leading to:

$$S_n = \sum_{i=0}^{n} G(\tau_i)[B(t_i) - B(t_{i-1})] \qquad t_{i-1} \le \tau_i \le t_i \tag{2}$$

and then take the continuum limit for $n \to \infty$. This, however, proves to be more difficult than expected, for the following reasons:

- First of all, the increments $B(t_i) - B(t_{i-1})$ are chosen *at random*. This means that S_n is a **random variable**. In fact, we could see S_t as the sum of points from $G(\tau)$, each *weighted* with a *randomly chosen weight*.
  So it is necessary to define what it means to take the limit of a sequence of random variables S_n. As we will see, there is no unique definition.

- It is not clear how to choose the *sampling instants* τ_i for $G(\tau)$ in the discretization (2). We could hope that in the limit of $n \to \infty$, any choice would lead to the same final result. This would be indeed true if $B(\tau)$ were a differentiable function - except it is only *continuous* and *nowhere differentiable*. So we need to pay attention to the *specific* (and arbitrary) rule to be used in computing the discretization.

0.2.1 Limits of sequences of random variables

Some basic definitions. Recall that a probability space is defined by a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set of outcomes (*sample space*), $\mathcal{F}$ is a σ-algebra on Ω, containing all possible *events*, that is *sets of outcomes*, and $\mathbb{P}\colon \mathcal{F} \to [0,1]$ is the *probability measure*. Then, a **random variable** is a *measurable function* $X\colon \Omega \to S$, with S denoting a *state space*.
For example, let Ω be the set of all possible results of rolling two dice, i.e. the set of ordered pairs (x_1, x_2) with $x_1, x_2 \in \{1,2,3,4,5,6\}$. Then $\mathcal{F}$ is the set of *all possible subsets* of Ω (including both Ω and $\varnothing$) and $\mathbb{P}\colon \mathcal{F} \ni f \mapsto \mathbb{P}(f)$ is given by:

$$\mathbb{P}(f) = \frac{|f|}{36}$$

where $|f|$ is the cardinality of the set f.
A random variable can be, for example, the *sum* of the two dice:

$$X(\omega) = x_1 + x_2 \quad \forall \omega = (x_1, x_2) \in \Omega$$

Then, we can compute the probability of X assuming a certain value by measuring with $\mathbb{P}$ the preimage set of X:

$$\mathbb{P}(X = 2) = \mathbb{P}(\omega \in \Omega | X(\omega) = 2) = \mathbb{P}(\{1, 1\}) = \frac{1}{36}$$

For discrete one-dimensional variables such as these all of this formalism does not lead to much gain, as there is an immediate and natural choice for $(\Omega, \mathcal{F}, \mathbb{P})$, which is usually denoted by the saying "random". However, in more complex cases it becomes imperative to precisely define Ω, $\mathcal{F}$ and $\mathbb{P}$, so to avoid ambiguous results (see Bertrand's paradox).

Consider a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables in a certain probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that X is another random variable, and we would like to give meaning to the concept of X_n "tending to" X:

$$X_n \xrightarrow[n\to\infty]{} X$$

There are several possibilities, here stated from the weakest to the strongest:

1. **Convergence in distribution**. In this case, we simply require that the distribution of S_n approaches that of S as $n \to \infty$. Let F_n and F be the cumulative distributions of S_n and S, respectively. Then:

   $$X_n \xrightarrow[n\to\infty]{D} X \Leftrightarrow \lim_{n\to\infty} F_n(x) = F(x) \qquad \forall x \in \mathbb{R} | F \text{ is continuous at } x$$

   (The cumulative distribution, or cdf, is defined as $F_X(x) = \mathbb{P}(X \leq x)$).

   Note that, as we are merely comparing functions, there is no need for X_n or X to be defined on the *same probability space*. Also, here the focus is on *integral properties* of the random variables, so there is no guarantee that sampling X_n and X will lead to *close* results, even for a large n. For example, consider X_n to be a sequence of standard gaussians, which obviously converges to a standard gaussian (X) in the distribution sense. If we sample a number from X_{100} and one from X, they could be arbitrarily far away from each other with a non-zero probability, that remains the same for all n. If we want to *exclude* that possibility we need a *stronger requirement*, which leads to the next definition.

2. **Convergence in probability** (*Stochastic limit*). If the probability of values of X_n being *far* from values of X vanishes as $n \to \infty$, then X_n converges *in probability* to X:

   $$X_n \xrightarrow[n\to\infty]{P} X \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

   Expanding the definition, this means that:

   $$\forall \epsilon > 0, \forall \delta > 0, \exists N(\epsilon, \delta) \text{ s.t. } \forall n \geq N, \mathbb{P}(|X_n - X| > \epsilon) < \delta$$

In other words, the probability of "a significant discordance" between values sampled from X_n and X vanishes as $n \to \infty$. Intuitively, X_n and X are *strongly related*, i.e. they not only distribute similarly, but also come from *similar processes*. For example, let X be the *true length* of a stick chosen *at random* from a population of sticks, and X_n be a measurement of that length made with an instrument that is more and more precise as $n \to \infty$. Then, for large n, it is clear that X_n will have a value that is really close to that of X. In this case, we say that X_n converges *in probability* to X, as $n \to \infty$.

3. **Almost sure convergence**. An even stronger limit requires that:

$$X_n \xrightarrow[n\to\infty]{\text{a.s.}} X \Leftrightarrow \mathbb{P}\left(\liminf_{n\to\infty}\{\omega \in \Omega \colon |X_n(\omega) - X(\omega)| < \epsilon\}\right) = 1 \qquad \forall \epsilon > 0$$

Here, the $\liminf$ of a sequence of sets A_n is defined as:

$$\liminf_{n\to\infty} A_n = \bigcup_{N=1}^{\infty} \bigcap_{n\geq N} A_n$$

A member of $\liminf A_n$ is a member of *all* sets A_n, *except* a *finite* number of them (i.e. it's *definitively* a member of the A_n, as it is $\in A_n$ for all $n \geq \bar{n}$). So the term inside the parentheses is the set of all outcomes $\omega \in \Omega$ for which $X_n(\omega)$ is *definitively* close to $X(\omega)$, i.e. it *covers* all events resulting in a sequence of X_n that *converges* to X.

If we take X_n and X to be *real-valued* random variables, then the definition is simpler:

$$X_n \xrightarrow[n\to\infty]{\text{a.s.}} X \Leftrightarrow \mathbb{P}\left(\omega \in \Omega \colon \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

Or, in other words:

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \qquad \forall \omega \in \Omega \setminus A$$

where $A \subset \Omega$ has 0 measure.

Almost sure convergence vs probability convergence. The difference between the two definitions is subtle, and can be somewhat seen from the following example, taken from `http://bit.ly/2u2E9Rk` and `http://bit.ly/2Zy66v0`. Consider a sequence $\{X_n\}$ of *independent* random variables with only two possible values, 0 and 1, such that:

$$\mathbb{P}(X_n = 1) = \frac{1}{n} \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$$

For $\epsilon > 0$:

$$\mathbb{P}(|X_n| \geq \epsilon) = \begin{cases} \frac{1}{n} & 0 < \epsilon \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

As $n \to \infty$, $\mathbb{P}(|X_n| \geq \epsilon) \to 0$, and so $X_n \xrightarrow[n\to\infty]{P} 0$.
However, X_n **does not** converge almost surely to 0. Consider a *realization* of the sequence X_n, i.e. the measured outcomes of all X_n during "one run" of the experiment. This will be a binary sequence, like $000101001\ldots$. Now, consider an *ensemble* of such sequences. What is the average number of ones in them?
We can estimate it by summing the probability to have a 1 in the first place, in the second, and so on:

$$\sum_{n=1}^{\infty} \frac{1}{n} = +\infty$$

This in fact implies, by the second Borel Cantelli theorem[a], that the probability of getting $X_n = 1$ *infinitely often* (i.o.) is 1, and so X_n *cannot* converge *almost surely* to 0.

[a] ∧See a proof at `http://bit.ly/2tcfZU4` The main idea is that, given a set of *independent* events ($X_n = 1$), the sum of their probabilities diverges, then *surely* an infinite number of them *do indeed occur*. Formally: if $\sum_{n=1}^{+\infty} \mathbb{P}(X_n = 1) = \infty$, then $\mathbb{P}(\limsup_{n\to\infty}\{X_n = 1\}) = \mathbb{P}(\cap_{N=1}^{\infty} \cup_{n\geq N} \{X_n = 1\}) = \mathbb{P}(\{X_n = 1\} \text{ i.o.}) = 1$

It can be proven that *almost sure convergence* implies *convergence in probability*, which implies *convergence in distribution*. However, for our purposes we are interested in *another kind of convergence*:

- ***L^q* convergence**:

$$X_n \xrightarrow[n\to\infty]{L^q} X \Leftrightarrow \lim_{n\to\infty} \langle |X_n - X|^q \rangle = 0 \qquad q \in \mathbb{N}$$

  Note that this implies *convergence in probability*. In fact:

$$\mathbb{P}(|X - X_n| > \epsilon) = \langle \mathbb{I}_{|X-X_n|>\epsilon} \rangle \leq \langle \underbrace{\mathbb{I}_{|X-X_n|>\epsilon}}_{0\leq\odot\leq 1} \underbrace{\left| \frac{X - X_n}{\epsilon} \right|^q}_{\geq 1} \rangle \tag{3}$$

  where is a *characteristic function*, i.e. the random variable that is 1 when $|X - X_n| > \epsilon$ and 0 otherwise - so that the second term is always ≥ 1 when it is not killed by the first one. Then, by substituting $\mathbb{I}$ with its *maximum* 1 we get a *greater* term:

$$(3) \leq \langle |X - X_n|^q \rangle \frac{1}{\epsilon^q} \xrightarrow[n\to\infty]{} 0 \qquad \forall \epsilon > 0$$

  where we used the linearity of the average to extract the constant ϵ^q, and then the L^q convergence (assumed by hypothesis).

  Also, L^q convergence implies the convergence (in the usual sense) of the q-th moment:

$$X_n \xrightarrow[n\to\infty]{L^q} X \Rightarrow \lim_{n\to\infty} \langle |X_n|^q \rangle = \langle |X|^q \rangle \tag{4}$$

If we choose $q = 2$, we obtain **mean square convergence**:

$$X_n \xrightarrow[n\to\infty]{\text{m.s.}} X \Leftrightarrow \lim_{n\to\infty} \langle |X_n - X|^2 \rangle = 0$$

In this case it is easy to prove (4) by using the Cauchy-Schwarz inequality:

$$(\mathbb{E}(XY))^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

If we let $X = X_n - X$ and $Y = 1$, and assume that X_n converges to X in mean square, we obtain:

$$0 \leq (\mathbb{E}(X_n - X))^2 \leq \mathbb{E}((X_n - X)^2)\mathbb{E}(1) \xrightarrow[n\to\infty]{} 0$$

And so:

$$\mathbb{E}(X_n - X) = \mathbb{E}(X_n) - \mathbb{E}(X) \xrightarrow[n\to\infty]{} 0 \Rightarrow \lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X) \qquad \square$$

Hölder inequality. Cauchy inequality is, in this case, a special case of the more general Hölder inequality. Consider a measure space (S, Σ, μ) (where S is the space, Σ a σ-algebra and μ a measure), and two measurable functions $f, g\colon S \to \mathbb{R}$:

$$\|fg\|_1 \leq \|f\|_p \|g\|_p \qquad \|\cdot\|_p = \left(\int_S |\cdot|^p \,\mathrm{d}\mu\right)^{1/p}$$

To compute a stochastic integral, we will proceed like the following:

- Discretize the integral as a finite (Riemann) sum, obtaining a sequence of *finer* and *finer* random variables $\{S_n\}_{n\in\mathbb{N}}$
- Use a *mean square* limit to compute the limit S of the sequence $\{S_n\}$

0.2.2 Prescriptions

All that's left is to choose a *rule* for the mid-points in the terms of the discretized sum. As we will see in the following example, there are several different possibilities, each leading to *different results.*

Example 1 (A simple stochastic integral)**:**

Suppose $G(\tau) = B(\tau)$, and consider the following integral:

$$S = \int_0^t B(\tau)\,\mathrm{d}B(\tau)$$

If $B(\tau)$ where differentiable, then we could simply change variables and solve:

$$S = \int_0^t B(\tau) \frac{\mathrm{d}B(\tau)}{\mathrm{d}\tau}\,\mathrm{d}\tau = \frac{1}{2} B^2(\tau)\Big|_0^t = \frac{B^2(t) - B^2(0)}{2} \quad \text{if } \exists \frac{\mathrm{d}B}{\mathrm{d}\tau}$$

However, here $B(\tau)$ is a *rapidly varying irregular function*, which is nowhere differentiable.
So, following our plan, we first discretize:

$$S_n = \sum_{i=1}^{n} B(\tau_i)[B(t_i) - B(t_{i-1})] \qquad t_0 \equiv 0;\ t_n \equiv t;\ t_{i-1} \leq \tau_i \leq t_i \tag{5}$$

We now need a rule for choosing the τ_i. The simplest possibility is to fix them in the "same relative position" in every interval $[t_{i-1}, t_i]$, that is:

$$\tau_i = \lambda t_i + (1-\lambda) t_{i-1} \qquad \lambda \in [0,1] \tag{6}$$

Depending on the value of λ, the limit S will be different. We can quickly check this *before* computing S, by focusing on the *expected values*. In fact, we know that if $S_n \xrightarrow[n\to\infty]{\text{m.s.}} S$, then $\langle S_n \rangle \xrightarrow[n\to\infty]{} S$ in the usual sense. So, we compute the average of S_n:

$$\langle S_n \rangle = \sum_{i=1}^{n} \langle B(\tau_i)(B(t_i) - B(t_{i-1})) \rangle = \sum_{i=1}^{n} (\langle B(\tau_i) B(t_i) \rangle - \langle B(\tau_i) B(t_{i-1}) \rangle)$$

We already computed the correlator function for the Brownian noise $B(t)$:

$$\langle B(t) B(t') \rangle = \min(t, t') \tag{7}$$

And so, as $t_{i-1} \leq \tau_i \leq t_i$, we get:

$$\langle S_n \rangle = \sum_{i=1}^{n} (\tau_i - t_{i-1})$$

Substituting the choice for τ (6):

$$\langle S_n \rangle = \lambda \sum_{i=1}^{n} (t_i - t_{i-1}) = \lambda t_n = \lambda t$$

Which does not depend on n, making the limit trivial:

$$\langle S \rangle = \lim_{n\to\infty} \langle S_n \rangle = \lambda t$$

This dependence on the **prescription** of τ_i is an important difference from ordinary calculus, meaning that many common results cannot be directly translated to stochastic calculus.
In practice, there are many possibilities for λ. The two most common are:

$$\lambda = \begin{cases} 0 & \text{Ito's prescription} \\ \frac{1}{2} & \text{Stratonovich's prescription (also called middle-point prescription)} \end{cases}$$

Leading to, as we will see:

$$S_n \xrightarrow[n\to\infty]{\text{m.s.}} S = \begin{cases} \frac{B^2(t) - B^2(0)}{2} - \frac{t}{2} & \lambda = 0 \\ \frac{B^2(t) - B^2(0)}{2} & \lambda = 1/2 \end{cases}$$

The Stratonovich prescription gives exactly the same result as ordinary calculus. However, note that it involves a dependence *on the future*, i.e. the next step of a path depends on the point that is a *half-step* later. This has no a real physical meaning (in a certain sense, it "violates causality"). That's why many physicists prefer the Ito's prescription.
Let's explicitly compute both results.

Ito's prescription. We want to prove the following result:

$$\sum_{i=1}^{n} B(t_{i-1})(B(t_i) - B(t_{i-1})) \xrightarrow[n\to\infty]{\text{m.s.}} \frac{B^2(t) - B^2(0)}{2} - \frac{t}{2} \tag{8}$$

Denoting:

$$B(t_i) = B_i; \qquad \Delta B_i = B_i - B_{i-1}$$

we can rewrite (5) as:

$$S_n = \sum_{i=1}^{n} B_{i-1}\Delta B_i$$

First of all, we *split* that product in a sum of terms, with the double-product trick:

$$ab = \frac{1}{2}[(a+b)^2 - a^2 - b^2]$$

So that:

$$\begin{aligned} S_n &= \sum_{i=1}^{n} B_{i-1}\Delta B_i = \frac{1}{2}\sum_{i=1}^{n}\Big[\underbrace{(B_{i-1}+\Delta B_i)^2}_{B_i^2} - B_{i-1}^2 - (\Delta B_i)^2\Big] = \\ &= \frac{1}{2}\sum_{i=1}^{n}\Big[B_i^2 - B_{i-1}^2 - (\Delta B_i)^2\Big] = \frac{1}{2}(B_n^2 - B_0^2) - \frac{1}{2}\sum_{i=1}^{n}(\Delta B_i)^2 = \\ &= \frac{1}{2}(B^2(t) - B^2(0)) - \frac{1}{2}\sum_{i=1}^{n}(\Delta B_i)^2 \end{aligned}$$

Now (**??**) becomes:

$$\frac{B^2(t) - B^2(0)}{2} - \frac{1}{2}\sum_{i=1}^{n}(\Delta B_i)^2 \xrightarrow[n\to\infty]{\text{m.s.}} \frac{B^2(t) - B(0)}{2} - \frac{t}{2} \qquad t_n = t;\ t_0 = 0$$

Applying the definition of *mean square limit*, this is equivalent to showing that:

$$\langle\Big|\cancel{\frac{B^2(t) - B^2(0)}{2}} - \frac{1}{2}\sum_{i=1}^{n}(\Delta B_i)^2 - \Big[\cancel{\frac{B^2(t) - B^2(0)}{2}} - \frac{t}{2}\Big]\Big|^2\rangle \xrightarrow[n\to\infty]{?} 0 \tag{9}$$

Expanding:

$$\begin{aligned} \frac{1}{4}\langle\Big[-\sum_{i=1}^{n}(\Delta B_i)^2 + t\Big]^2\rangle &= \frac{1}{4}\langle\Big[t - \sum_{i=1}^{n}(\Delta B_i)^2\Big]^2\rangle \underset{(a)}{=} \frac{1}{4}\langle\Big[\sum_{i=1}^{n}(\Delta t_i - \Delta B_i^2)\Big]^2\rangle = \\ &\underset{(b)}{=} \frac{1}{4}\sum_{i,j=1}^{n}\langle[\Delta t_i - (\Delta B_i)^2][\Delta t_j - (\Delta B_j)^2]\rangle \end{aligned} \tag{10}$$

where in (a) we used $t = \sum_{i=1}^n \Delta t_i$, and in (b) $(\sum_i a_i)^2 = \sum_{ij} a_i a_j$. We can rewrite the sum highlighting the case where $i = j$:

$$(10) = \frac{1}{4}\left[\sum_{i=1}^n \langle [\Delta t_i - (\Delta B_i)^2]^2 \rangle + \sum_{i \neq j}^n \langle [\Delta t_i - (\Delta B_i)^2][\Delta t_j - (\Delta B_j)^2] \rangle \right] \tag{11}$$

Noting that the ΔB_i come from *independent gaussians*, we have that the expected values integrals factorize:

$$\langle A \rangle = \int \mathrm{d}\Delta B_i \dots \mathrm{d}\Delta B_n \, A \prod_{i=1}^n \frac{1}{\sqrt{2\pi \Delta t_i}} \exp\left(-\frac{(\Delta B_i)^2}{2\Delta t_i}\right)$$

In other words, this means that the average of the product is just the product of the averages:

$$\begin{aligned}\langle (\Delta t_i - (\Delta B_i)^2)(\Delta t_j - (\Delta B_j)^2) \rangle &= \langle (\Delta t_i - (\Delta B_i)^2) \rangle \langle (\Delta t_j - (\Delta B_j)^2) \rangle = \\ &= [\Delta t_i - \langle (\Delta B_i)^2 \rangle][\Delta t_j - \langle (\Delta B_j)^2 \rangle]\end{aligned}$$

We already computed the second moment of that gaussian:

$$\langle (\Delta B_i)^2 \rangle = \int \frac{\mathrm{d}\Delta B_i}{\sqrt{2\pi \Delta t_i}} \Delta B_i^2 \exp\left(-\frac{\Delta B_i^2}{2\Delta t_i}\right) = \Delta t_i$$

and so:

$$\langle (\Delta t_i - (\Delta B_i)^2) \rangle = 0$$

So we are left only with the first term of (11):

$$(11) = \frac{1}{4}\sum_{i=1}^n \langle [\Delta t_i - (\Delta B_i)^2]^2 \rangle = \frac{1}{4}\sum_{i=1}^n \left[\Delta t_i^2 - 2\Delta t_i \underbrace{\langle (\Delta B_i)^2 \rangle}_{\Delta t_i} + \langle \Delta B_i^4 \rangle \right] \tag{12}$$

Recall that, for a random variable x sampled from a gaussian $\mathcal{N}(0, \sigma)$:

$$\langle x^{2n} \rangle = \sigma^{2n} \frac{(2n)!}{2^n n!} = \begin{cases} \sigma^2 & n = 1 \\ \sigma^4 \frac{4!}{4 \cdot 2!} = 3\sigma^4 & n = 2 \end{cases}$$

In our case, this means that $\langle (\Delta B_i)^4 \rangle = \Delta t_i^2$, leading to:

$$(12) = \frac{1}{2}\sum_{i=1}^n \Delta t_i^2$$

When taking the limit of the mesh ($n \to \infty$), the number of summed terms become infinite, but also the size of each of them vanishes:

$$\max_i \Delta t_i \xrightarrow[n \to \infty]{} 0$$

To resolve that limit we need to use the fact that the end-point is fixed ($t_n \equiv t$) and so:

$$\frac{1}{2}\sum_{i=1}^n \Delta t_i^2 \leq \frac{1}{2}\left(\sum_{i=1}^n \Delta t_i\right)^2 = \frac{1}{2}\left(\sum_{i=1}^n \Delta t_i\right)\underbrace{\left(\sum_{j=1}^n \Delta t_j\right)}_{t} \leq \frac{t}{2}\left(\max_i \Delta t_i\right) \xrightarrow[n\to\infty]{} 0$$

This proves (9), and so the desired result (8).

Stratonovich's prescription. In this case, we want to show that:

$$S_n = \sum_{i=1}^n B\left(\frac{t_i + t_{i-1}}{2}\right)[B(t_i) - B(t_{i-1})] \xrightarrow[n\to\infty]{\text{m.s.}} \frac{B^2(t) - B^2(0)}{2}$$

Note that now we need a set of *middle points* in the mesh, which leads to some complications.
One trick is to simply *double* the "resolution" of the discretization, and choose the *middle points* to be the *odd* indices. We then define:

$$S'_{2n} = \sum_{i=1}^{2n} B_{2i-1}(B_{2i} - B_{2(i-1)})$$

with $t_{2i-1} \equiv (t_{2i} + t_{2(i-1)})/2$, while the t_{2i} may be distributed arbitrarily. The full computation is very long and tedious, and not much enlightening, and is therefore omitted.

A shorter way to compute that, but not as rigorous, is by stating that:

$$S_n = \sum_{i=1}^n \frac{B(t_i) + B(t_{i-1})}{2}(B(t_i) - B(t_{i-1}))$$

However it is not obvious that is possible to *approximate* a midpoint of B with an average, as $B(t_i$ are all random variables. In fact, it is possible to show that the two expressions *have the same distribution*, but they *are not the same random variable*! In any way, if we do this, the thesis immediately follows:

$$= \frac{1}{2}\sum_{i=1}^n (B^2(t_i) - B^2(t_{i-1}))$$

0.2.3 Ito's calculus

In our calculations, we will be usually concerned with the following kinds of stochastic integrals $G(t)$:

1. $\int_0^t F(B(\tau))\, \mathrm{d}B(\tau)$

2. $\int_0^t g(\tau)\,\mathrm{d}B(\tau)$

3. $\int_0^t g(\tau)\,\mathrm{d}\tau$ (usual integrals)

These $G(t)$ are called **non-anticipating functions**, because they are independent of $B(t') - B(t)$ for $t' > t$, meaning that they do not dependent on what happens in the Brownian motion at times later than t (i.e. they do not depend on the future). So, by using Ito's prescription (I.p.) in the discretization and *mean square* (m.s.) for the continuum limit we get:

$$\int_0^t F(B(\tau))\,\mathrm{d}B(\tau) \underset{\text{m.s.}}{\overset{\text{I.p.}}{=}} \sum_{i=1}^{n} F(B_{i-1})\Delta B_i$$

Note how $F(B_{i-1})$ and ΔB_i are independent of each other, simplifying the calculations. (Note that the Stratonovich prescription here causes *troubles* during evaluation, as it introduces some interdependence between different terms).