Stochastic Calculus, Courant Institute, Fall 2019
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2019/StochCalc.html
Jonathan Goodman, September, 2019

Class 1, Brownian motion

1 Introduction to the course

These are class notes for the Stochastic Calculus class of Fall, 2019. They contain the material from the lecture, and probably a little more that I didn't have time to say in class. I plan to post the notes shortly after each class so that it can represent what happened that day.

This course uses the term *stochastic calculus* in two senses. In one sense, *stochastic calculus* refers to a set of tricks for calculating things related to random processes. One such trick is using the recursive backward equation to calculate expected values. Most of the information we have about stochastic processes comes from calculations like these. A clever proof usually relies on a clever calculation. When these calculations are done by computer, they become algorithms for computing things about random processes.

In another sense, *stochastic calculus* refers to the *Ito calculus* and relationships between certain stochastic processes and partial differential equations. The basic operations of ordinary differential and integral calculus may not work when applied to *diffusion processes* because they are not differentiable. The chain rule for diffusion processes, which is called *Ito's lemma*, requires you to calculate to second order in Taylor series in order to get the first "Ito derivative". The Ito calculus not only is a way to study diffusion processes. It is also a way to create *stochastic differential equation* (*SDE*) models of physical and economic systems.

Ordinary differential equation (ODE) models may be derived by asking: "What happens in a small time increment dt?". Here, and throughout the course, we use informal physical reasoning. If the state of the system at time t is modeled by $x(t)$, the *increment* is $dx = x(t+dt) - x(t)$. An ODE model would be a formula for dx as a function of x. This can take the form $dx(t) = f(x(t))dt$. The model is derived (the function f is determined) by asking how x changes in a small increment of time. Models of this form apply when the relationship between dx and x is not random.

Stochastic differential equations model dynamics under uncertainty. Random paths are often written with capital letters with the time argument as a subscript, such as X_t. Even if X_t is known, the increment $dX_t = X_{t+dt} - X_t$ is random. The expected value is $a(X_t)dt$ and the variance is $\mu(X_t)dt$. Assuming $dt > 0$ (so we are talking about the future), this is

$$\mathrm{E}[\, dX_t \mid X_t] = a(X_t)dt \ , \quad \mathrm{var}(\, dX_t \mid X_t) = \mu(X_t)\, dt \ . \tag{1}$$

We will call $a(x)$ the *infinitesimal mean* and $\mu(x)$ the *infinitesimal variance*. An SDE model of a stochastic process is specified by giving the infinitesimal mean

and variance. It may be surprising that the (infinitesimal) mean and variance completely describes the SDE model, since you have to give more than the mean and variance to specify the probability distribution of a random variable. However, a diffusion process is uniquely specified by its infinitesimal mean and variance. Therefore $a(x)$ and $\mu(x)$ are all that is needed to model a stochastic process as a diffusion or to simulate the diffusion model.

A diffusion model (an SDE) may be found by asking, approximately, how the system might change in a small amount of time. This often involves "coarse graining" – describing a complex system by just one or a few numbers. For example, stock trading in financial markets involves an *order book* where offers to buy or sell shares of stock are recorded. The complex behavior of the order book is replaced by a single number, S_t, that represents the stock price at time t. Actual stock prices are discrete, with prices given (in the US) in dollars and cents but not fractions of cents. *Geometric Brownian motion* is a simple SDE model in which t and S_t are continuous variables. The assumption is that $\mathrm{E}[\, dS_t \mid S_t]$ is proportional to S_t and $\mathrm{var}[\, dS_t \mid S_t]$ is proportional to S_t^2.

The mathematics of diffusion processes is a branch of probability theory. In elementary probability, the random object is a random variable X or a collection of random variables $X_1, \ldots, X_n$. Probability densities $p(x)$ or $p(x_1, \ldots, x_n)$ are a complete description of what we know about the random variable or variables before they are measured. A collection of n random variables may be thought of as a single n dimensional random variable, in which case we write $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$. The dimension is n. Diffusions have infinite dimension in this sense. The random object is a whole *path*, which is a "random function" of t. This might mean giving values X_t for all times t in the range $0 \leq t \leq T$. The numbers X_t are the "values" or "positions" at specific times. The whole path may be written $X_{[0,T]}$. The path cannot be specified completely by giving finitely many numbers, except is some uselessly abstract sense. Therefore, the mathematics of diffusions involves more than calculations with probability densities. We will see, for example, that partial differential equations play a big role.

The class was originally created for the Mathematics in Finance program, but it was always meant to be generic and useful to others with different applications in mind. Many of the examples are not from finance. Still, the choice of topics was influenced by financial applications. In particular, there is less about steady states and correlation functions than a course aimed at physics or chemistry students would have.

The reasoning in this class isn't rigorous in the pure math sense, but it is serious. Someone with the right background in measure theory would be able to make many of the arguments rigorous if she or he were interested. My wish for the class is to add as much "value" to students as possible in 13 lectures. That means sacrificing proofs to make time for applications.

Computing is an essential part of present and future applied mathematics. Since this class is applied mathematics, it would be wrong to do it without computing. In fact, the computational methods – simulation and PDE solving, etc. – are core elements of modern stochastic calculus.

2 Introduction to Brownian motion

Note: much of the first class was spent talking about mechanics and policies. That's the reason there is not so much technical material this week.

Brownian motion (for mathematicians) is the diffusion process with $a(x) = 0$ and $\mu(x) = 1$. Brownian motion (for scientists) is the name of the phenomenon that small particles in water, when you look at them with a powerful enough microscope, seem to move in a random fashion. It is named after an English doctor named Brown, but the Wikipedia page suggests that it was first observed elsewhere (France?).

The importance of Brownian motion goes beyond this particular physical model.

- Many other physical and mathematical processes are well approximated by Brownian motion. Brownian motion is a good model for many processes in addition to a small particle wandering in a fluid.

- The properties of Brownian motion, and the mathematical tools used to understand it, apply to many other stochastic processes, particularly to *diffusion processes* studied in Stochastic Calculus.

- Brownian motion is used as a "source of noise" to generate any other diffusion. It is useful to express any other diffusion path (definitions to come) as a function of a Brownian motion path. The *Ito calculus* allows us to express the stochastic dynamics of a diffusion process X_t in terms of a Brownian motion process W_t by writing the *Ito differential equation*, also called *stochastic differential equation* or *SDE*, as

$$dX_t = a(X_t)dt + b(X_t)dW_t \ . \tag{2}$$

  In this informal expression (as explained later in the course), $a(x)$ is the infinitesimal mean (also called *drift*). The infinitesimal variance is $\mu(x) = b^2(x)$.

In the last point above, X_t represents a general diffusion process and W_t is a Brownian motion. The W is for *Norbert Wiener*, who, along with Kolmogorov, created the mathematical structure we use to describe Brownian motion. It is sometimes called a *Wiener process*. We write X_t for Brownian motion when it is being studied on its own and W_t when it is being used to study another diffusion. Therefore, this Lecture uses X_t and future Lectures will usually use W_t.

Einstein in 1905 gave a physical explanation of the motion Brown saw in his microscope. He suggested thinking of the "small" particle in the microscope as a big object constantly being pushed in a random way by (small) water molecules around it. His model is now called the *Ornstein Uhlenbeck* process, after physicists who studied Einstein's model more abstractly later. The mathematician's Brownian motion is a limit of the Ornstein Uhlenbeck process, as we will see.

2.1 Motivation: Brownian motion as accumulation of noise.

By "pure noise", we often mean a sequence of "completely random" numbers of the same type. Here, let $\xi_1, \ldots, \xi_k, \ldots$ be an i.i.d., (independent and identically distributed) sequence of random variables with mean zero and finite variance

$$\mathrm{var}(\xi_k) = \sigma_\xi^2 = \mathrm{E}\big[\, \xi_k^2 \big] \ .$$

The variance is equal to the expected square because the mean is zero. The mean being zero is part of the definition of "noise". If the mean were not zero, we would call the mean a "signal". It would be the deterministic (non-random) part of the ξ_k sequence.

If you add together all the noise up to time n, you get

$$S_n = \sum_{k=1}^{n} \xi_k \ . \tag{3}$$

Brownian motion (the mathematician's version, from now on) is a continuous time version of this. The S_n sequence has several properties that have analogues for continuous time Brownian motion. First, $S_0 = 0$. We will see that "standard" Brownian motion (to be explained) has $X_0 = 0$. At time zero, you have added together zero noise, none. The means are zero:

$$\mathrm{E}[\, S_n] = 0 \quad (\text{for } n \geq 0) \ , \quad \mathrm{E}[\, X_t] = 0 \quad (\text{for } t \geq 0) \ .$$

Neither the S_n process not the X_t process has any signal.

The variance of both process grown linearly in time. For the S_n process, this follows from the definition and the fact that the ξ_k are independent:

$$\mathrm{var}(S_n) = \sigma_\xi^2 \, n \ .$$

The variance is proportional to n (the discrete "time" variable) with σ_ξ^2 being the constant of proportionality. The variance property for Brownian motion is

$$\mathrm{var}(X_t) = \mathrm{E}\big[\, X_t^2 \big] = t \ . \tag{4}$$

The variance is the expected square because the mean is zero. The variance is proportional to t. The constant of proportionality is equal to one, for "standard" or "normalized" Brownian motion.

The S_n and X_t processes both have the *independent increments* property. Let $n_1 < n_2$ be two positive times. The *increment* of S_n from time n_1 to time n_2 is

$$Y_{12} = S_{n_2} - S_{n_1} \ .$$

If $n_3 > n_2$ is another time, then there is another increment $Y_{23} = S_{n_3} - S_{n_2}$. The independent increments property is that increments of S_n from non-overlapping intervals of time are independent. In particular the increments Y_{12} and Y_{23} are

independent. This follows from the definition (3). The increments Y_{12} and Y_{23} are the sums of the noise arriving from n_1 to n_2 and from n_2 to n_3 respectively

$$Y_{12} = \sum_{k=n_1+1}^{n_2} \xi_k \ , \quad Y_{23} = \sum_{k=n_2+1}^{n_3} \xi_k \ .$$

None of the random variables ξ_k in the Y_{12} sum appear in the Y_{23} sum. Since the ξ_k are independent, the "disjoint" sums are independent too. This independence applies to any number of increments, as long as the corresponding intervals of time have no overlap. The increments have mean zero. The variance of an increment is proportional to the length of time in the increment

$$\mathrm{var}(Y_{12}) = \sigma_\xi^2 \ (n_2 - n_1) \ .$$

Brownian motion has a similar independent increments property. If $Y_{12} = X_{t_2} - X_{t_1}$ and $Y_{23} = X_{t_3} - X_{t_2}$, then Y_{12} and Y_{23} are independent. The independent increments property holds for any number of increments as long as the corresponding time intervals are disjoint. The increments have variance equal to (proportional to with constant of proportionality equal to 1) the amount of time:

$$\mathrm{var}(X_{t_2} - X_{t_1}) = t_2 - t_1 \ . \tag{5}$$

Brownian motion has two properties that are different from the discrete noise sum S_n. One is that increments of Brownian motion are Gaussian. The other is that the path X_t is a continuous function of t. Brownian motion X_t is supposed to represent the sum of infinitely many (the limit as $\Delta t \to 0$ of a large number of) infinitely small (increasingly small as $\Delta \to 0$) pieces of noise. The sum is Gaussian because of the central limit theorem. As $\Delta t \to 0$ (as we will see in future classes), the distribution of the approximate Brownian motion converges to a Gaussian, because of the central limit theorem.

Let us look at the Brownian motion *path* as a function of t. The path is a continuous function of t, as we will argue in the next paragraph and demonstrate in a later class. But the path is "rough" in the sense that it is not differentiable. In fact, the "velocity" of X_t is (in some sense) infinite. In fact,

$$\lim_{\Delta t \to 0} \left| \frac{X_{t+\Delta t} - X_t}{\Delta t} \right| = \infty \ .$$

It is common to say "almost surely" for statements like this. This will be explained later in the course.

You can understand these facts from the increment variance formula (). Write the increment as

$$\Delta X = X_{t+\Delta t} - X_t \ .$$

Then $\mathrm{E}\big[\, \Delta X^2 \big] = \Delta t$ (the length of the time interval). This suggests that a typical increment has the size roughly

$$|\Delta X| \sim \sqrt{\Delta t} \ .$$

The symbol $\sim$ here just means that the sizes on both sides are comparable. If $\Delta t = .01$ then $|\Delta X|$ might be .2 or .05, but it is unlikely to be as large as .5 or as small as .01. Using this in a calculation suggests that

$$\frac{|\Delta X|}{\Delta t} \sim \frac{1}{\sqrt{\Delta t}} \to \infty \ , \quad \text{as } \Delta t \to 0 \ .$$

This is the infinite velocity thing.

Question: If Brownian motion moves infinitely fast all the time, why does it not go infinitely far? How can X_t be finite if $X_0 = 0$ and $|dX/dt| = \infty$ all the time? The answer is *cancellation*. Sometimes ΔX is a large positive number and sometimes it's a nearly as large negative number. This makes the algebraic sum of increments (with signs included) much smaller than the sum of the absolute values of the increments:

$$\sum_{j=1}^{m} \left(X_{t_{j+1}} - X_{t_j} \right) \ll \sum_{j=1}^{m} \left| X_{t_{j+1}} - X_{t_j} \right| \ .$$

The $\ll$ is an informal symbol meaning "is much less than". There are approximately the same number of positive and negative terms on the left, so they "cancel" (approximately) in the sum.

Now ask about large t. Since X_t is Gaussian with mean zero and variance t, a calculation with the Gaussian PDF (probability density function) shows that

$$\mathrm{E}[\,|X_t|] = C\sqrt{t} \ .$$

The C involves π. The exact calculation is in assignment 1. If Brownian motion were to move at a uniform speed, then X_t would be proportional to t. Instead, it's proportional to something much smaller, $\sqrt{t}$. The reason, again, is cancellation.

2.2 Rigor, or not

This class is mathematically careful but not rigorous. Much of stochastic calculus involves understanding limits, particularly the limit as $\Delta t \to 0$. Derivatives and integrals in ordinary calculus are limits like this. We will not prove facts about limits, or even give the mathematical definition of a limit, the famous ϵ, δ definition from analysis (which you don't need to look up if you don't know). Suppose we have $Q_{\Delta t}$ and we want to show that $Q_{\Delta t} \to q$ as $\Delta t \to 0$. We will do this using inequalities, such as

$$|Q_{\Delta t} - q| \leq C\sqrt{\Delta t} \ . \tag{6}$$

Typically, a calculation will show that "there is a C so that .... If there is a C, even if we don't know what it is, this shows that $Q_{\Delta t} \to q$. For this class, the hard part of a "proof" is the calculation that leads to an inequality like (6).

An inequality like (6) if often called an "estimate". This use of the word "estimate" is different from how the word is normally used in English. Normally,

if a number is 4.72 and you estimate 4.71 or 4.73, that would be a good estimate. Here, on the other hand, it doesn't matter whether $C\sqrt{\Delta t}$ is close to the number $|Q_{\Delta t} - q|$. It only matters that it is larger. In this sense, 472 is an estimate of 4.72 (because it's larger). It would be better to call it "upper bound", which people used to do.

Finding a useful "estimate" (inequality) can take some cleverness. The challenge is to find a simple expression like $C\sqrt{\Delta t}$ that is larger than the complicated expression involving $Q_{\Delta t}$. A new theorem in stochastic calculus usually relies on a new clever inequality like this. Inequalities like these may be the hardest part of the class.

We gave properties of Brownian motion, but we did not show that Brownian motion exists. The discrete process S_n obviously exists, because we gave a formula for it. The properties (independent increments, variance) are consequences of the definition. The continuous time X_t is harder to define. In fact, it was around 20 years from the time Einstein made Brownian motion popular to the first proofs (independent and quite different) from Kolmogorov and Wiener that Brownian motion exists. In the approach of Wiener, you make an approximation $X_t^{\Delta t}$ and show that the limit exists as $\Delta t \to 0$. The Kolmogorov approach works directly with the "measure space" for Brownian motion. You can learn all about this in our courses Probability Limit Theorems I and II.