

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 8

8. BROWNIAN MOTION

We will be spending the rest of the course on Brownian motion and integration with respect to Brownian motion (stochastic integrals). The topics in the book are closely interrelated so we need to go over everything in the book plus additional material such as Lévy's remarkable theorem which I will explain today. Here is an outline of the chapter.

- (0) Definition of Brownian motion
- (1) Martingales and Lévy's theorem
- (2) Strong Markov property and reflection principle
- (3) Fractal dimension of the zero set
- (4) Brownian motion and the heat equation in several dimensions
- (5) Recurrence and transience
- (6) Fractal dimension of the path
- (7) Scaling and the Cauchy distribution
- (8) Drift

8.0. Definition of Brownian motion. First of all this is a random process in continuous time and continuous space. We will start with dimension one: Brownian motion on the real line.

The idea is pretty simple. A particle is bouncing around and its position at time t is X_t . The process is

- (1) *memoryless*: What happens for time $s > t$ depends only on its position X_t and not on how it got there.
- (2) *time and space homogeneous*: The behavior of the particle remains the same if we reset both time and space coordinates. I.e., the distribution of $Y_t = X_{s+t} - X_s$ depends only on t and is independent of the time s and position X_s .
- (3) *continuity*: The particle moves on a continuous path (without jumping from one point to another).

These conditions almost guarantee that we have Brownian motion. But we need a little more. Here is the definition.

8.0.1. *definition.*

Definition 8.1. A random function $X : [0, \infty) \rightarrow \mathbb{R}$ written X_t is Brownian motion with variance σ^2 starting at 0 if:

- (1) $X_0 = 0$
- (2) For any $s < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n$ the random variables $X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}, \dots, X_{t_n} - X_{s_n}$ are independent.
- (3) The path X_t is continuous
- (4) For $s < t$,

$$X_t - X_s \sim N(0, (t - s)\sigma^2)$$

i.e., $X_t - X_s$ is normally distributed with mean 0 and variance $(t - s)\sigma^2$.

Theorem 8.2. The last condition is equivalent to the condition:

- (4') $X_t - X_s$ and $X_{t+c} - X_{s+c}$ are identically distributed with mean 0 and variance $(t - s)\sigma^2$.

Proof. (Outlined in the book.) Certainly (4) \Rightarrow (4'). To prove the converse, assume (4'). Let $\Delta t = (t - s)/N$ for large N . Then

$$X_t - X_s = (X_{s+\Delta t} - X_s) + (X_{s+2\Delta t} - X_{s+\Delta t}) + \dots + (X_t - X_{t-\Delta t})$$

This is a sum of N independent identically distributed random variables with mean 0 and variance $(\Delta t)\sigma^2$. By the central limit theorem we get

$$X_t - X_s \approx N(0, N\Delta t\sigma^2) = N(0, (t - s)\sigma^2)$$

Now take the limit as $N \rightarrow \infty$. (This is not rigorous because we are not using the precise statement of the CLT.) \square

Recall that the *variance* of a random variable X is defined by

$$\text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

and it has the property that it is additive for *independent* random variables:

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

8.0.2. *as limit of random walk.* Brownian motion can be obtained as a limit of random walks: Take time to be integer multiples of a fixed interval Δt and we take points on the line at integer multiples of $\sigma\sqrt{\Delta t}$. For each unit time assume that position changes by $\pm\sigma\sqrt{\Delta t}$ with equal probability. This is Bernoulli with mean 0 and variance

$$(\pm\sigma\sqrt{\Delta t})^2 = \sigma^2\Delta t$$

In a time interval $N\Delta t$ the change of position is given by a sum of N independent random variables. So, the mean would be 0 and the

variance would be $N\Delta t\sigma^2$. The point is that this is σ^2 times the length of the time interval. As $\Delta t \rightarrow 0$, assuming the sequence of random walks converges to a continuous function, the limit gives Brownian motion with variance σ^2 by the theorem.

8.0.3. *nowhere differentiable.* Notice that, as Δt goes to zero, the change in position is approximately $\sigma\sqrt{\Delta t}$ which is much bigger than Δt . This implies that the limit

$$\lim_{t \rightarrow 0} \frac{X_t}{t}$$

diverges. So, Brownian motion is, almost surely, nowhere differentiable. (*Almost surely* or a.s. means “with probability one.”)

8.1. Martingales and Lévy’s theorem.

Theorem 8.3. *Suppose that X_t is Brownian motion. Then*

- (1) X_t is a continuous martingale.
- (2) $X_t^2 - t\sigma^2$ is a martingale

Proof. (1) is easy: If $t > s$ then

$$\begin{aligned} \mathbb{E}(X_t | \mathcal{F}_s) &= \mathbb{E}(X_t - X_s | \mathcal{F}_s) + \mathbb{E}(X_s | \mathcal{F}_s) \\ &= 0 + X_s \end{aligned}$$

where \mathcal{F}_s is the information contained in X_r for all $r \leq s$.

For (2), we need the equation:

$$\begin{aligned} (X_t - X_s)^2 &= X_t^2 - 2X_tX_s + 2X_s^2 - X_s^2 \\ &= X_t^2 - 2(X_t - X_s)X_s - X_s^2 \end{aligned}$$

Taking $\mathbb{E}(\cdot | \mathcal{F}_s)$ of both sides gives:

$$\begin{aligned} \mathbb{E}((X_t - X_s)^2 | \mathcal{F}_s) &= \text{Var}(X_t - X_s) = (t - s)\sigma^2 = \\ \mathbb{E}(X_t^2 | \mathcal{F}_s) - 2\mathbb{E}(X_t - X_s | \mathcal{F}_s)X_s - X_s^2 &= \mathbb{E}(X_t^2 | \mathcal{F}_s) - X_s^2 \end{aligned}$$

Which gives

$$\mathbb{E}(X_t^2 - t\sigma^2 | \mathcal{F}_s) = X_s^2 - s\sigma^2$$

□

Lévy’s theorem is the converse:

Theorem 8.4 (Lévy). *Suppose that X_t is a continuous martingale and $X_t^2 - t\sigma^2$ is also a martingale. Then X_t is Brownian motion with variance σ^2*

This famous theorem has been proved many times. I will try to find the proof using stochastic integrals. One amazing consequence is the following.

Corollary 8.5. *Any continuous martingale M_t is Brownian motion reparametrized and starting at $C = M_0$. I.e.*

$$M_t = X_{\phi(t)} + C$$

where X_s is standard Brownian motion (with $\sigma = 1$).

Proof. (When I did this in class I forgot to “center” the martingale by subtracting M_0 .) The idea is to let $\phi(t) = \mathbb{E}((M_t - C)^2)$ and apply Lévy’s theorem. (I’ll look for the details). \square

8.2. Strong Markov property and Reflection principle.

Theorem 8.6 (strong Markov property). *Let T be a stopping time for Brownian motion X_t . Let*

$$Y_t = X_{t+T} - X_T$$

Then Y_t is independent of \mathcal{F}_T (for $t > 0$).

One consequence of this is the *reflection principle* which the book uses over and over.

8.2.1. reflection principle.

Corollary 8.7 (reflection principle). *Suppose that $a < b$ then the probability that you will reach b from a within time t is twice the probability that at time t you will be past b . I.e.:*

$$\mathbb{P}(X_s = b \text{ for some } 0 < s < t \mid X_0 = a) = 2\mathbb{P}(X_t > b \mid X_0 = a)$$

Proof. If you reach the point b at some time before time t then half the time you will end up above b and half the time you will end up below b since the probability that $X_t = b$ is zero. So,

$$\begin{aligned} \mathbb{P}(X_s \text{ reaches } b \text{ sometime before } t \text{ and ends up higher} \mid X_0 = a) \\ = \frac{1}{2}\mathbb{P}(X_s = b \text{ for some } 0 < s < t \mid X_0 = a) \end{aligned}$$

But the event “ X_s reaches b sometime before t and ends up higher” is the same as the event “ $X_t > b$ ” since X_t is continuous and therefore cannot get to a point $X_t > b$ starting at $a < b$ without passing through b . This proves the reflection principle.

Why is the reflection principle a corollary of the strong Markov property? The reason is that we are using the stopping time $T =$ the first time that $X_T = b$. And $Y = X_t - X_T$. For every fixed T this is normally distributed with mean 0 and variance $(t - T)\sigma^2$. So,

$$\mathbb{P}(Y > 0 \mid T < t) = \frac{1}{2}$$

By the formula for conditional probability, this is

$$\frac{1}{2} = \mathbb{P}(Y > 0 | T < t) = \frac{\mathbb{P}(Y > 0 \text{ and } T < t)}{\mathbb{P}(T < t)}$$

But “ $Y > 0$ and $T < t$ ” is the same as “ $X_t > b$ ” and “ $T < t$ ” is the same as “ X_s reaches b sometime before t .” So, this gives the reflection principle again. \square

8.2.2. *density function.* If $X_0 = a$, then $X_t - a \sim N(0, t\sigma^2)$. The normal distribution $N(0, t\sigma^2)$ has density function

$$\phi_t(x) = \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-x^2/2\sigma^2 t}$$

The probability density function for X_t is given by shifting the normal distribution by $a = X_0$.

$$f_{X_t}(x) = \phi_t(x - a) = p_t(a, x) = \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(x-a)^2/2\sigma^2 t}$$

It is called $p_t(a, x)$ because it is the (infinitesimal) transition matrix:

$$p_t(a, x)dx = \mathbb{P}(x < X_t \leq x + dx | X_0 = a)$$

The integral of this over any interval I is equal to the probability that X_t will lie in I . E.g.,

$$\mathbb{P}(X_t > b | X_0 = a) = \int_b^\infty p_t(a, x) dx = \int_b^\infty \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(x-a)^2/2\sigma^2 t} dx$$

In the reflection principle we get twice this number:

$$\mathbb{P}(X_s = b \text{ for some } 0 < s < t | X_0 = a) = 2 \int_b^\infty \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(x-a)^2/2\sigma^2 t} dx$$

If we make the substitution

$$y = \frac{x - a}{\sigma\sqrt{t}}, \quad dy = \frac{dx}{\sigma\sqrt{t}}$$

we get the standard normal distribution:

$$2 \int_{(b-a)/\sigma\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

This is the well-know rule: *To convert to standard normal, subtract the mean then divide by the standard deviation.* It works for integrals.

8.2.3. *The Chapman-Kolmogorov equation.* is an example of a formula which is easy to understand using Brownian motion:

$$p_{s+t}(x, y) = \int_{-\infty}^{\infty} p_s(x, z)p_t(z, y) dz$$

This is just a continuous version of the matrix equation

$$P_{s+t} = P_s P_t$$

and it holds for *all Markov processes*.

In the particular case of Brownian motions, the integral is a convolution and the Chapman-Kolmogorov equation can be rewritten as:

$$\phi_{s+t} = \phi_s * \phi_t$$

As I explained to you last week, convolution of density functions gives the density function for the sum of two random variables. In this case:

$$pdf(X_{s+t} - X_0) = pdf(X_s - X_0) * pdf(X_{t+s} - X_s)$$

8.2.4. *example 1.* Here we want the probability that standard Brownian motion, starting at 0, will return to 0 sometime between time 1 and time $t > 1$.

$$\mathbb{P}(X_s = 0 \text{ for some } 1 < s < t \mid X_0 = 0) = ?$$

We first look at where the particle is at time 1. Half the time X_1 will be positive and half the time it will be negative. So, we will assume that $X_1 > 0$ then multiply by 2. By the reflection principle, the probability of returning to 0 before time t is twice the probability that $X_t < 0$. So, the answer will be

$$\begin{aligned} & 4\mathbb{P}(X_1 > 0 \text{ and } X_t < 0 \mid X_0 = 0) \\ & = 4\mathbb{P}(X_1 = b > 0 \text{ and } X_t - X_1 < -b \mid X_0 = 0) \end{aligned}$$

The probability for fixed b (in the interval $(b, b + db]$) is

$$\phi_1(b)db \Phi_{t-1}(-b)$$

where Φ_{t-1} is the cumulative distribution function:

$$\Phi_{t-1}(-b) = \int_{-\infty}^{-b} \phi_{t-1}(x) dx = \int_b^{\infty} \phi_{t-1}(x) dx = \int_{b/\sqrt{t-1}}^{\infty} \phi_1(y) dy$$

where we used the “convert to standard normal” rule. The answer is now given by integrating over all $b > 0$ and multiplying by 4:

$$4 \int_0^{\infty} \int_{b/\sqrt{t-1}}^{\infty} \phi_1(b)\phi_1(y) dydb$$

The integrand is

$$\phi_1(b)\phi_1(y) = \frac{1}{\sqrt{2\pi}} e^{-b^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(b^2+y^2)/2}$$

Now convert to polar coordinates:

$$\phi_1(b)\phi_1(y) dy db = \frac{1}{2\pi} e^{-r^2/2} r dr d\theta$$

The answer is

$$\begin{aligned} & 4 \int_0^\infty \int_{\arctan(1/\sqrt{t-1})}^{\pi/2} \frac{1}{2\pi} e^{-r^2/2} r d\theta dr \\ &= 4 \left(\pi/2 - \arctan(1/\sqrt{t-1}) \right) \frac{1}{2\pi} \int_0^\infty e^{-r^2/2} r dr \\ &= 1 - \frac{2}{\pi} \arctan \frac{1}{\sqrt{t-1}} \end{aligned}$$

8.2.5. *example 2.* In this example we have to show that, a.s.,

$$\lim_{t \rightarrow \infty} \frac{X_t}{t} = 0$$

where X_t is standard Brownian motion.

First, let $t = n$ be an integer going to infinity. Then

$$X_n = (X_1 - X_0) + (X_2 - X_1) + \cdots + (X_n - X_{n-1})$$

This is a sum of n iid random variables. So, by the strong law of large numbers, the average will converge to the expected value with probability one:

$$\lim_{n \rightarrow \infty} \frac{X_n}{n} = \mathbb{E}(X_1 - X_0) = 0$$

Next, we have to show that, as t goes from one integer n to the next $n + 1$, X_t doesn't deviate too far from X_n . What the book shows is that, a.s., for all but a finite number of n , the difference

$$|X_t - X_n| < 2\sqrt{\ln n} = a$$

When we divide by n the difference between X_t and X_n will go to zero. Dividing by t is even better because $t > n$.

We want the probability that at some time $t \in (n, n + 1)$, $X_t - X_n$ goes above a or below $-a$. By symmetry this is \leq twice the probability that it will go above a . By the reflection principle this is 4 times the probability that it will end up above a :

$$\begin{aligned} \mathbb{P}(|X_t - X_n| > a \text{ for some } n < t < n + 1) &\leq 4\mathbb{P}(X_{n+1} - X_n > a) \\ &= 4 \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq 4 \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-ax/2} \end{aligned}$$

$$= \frac{8}{a\sqrt{2\pi}} e^{-a^2/2} = \frac{8}{2\sqrt{\ln n}\sqrt{2\pi n^2}}$$

(because $-a^2/2 = -2\ln n = \ln n^{-2}$)

This is an upper bound for the probability that X_t deviates too far from X_n for a single n . If we sum over all n we get an upper bound for the expected number of times that this will happen. But

$$\sum_{n=2}^{\infty} \frac{8}{2\sqrt{\ln n}\sqrt{2\pi n^2}} \leq C \sum \frac{1}{n^2} < \infty$$

So, the expected number is finite. So, a.s., the deviation occurs only finitely many times. (If there were a nonzero probability p that the event occurs an infinite number of times then the expected number would be at least $p \cdot \infty = \infty$.)

8.3. Fractal dimension of the zero set. The *zero set* Z is just the set of all times t so that $X_t = 0$. This is a subset of the positive real line: $Z \subset [0, \infty)$. The zero set is a *fractal* in the sense that it looks the same on the small scale as it does on the big scale. The “fractal dimension” of the set measures the scale at which the set is self-similar. We use the box dimension definition.

8.3.1. *self-similarity of Z .*

Theorem 8.8. *Suppose that X_t is Brownian motion with variance σ^2 . Then*

- (1) $Y_t = bX_{at}$ is Brownian motion with variance $b^2a\sigma^2$ ($= \sigma^2$ if $b^2 = 1/a$).
- (2) $Y_t = tX_{1/t}$ is Brownian motion with variance σ^2 .

Proof. For (2) you need to use the fact that, a.s.,

$$\lim_{t \rightarrow 0} tX_{1/t} = \lim_{1/t \rightarrow \infty} \frac{X_{1/t}}{1/t} = 0$$

Therefore, $Y_t = tX_{1/t}$ is continuous at $t = 0$. This settles the continuity condition. The other conditions are clear: Since the functions at and $1/t$ are monotonic, Y_t is a memoryless process in both cases. We just have to calculate the variance (and see that it is constant). This is easy:

$$\text{Var}(bX_{at}) = \mathbb{E}(b^2X_{at}^2) = b^2at\sigma^2$$

$$\text{Var}(tX_{1/t}) = \mathbb{E}(t^2X_{1/t}^2) = t^2(1/t)\sigma^2 = t\sigma^2$$

Here is a more careful proof (mainly for my reference).

We have to show that $Y_t - Y_s$ is normally distributed with variance proportional to $t - s$. (It obviously has zero mean.) In case (1) this is easy:

$$Y_t - Y_s = bX_{at} - bX_{as} \sim bN(0, (at - as)\sigma^2) = N(0, b^2a(t - s)\sigma^2)$$

In case (2) we have:

$$\begin{aligned} Y_t - Y_s &= tX_{1/t} - sX_{1/s} = (t - s)X_{1/t} + s(X_{1/t} - X_{1/s}) \\ &\sim N(0, (t - s)^2 \frac{1}{t} \sigma^2) + N(0, s^2 \left(\frac{1}{t} - \frac{1}{s}\right) \sigma^2) \end{aligned}$$

Then use the fact that the sum of independent normal distributions is normal with mean the sum of the means and variance the sum of the variances. Then calculate:

$$(t - s)^2 \frac{1}{t} \sigma^2 + s^2 \left(\frac{1}{t} - \frac{1}{s}\right) \sigma^2 = (t - s)\sigma^2$$

□

What does this mean in terms of the set Z ?

First of all, if we multiply X_t by a constant, the zero set is unchanged since $bX_t = 0 \Leftrightarrow X_t = 0$. Therefore, the theorem says:

- (1) Z looks like aZ for any positive constant a .
- (2) Z looks like $1/Z$.
- (3) Z does not depend of the variance σ^2 . (So, we assume $\sigma^2 = 1$.)

When I say “looks like” I mean it “has the same probability distribution as.”

8.3.2. *gaps in Z* . Example 1 from 8.2 calculates the probability that Z meets the set $[1, t]$

$$\mathbb{P}(Z \cap [1, t] \neq \emptyset) = 1 - \frac{2}{\pi} \arctan\left(\frac{1}{\sqrt{t-1}}\right)$$

or:

$$\mathbb{P}(Z \cap [1, t] = \emptyset) = \frac{2}{\pi} \arctan\left(\frac{1}{\sqrt{t-1}}\right)$$

This is equal to $1/2$ when $t = 2$. And as $t \rightarrow \infty$ this probability goes to zero ($\arctan 0 = 0$). So,

$$\mathbb{P}(Z \cap [1, \infty) = \emptyset) = 0$$

The scaling theorem now says the same is true for any rescaling:

$$\mathbb{P}(Z \cap [t, 2t] = \emptyset) = \frac{1}{2}$$

and

$$\mathbb{P}(Z \cap [t, \infty) = \emptyset) = 0$$

for any $t > 0$.

8.3.3. *fractal dimension.* First I'll explain the simple version of the idea and do some examples. Then I'll give the mathematical definition.

Take the unit square. If we scale this by a factor of 10 then we get something which can be cut into $10^2 = 100$ squares. If we take a cube and scale by 10 we get $10^3 = 1,000$ cubes. The dimension is equal to the exponent that we need to take the scaling factor to. For the Cantor set C , if you scale by a factor of 3 then you get 2 Cantor sets. So, its dimension D is the solution of the equation

$$2 = 3^D$$

Taking the log of both sides gives

$$D = \dim C = \frac{\ln 2}{\ln 3} \approx 0.631$$

Instead of scaling the object and seeing how big it gets, you could just as well scale the units down and see how many smaller units you need. For example take a unit square. How many little $1/10 \times 1/10$ squares do you need to cover it? It is $10^2 = (1/10)^{-2}$ just like before except that now the dimension is the negative of the exponent of the scale of the little pieces. It is the same concept.

Definition 8.9. *The box dimension of a bounded subset A of \mathbb{R}^d is equal to the infimum of $D > 0$ so that as $\epsilon \rightarrow 0$, the number of cubes with sides ϵ needed to cover A becomes $< C\epsilon^{-D}$ where C is a constant.*

The set A needs to be bounded otherwise you need an infinite number of little cubes to cover it.

8.3.4. *dimension of Z .* Take the bounded set $Z_1 = Z \cap [0, 1]$. What is the expected number of intervals of length $\epsilon = 1/n$ needed to cover Z_1 ? It should be $\sim n^D$ where D is the dimension of Z_1 .

The expected number of intervals needed is equal to the sum of probabilities

$$\begin{aligned} & \mathbb{E}(\text{number of intervals } [k/n, (k+1)/n] \text{ that meet } Z_1) \\ &= \sum_{k=0}^{n-1} \mathbb{P}(Z_1 \cap \left[\frac{k}{n}, \frac{k+1}{n}\right] \neq \emptyset) \end{aligned}$$

But the scaling theorem tells us that

$$\begin{aligned} \mathbb{P}(Z_1 \cap \left[\frac{k}{n}, \frac{k+1}{n}\right] \neq \emptyset) &= \mathbb{P}(Z \cap \left[1, \frac{k+1}{k}\right] \neq \emptyset) \\ &= 1 - \frac{2}{\pi} \arctan \sqrt{k} \end{aligned}$$

So, the expected number of intervals is

$$\sum_{k=0}^{n-1} 1 - \frac{2}{\pi} \arctan \sqrt{k}$$

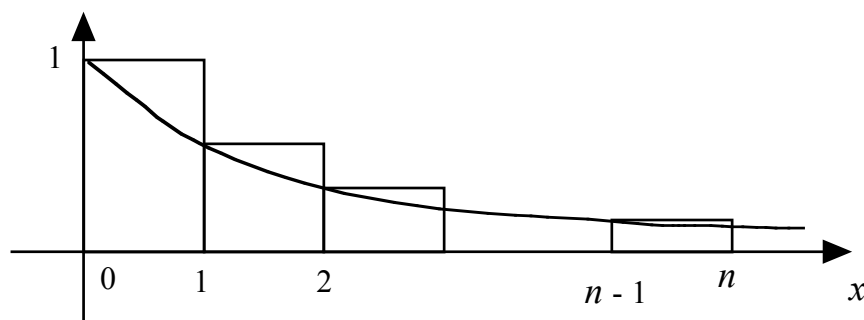


FIGURE 1. The sum is equal to the integral plus the little triangles which can be stacked up to give about $1/2$

This is a Riemann sum. (See Figure 1.) So, it is approximately equal to

$$\begin{aligned} & \frac{1}{2} + \int_0^n 1 - \frac{2}{\pi} \arctan \sqrt{x} dx \\ &= \frac{1}{2} + n - \frac{2}{\pi} (n+1) \arctan \sqrt{n} + 2\sqrt{n} \end{aligned}$$

Using the approximation

$$\arctan x \approx \frac{\pi}{2} - \frac{1}{x}$$

this becomes

$$\approx \frac{1}{2} + n - \left(n+1 - \frac{2n+2}{\pi\sqrt{n}} \right) + 2\sqrt{n} \approx \sqrt{n} \left(2 - \frac{2}{\pi} \right) \approx 1.36\sqrt{n}$$

The dimension of Z_1 is the exponent of n which is $D = 1/2$.

8.4. Brownian motion and the heat equation in several dimensions. (When you read this section ask yourself: What do people mean when they say:

“The infinitesimal generator of Brownian motion is $\frac{1}{2} \frac{\partial^2}{\partial x^2}$ ” ?)

8.4.1. *definition.* With several variables, Brownian motion can be written in rectangular or polar coordinates. I prefer the version which is obviously rotationally invariant. (You can rotate the coordinates and the equation does not change.)

Definition 8.10. Standard d -dimensional Brownian motion is a vector valued stochastic process, i.e., random function $\mathbf{X} : [0, \infty) \rightarrow \mathbb{R}^d$ so that

- (1) $\mathbf{X}_0 = 0$
- (2) For any $s < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n$ the random variables $\mathbf{X}_{t_1} - \mathbf{X}_{s_1}, \mathbf{X}_{t_2} - \mathbf{X}_{s_2}, \dots, \mathbf{X}_{t_n} - \mathbf{X}_{s_n}$ are independent.
- (3) The path \mathbf{X}_t is continuous
- (4) For $s < t$, the random variable $\mathbf{X}_t - \mathbf{X}_s$ has density function

$$\phi_r(\mathbf{x}) = \frac{1}{(2\pi r)^{d/2}} e^{-\|\mathbf{x}\|^2/2r}$$

where $r = t - s$ and $\|\mathbf{x}\|$ is the length of the vector $\mathbf{x} \in \mathbb{R}^d$.

The coordinates $X_t^j - X_s^j$ of the vector $\mathbf{X}_t - \mathbf{X}_s$ are independent standard 1-dimensional Brownian motions with densities

$$\frac{1}{(2\pi r)^{1/2}} e^{-x_j^2/2r}$$

whose product is $\phi_r(\mathbf{x})$. The elapsed time or time increment is denoted $r = t - s$ in the definition. The *covariance matrix* is the $d \times d$ matrix

$$[\mathbb{E}((X_t^i - X_s^i)(X_t^j - X_s^j))] = (t - s)\mathbf{I}$$

The *transition density* for time increment t is

$$p_t(\mathbf{x}, \mathbf{y}) = \phi_t(\mathbf{y} - \mathbf{x}) = \frac{1}{(2\pi t)^{d/2}} e^{-\|\mathbf{y} - \mathbf{x}\|^2/2t}$$

One important feature is that the transition density is *symmetric*, i.e.,

$$p_t(\mathbf{x}, \mathbf{y}) = p_t(\mathbf{y}, \mathbf{x})$$

This satisfies the *Chapman-Kolmogorov* equation just like any other transition density.

8.4.2. *diffusion.* If you have a large number of particles moving independently according to the rules of Brownian motion, then the distribution of these particles will change in a deterministic process called *diffusion*.

Let $f_t(\mathbf{x})$ denote the density of particles at time t and position \mathbf{x} . After an increment of time δt , the density will change to

$$(8.1) \quad f_{t+\delta t}(\mathbf{y}) = \int_{\mathbb{R}^d} f_t(\mathbf{x}) p_{\delta t}(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}$$

where I used the abbreviation $d\mathbf{x} = dx_1 dx_2 \cdots dx_d$. Since $p_{\delta t}(\mathbf{x}, \mathbf{y}) = p_{\delta t}(\mathbf{y}, \mathbf{x})$ we can rewrite this as

$$f_{t+\delta t}(\mathbf{y}) = \int_{\mathbb{R}^d} f_t(\mathbf{x}) p_{\delta t}(\mathbf{y}, \mathbf{x}) \, d\mathbf{x}$$

Now switch “ \mathbf{x} ” and “ \mathbf{y} ”:

$$(8.2) \quad f_{t+\delta t}(\mathbf{x}) = \int_{\mathbb{R}^d} f_t(\mathbf{y}) p_{\delta t}(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}$$

These equations look similar but they mean different things. The first equation (8.1) gives the density of particles as an sum over all places where the particles came from. Equation (8.2) says that the future density at \mathbf{x} will be equal to the expected value of the present density function of the new (random) location of a single particle starting at the point \mathbf{x} . The first equation is deterministic and the second is probabilistic!

Equation (8.2) can be written:

$$f_{t+\delta t}(\mathbf{x}) = \mathbb{E}^{\mathbf{x}}(f_t(\mathbf{X}_{\delta t})) = \mathbb{E}(f_t(\mathbf{X}_{\delta t}) \mid \mathbf{X}_0 = \mathbf{x})$$

where we use the abbreviation $\mathbb{E}^{\mathbf{x}} = \mathbb{E}(- \mid \mathbf{X}_0 = \mathbf{x})$. I changed the equation from that in the book to clarify what is absolute time and what is relative time.

8.4.3. *the differential equation.* Now take the limit as $\delta t \rightarrow 0$:

$$(8.3) \quad \frac{\partial}{\partial t} f_t(\mathbf{x}) = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \mathbb{E}^{\mathbf{x}}(f_t(\mathbf{X}_{\delta t}) - f_t(\mathbf{X}_0))$$

On the RHS we are taking density at a fixed time and variable position. The book estimates the density first in the case $d = 1$:

$$f_t(X_{\delta t}) = f_t(X_0) + \frac{\partial}{\partial x} f_t(x)(X_{\delta t} - X_0) + \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x)(X_{\delta t} - X_0)^2 + o((X_{\delta t} - X_0)^2)$$

Now take expected value

$$\mathbb{E}^x(f_t(X_{\delta t}) - f_t(X_0)) = \frac{\partial}{\partial x} f_t(x) \mathbb{E}^x(X_{\delta t} - X_0) + \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x) \mathbb{E}^x((X_{\delta t} - X_0)^2) + o(\delta t)$$

But $\mathbb{E}^x(X_{\delta t} - X_0) = 0$ and $\mathbb{E}^x((X_{\delta t} - X_0)^2) = (\delta t - 0)\sigma^2 = \delta t$. So,

$$\mathbb{E}^x(f_t(X_{\delta t}) - f_t(X_0)) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x) \delta t + o(\delta t)$$

Dividing by δt and taking the limit as $\delta t \rightarrow 0$ gives

$$\frac{\partial}{\partial t} f_t(x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x)$$

For the random vector $\mathbf{X}_{\delta t}$ we get:

$$\begin{aligned} & f_t(\mathbf{X}_{\delta t}) - f_t(\mathbf{X}_0) = \\ & \sum_i \frac{\partial}{\partial x_i} f_t(\mathbf{x})(X_{\delta t}^i - X_0^i) + \sum_{i,j} \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} f_t(\mathbf{x})((X_{\delta t}^i - X_0^i)(X_{\delta t}^j - X_0^j)) + o(\delta t) \end{aligned}$$

Taking expected value, the first term gives zero. The expected value of the second term is given by the covariance matrix $\delta t \mathbf{I}$. So

$$\mathbb{E}^x(f_t(\mathbf{X}_{\delta t}) - f_t(\mathbf{X}_0)) = \sum_i \frac{1}{2} \frac{\partial^2}{\partial x_i^2} f_t(\mathbf{x}) \delta t + o(\delta t)$$

Referring back to the original differential equation (8.3) we get

$$\frac{\partial}{\partial t} f_t(\mathbf{x}) = \frac{1}{2} \Delta f_t(\mathbf{x})$$

where Δ is the *Laplacian*

$$\Delta = \sum_{i=1}^d \frac{1}{2} \frac{\partial^2}{\partial x_i^2}$$

If the particles are moving according to Browning motion with variance σ^2 then the density changes according to the equation

$$(8.4) \quad \frac{\partial}{\partial t} f_t(\mathbf{x}) = \frac{D}{2} \Delta f_t(\mathbf{x})$$

where $D = \sigma^2$.

8.4.4. *heat equation.* The equation (8.4) is called the *heat equation* with *diffusion constant* D . We will see how the time reversal trick explained above can be used to solve this equation using stochastic methods.

Suppose that B is a region in \mathbb{R}^d with boundary ∂B . Suppose we start with a heat density function f on B which changes according to the heat equation and a constant heat density of g on ∂B . If the temperature in the interior point \mathbf{x} at time t is $u(t, \mathbf{x})$ then the probabilistic interpretation is that $u(t, \mathbf{x})$ is equal to the expected value of the density at time 0 at the position a particle starting at \mathbf{x} will end up at time t .

The boundary, which in forward time is emitting heat at a constant rate, will, in backward time, act like flypaper for randomly moving particles. The particle starts at $\mathbf{X}_0 = \mathbf{x}$ and moves according to Brownian motion and stops the first time it hits the boundary ∂B . This is a stopping time. Call it $\tau = \tau_{\partial B}$.

The equation for this expected value is

$$u(t, \mathbf{x}) = \mathbb{E}^{\mathbf{x}}(g(\mathbf{X}_{\tau})I(\tau < t) + f(\mathbf{X}_t)I(t \leq \tau))$$

As t goes to infinity, the temperature reaches a steady state given by

$$v(\mathbf{x}) = \mathbb{E}^{\mathbf{x}}(g(\mathbf{X}_{\tau}))$$

So, this is the solution of the equation:

$$\Delta v(\mathbf{x}) = 0$$

(in the interior of B) with boundary condition $v = g$ on ∂B .

8.4.5. *example 1: probabilistic method.* Let $d = 1$ and $B = (a, b) \subset \mathbb{R}$ where $0 \leq a < b < \infty$. Then the boundary is just two points $\partial B = \{a, b\}$. Suppose the function on the boundary is $g(a) = 0, g(b) = 1$. We start at some point $x \in (a, b)$ and stopping time τ is when we reach either a or b . This is the “gambler’s ruin” because it describes what happens to a gambler playing a fair game who starts with $\$x$ and quits when he reaches either a or b .

$$v(x) = \mathbb{E}^x(g(X_{\tau})) = \mathbb{P}^x(X_{\tau} = b)$$

By the strong Markov property, $\mathbb{E}^x(X_{\tau}) = \mathbb{E}^x(X_0) = x$. So,

$$\mathbb{E}^x(X_{\tau}) := a\mathbb{P}^x(X_{\tau} = a) + b\mathbb{P}^x(X_{\tau} = b) = a(1 - v(x)) + bv(x) = x$$

$$v(x) = \frac{x - a}{b - a}$$

8.4.6. *example 2: analytic method.* Now consider the case when $g(0) = g(2\pi) = 0$ and $B = (0, \pi)$. The heat equation

$$\frac{\partial}{\partial t}u(t, x) = \frac{1}{2} \frac{\partial^2}{\partial x^2}v(t, x)$$

is homogeneous and has a complete list of basic solutions:

$$u(t, x) = e^{-tn^2/2} \sin(nx)$$

and any solution is an infinite linear combination

$$u(t, x) = \sum_{n=1}^{\infty} C_n e^{-tn^2/2} \sin(nx)$$

Where does this come from? The idea is to write $u(t, x) = \psi(t)\phi(x)$ where

$$\frac{\partial}{\partial t}\psi(t) = -\frac{\lambda^2}{2}\psi(t) \quad (\psi(t) = e^{-\lambda^2 t/2} \text{ works})$$

$$\Delta\phi(x) = -\lambda^2\phi(x) \quad (\phi(x) = \sin(\lambda x) \text{ works as does } \cos(\lambda x))$$

Then $u(t, x) = \psi(t)\phi(x)$ is a solution of the heat equation.

Start with $f(x) = \delta_y(x)$ being a Dirac delta function at some point $y \in (0, 2\pi)$. This means that

$$\delta_y(x) = \sum C_n \sin(nx)$$

To determine the coefficients C_n we multiply by $\sin(mx)$ and integrate:

$$\int_0^\pi \delta_y(x) \sin(mx) dx = \sum \int_0^\pi C_n \sin(nx) \sin(mx) dx$$

$$\sin(my) = \frac{\pi}{2} C_m$$

So $C_n = \frac{2}{\pi} \sin(ny)$ and

$$u(t, x) = \sum_{n=1}^{\infty} \frac{2}{\pi} e^{-tn^2/2} \sin(ny) \sin(nx)$$

The book points out that one of these terms lasts longer than the others. For large values of t , the term $e^{-tn^2/2}$ tends to be really small for larger n . So, the $n = 1$ term will dominate and we get the following approximation for large t .

$$u(t, x) \approx \frac{2}{\pi} e^{-t/2} \sin(y) \sin(x)$$

8.4.7. *example 3: general solution.* We already pointed out that we should look for solutions of the form

$$e^{-t\lambda_n^2/2} \phi_n(x)$$

where

$$\Delta\phi_n(x) = -\lambda_n^2\phi_n(x)$$

We can do the trick of multiplying by the m -th function $\phi_m(x)$ and integrating to get the coefficient C_m provided that the functions are *orthogonal* in the sense that

$$\int_a^b \phi_n(x)\phi_m(x) dx = 0$$

if $n \neq m$. We also need to have the complete list of functions, i.e., the only function which is orthogonal to all the $\phi_n(x)$ should be zero. In other words, we want a Hilbert space basis.

8.5. **Recurrence and transience.** The question is: Does Brownian motion make particle go off to ∞ ?

- (1) Set up the probabilistic equation
- (2) Convert to a differential equation by time reversal
- (3) Solve the differential equation
- (4) Reinterpret probabilistically

8.5.1. *set up.* We start at a point x which is off center between two circles (or spheres in dimensions ≥ 3)

$$\mathbf{x} \in B = \{\mathbf{x} \in \mathbb{R}^d \mid R_1 < \|\mathbf{x}\| < R_2\}$$

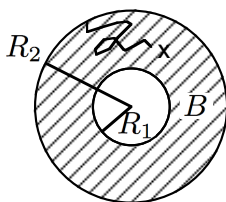


FIGURE 2. Will \mathbf{x} reach the outer circle before reaching the inner circle?

Take the stopping time T to be the smallest time so that $\mathbf{X}_T \in \partial B$ given that $\mathbf{X}_0 = \mathbf{x} \in B$. We now want to know: What is the probability that $\|\mathbf{X}_T\| = R_2$? The answer is

$$f(\mathbf{x}) = \mathbb{P}^{\mathbf{x}}(\|\mathbf{X}_T\| = R_2) = \mathbb{E}^{\mathbf{x}}(g(\mathbf{X}_T))$$

where $g(\mathbf{y})$ is given by

$$g(\mathbf{y}) = \begin{cases} 1 & \text{if } \|\mathbf{y}\| = R_2 \\ 0 & \text{if } \|\mathbf{y}\| = R_1 \end{cases}$$

8.5.2. *differential equation.* By the time reversal argument explained last time, $f(\mathbf{x})$ is the solution of the differential equation

$$\Delta f = 0$$

on B with boundary condition

$$f(\mathbf{y}) = \begin{cases} 1 & \text{if } \|\mathbf{y}\| = R_2 \\ 0 & \text{if } \|\mathbf{y}\| = R_1 \end{cases}$$

Since everything is rotationally symmetric, we know that the solution will be a function of $\|\mathbf{x}\|$. It is also a function of $z = \|\mathbf{x}\|^2 = \sum x_i^2$ which I much prefer since it has no nasty square roots.

$$f(\mathbf{x}) = \phi(z) = \phi\left(\sum x_i^2\right)$$

$$\frac{\partial f}{\partial x_i} = 2x_i\phi'(z)$$

$$\frac{\partial^2 f}{\partial x_i^2} = 2\phi'(z) + 4x_i^2\phi''(z)$$

Sum over all $i = 1, 2, \dots, d$ to get

$$\Delta f(\mathbf{x}) = 2d\phi'(z) + 4z\phi''(z) = 0$$

8.5.3. *solution of diffeq.* Put $\psi(z) = \phi'(z)$. Then the equation is

$$2dim\psi(z) + 4z\psi'(z) = 0$$

where I replaced the dimension d by “ dim ” temporarily so that I can write this as:

$$4z d\psi/dz = -2dim\psi$$

$$\frac{d\psi}{\psi} = -\frac{dim}{2} \frac{dz}{z}$$

Integrate both sides to get

$$\ln \psi = -\frac{dim}{2} \ln z + C_0$$

or:

$$\psi = \phi' = K_0 z^{-dim/2}$$

where $K_0 = e^{C_0}$. Integrate to get $f = \phi$:

$$f(\mathbf{x}) = \phi(z) = K_0 \frac{2z^{(2-d)/2}}{2-d} + C = K \|\mathbf{x}\|^{2-d} + C$$

if $d = dim \neq 2$ and $K = 2K_0/(2-d)$.

Now we put in the boundary conditions. First, $f(\mathbf{x}) = 0$ if $\|\mathbf{x}\| = R_1$. This gives

$$C = -KR_1^{2-d}$$

The other boundary condition is $f(\mathbf{x}) = 1$ when $\|\mathbf{x}\| = R_2$. This gives

$$1 = KR_2^{2-d} - KR_1^{2-d}$$

or

$$K = \frac{1}{R_2^{2-d} - R_1^{2-d}}$$

So, the solution (for $d \neq 2$) is

$$f(\mathbf{x}) = \frac{\|\mathbf{x}\|^{2-d} - R_1^{2-d}}{R_2^{2-d} - R_1^{2-d}}$$

If we put $d = 2$ we get $\frac{0}{0}$ and we can get the answer in the book by taking the limit as $d \rightarrow 2$ using l'Hospital's rule. (That's called "dimensional regularization." It isn't rigorous but it works.) The answer is:

$$f(\mathbf{x}) = \frac{\ln \|\mathbf{x}\| - \ln R_1}{\ln R_2 - \ln R_1}$$

8.5.4. *interpret solution.* Remember that $f(\mathbf{x})$ is the probability that $\|\mathbf{x}\|$ will reach R_2 before it reaches R_1 . So, we want to take the limit as $R_1 \rightarrow 0$ and $R_2 \rightarrow \infty$.

a) Take $R_1 \rightarrow 0$. When $d > 2$,

$$R_1^{2-d} = \frac{1}{R_1^{d-2}} \rightarrow \infty$$

So, $f(\mathbf{x}) \rightarrow 1$. Similarly, for $d = 2$,

$$\ln R_1 \rightarrow -\infty$$

So, $f(\mathbf{x}) \rightarrow 1$.

This means that, for $d \geq 2$, the probability is zero that the particle will ever return to the origin. When $d = 1$,

$$\lim_{R_1 \rightarrow 0} f(\mathbf{x}) = \|\mathbf{x}\|/R_2 < 1$$

The particle has a chance to go to the origin and therefore it eventually will with probability one. Then it will keep coming back because it can't avoid probability one events.

b) Take $R_2 \rightarrow \infty$ When $d > 2$,

$$R_2^{2-d} = \frac{1}{R_2^{d-2}} \rightarrow 0$$

So,

$$f(\mathbf{x}) \rightarrow \frac{R_1^{2-d} - \|\mathbf{x}\|^{2-d}}{R_1^{2-d}} = 1 - \left(\frac{R_1}{\|\mathbf{x}\|}\right)^{d-2} > 0$$

This means the particle has a chance to go to infinity. So, eventually it will with probability one. So, Brownian motion in dimensions > 2 is *transient*.

When $d = 2$,

$$\ln R_2 \rightarrow \infty$$

So,

$$f(\mathbf{x}) \rightarrow 0$$

The particle will never go to infinity. It will keep returning to the circle of radius R_1 about the origin no matter how small $R_1 > 0$ is. So, Brownian motion in \mathbb{R}^2 is (*neighborhood*) *recurrent*.

8.6. Fractal dimension of the path. The question is: What is the dimension of the path of Brownian motion?

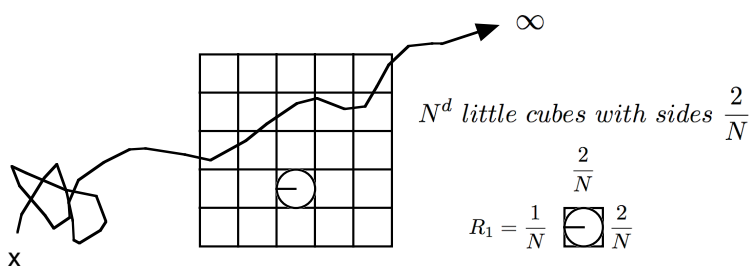


FIGURE 3. Count the number of little cubes needed to cover the path.

Take a $2 \times 2 \times \cdots \times 2$ cube in \mathbb{R}^d . Cut it into N^d little cubes with sides $2/N$ (so that it contains a disk of radius $1/N$). According to the definition, the box dimension of the path is given by counting the number of little squares needed to cover the path and looking at the exponent of N .

8.6.1. $d = 1$. In \mathbb{R}^1 , the path is equal to all of the line \mathbb{R}^1 . So, the dimension is 1.

8.6.2. $d = 2$. In \mathbb{R}^2 the path is dense. I.e., it gets arbitrarily close to every point. Therefore, we need all $N^d = N^2$ little cubes and the dimension is 2.

8.6.3. $d > 2$. For $d > 2$ we need to count. The expected number of little cubes that will be needed to cover the path is equal to the sum of the probability that the path will hit each little cube.

$$\mathbb{E}(\#\text{little cubes needed}) = \sum \mathbb{P}(\text{path hits one } (2/N)^d\text{-cube})$$

Since there are N^d little cubes this is approximately

$$N^d \mathbb{P}(\text{path hits one } (2/N)^d\text{-cube})$$

But we have the formula for the probability that a point will hit a sphere. So, I inserted the ratio between the volume of a cube and the volume of a ball and I got:

$$N^d \frac{2^d \Gamma\left(\frac{d+2}{2}\right)}{\pi^{d/2}} \mathbb{P}(\text{path hits ball of radius } \frac{1}{N})$$

We just calculated the probability of hitting the ball of radius $R_1 = 1/N$ before going off to infinity. (This was when we took the limit as $R_2 \rightarrow \infty$.) It was

$$\mathbb{P}(\text{hit ball of radius } \frac{1}{N}) = \left(\frac{R_1}{\|x\|}\right)^{d-2} = \frac{1}{\|x\|^{d-2}} \cdot N^{2-d}$$

So the number of cubes needed is a constant times $N^d N^{2-d} = N^2$. So the dimension of the path is 2.

8.7. Scaling and the Cauchy distribution. We skipped this because we talked about scaling at the beginning of the chapter and I don't think we need to know the Cauchy distribution.

8.8. Drift. *Brownian motion with (constant) drift* in \mathbb{R}^d is given by

$$\mathbf{Y}_t = \mathbf{X}_t + \mu t$$

where $\mu \in \mathbb{R}^d$ is a vector.

Suppose we are given the information \mathcal{F}_t up to time t . This is all contained in the single vector $\mathbf{Y}_t = \mathbf{x}$ in the sense that

$$\mathbb{E}^{\mathbf{x}}(-) = \mathbb{E}(- | \mathbf{Y}_t = \mathbf{x}) = \mathbb{E}(- | \mathcal{F}_t)$$

Then

$$\mathbf{Y}_{t+\delta t} = \mathbf{X}_{\delta t} + \mu \delta t + \mathbf{x}$$

where $\mathbf{X}_{\delta t}$ is a recentered standard Brownian motion.

Suppose that $f(\mathbf{x})$ is the particle density at \mathbf{x} at time t . (Density is actually $f(t, \mathbf{x})$ where the time t is just not written.)

Here I converted to $d = 1$. Then

$$f(Y_{t+\delta t}) = f(x) + f'(x)(X_{\delta t} + \mu \delta t) + \frac{1}{2} f''(x)(X_{\delta t} + \mu \delta t)^2 + o((X_{\delta t} + \mu \delta t)^2)$$

I denoted the change in f by

$$\delta f = f(Y_{t+\delta t}) - f(x)$$

So, the expected value of this is

$$\mathbb{E}^x(\delta f) = f'(x) \mathbb{E}^x(X_{\delta t} + \mu \delta t) + \frac{1}{2} f''(x) \mathbb{E}^x((X_{\delta t} + \mu \delta t)^2) + o$$

The first expected value is

$$\mathbb{E}^x(X_{\delta t} + \mu \delta t) = \mu \delta t$$

Next, I used the formula $\mathbb{E}(Z^2) = \mathbb{E}(Z)^2 + \text{Var}(Z)$ to get

$$\begin{aligned}\mathbb{E}^x((X_{\delta t} + \mu\delta t)^2) &= \mathbb{E}^x(X_{\delta t} + \mu\delta t)^2 + \text{Var}(X_{\delta t} + \mu\delta t) \\ &= (\mu\delta t)^2 + \text{Var}(X_{\delta t}) + \text{Var}(\mu\delta t) \\ &= \mu^2\delta t^2 + \delta t\end{aligned}$$

and I pointed out that the term $\mu^2\delta t^2$ is negligible since it is a $o(\delta t)$. This also means that

$$o((X_{\delta t} + \mu\delta t)^2) = o(\mu^2\delta t^2 + \delta t) = o(\delta t)$$

and

$$\mathbb{E}^x(\delta f(x)) = f'(x)\mu\delta t + \frac{1}{2}f''(x)\delta t + o(\delta t)$$

Dividing by δt and taking limit as $\delta t \rightarrow 0$ we get

$$\dot{f}(x) = \mu f'(x) + \frac{1}{2}f''(x)$$

where the dot is time derivative and the primes are space derivatives. This was for $d = 1$. In higher dimensions we get

$$\frac{\partial f}{\partial t} = \sum_{i=1}^d \mu_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\Delta f$$