Stochastic Calculus, Courant Institute, Fall 2020
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2020/index.html

Week 1

Jonathan Goodman, September, 2020

1 About this Stochastic Calculus course

This short course is about building and understanding models of random processes. The models are *stochastic differential equations* (usually called *SDEs*), which are a random process analogue of ordinary differential equations (ODEs). A function of time will be called a *process*. It is either *deterministic* or *random*, so there are deterministic processes and random processes, the latter also called *stochastic processes*. Ordinary calculus (derivatives and integrals) is the mathematical machinery used to create and understand ordinary differential equation models of deterministic processes. Stochastic calculus serves the same function for stochastic processes.

A differential equation model describes the small changes that happen in small amounts of time. The "system" is the thing being modeled. Let X_t be a variable that describes the system at time t. The tradition in stochastic calculus is to use capital letters for random quantities and to put the time as a subscript. In ordinary differential equations, the state at time t might be written as $x(t)$. In the informal language of scientists and engineers, let dt be a small increment of time. The corresponding increment of X is $dX = X_{t+dt} - X_t$. A deterministic process model gives dX as a function of X_t, which may be written as

$$dX_t = a(X_t)dt \ . \tag{1}$$

This is an ordinary differential equation. It is traditional to put the dt on the other side, as

$$\frac{dX_t}{dt} = a(X_t) \ .$$

A stochastic process model is a description of dX as a random variable. The distribution of dX depends on X_t and dt.

This course describes a class of stochastic processes called *diffusions* or *diffusion processes*. A diffusion process is determined by the mean and variance of dX. More precisely, you have to specify the conditional *infinitesimal mean*, or *drift*, and the *infinitesimal variance*, or *quadratic variation*. The increment in the ordinary differential equation is proportional to the time increment dt. This also holds for the infinitesimal mean and variance.

$$a(x)dt = \mathrm{E}[\, dX \mid X_t = x] \tag{2}$$

$$\mu(x)dt = \mathrm{var}[\, dX \mid X_t = x] \ . \tag{3}$$

This might seem surprising, since the probability distribution of a random variable is not determined by just the mean and variance. It turns out, for diffusion processes, that dX is close enough to being Gaussian that the mean and variance is enough. To summarize: a deterministic ordinary differential equation model is specified by the function $a(x)$. A stochastic diffusion process model is specified by the infinitesimal mean $a(x)$ and the infinitesimal variance $\mu(x)$.

The simplest way to enter the world of stochastic diffusion process models is to take $\mu(x) = 1$ and $a(x) = 0$. This process is *Brownian motion*. It models a "system" that is moved by a steady stream of infinitesimal "shocks" that have the same distribution. We require only that the mean is zero. The constant infinitesimal variance can be *normalized* to $\mu = 1$ by a *scaling*. Much of this class is about this example.

Geometric Brownian motion is a simple diffusion process in which dX is proportional to X. This may be appropriate for modeling a population where "birth" ($dX > 0$) and "death" are specified as percentages of the population. It is used to model the random changes in the price of a "stock" (financial asset). Here too, prices go up and down in percentages. The infinitesimal mean being proportional to x means

$$a(x) = rx \ . \tag{4}$$

The parameter r is the expected growth rate (for populations) or the expected rate of return (for prices). The standard deviation of dX also should be proportional to x. The standard deviation is the square root of the variance. We want the standard deviation to be proportional to x, so $\sqrt{\mu(x)} = \sigma x$, for some constant σ. You can take the square root of both sides of (3), and the result is

$$(\mathrm{var}[\, dX \mid X_t = x])^{\frac{1}{2}} = \sigma x \sqrt{dt} \ . \tag{5}$$

We use this in the form

$$\mathrm{var}[\, dX \mid X_t = x] = \sigma^2 x^2 dt \implies \mu(x) = \sigma^2 x^2 \ . \tag{6}$$

The parameter σ is called the *volatilty* in finance and economics.

The term *geometric* comes from the fact that the "log process" $Y_t = \log(X_t)$ is closely related to ordinary Brownian motion. After Week 2, you will be able to verify that it has constant infinitesimal variance $\mu = \sigma^2$ and constant infinitesimal mean

$$a = r - \frac{1}{2}\sigma^2 \ . \tag{7}$$

The $\frac{1}{2}\sigma^2$ is the "Ito term" that we will find in Week 2 using *Ito's lemma*. Let W_t be a Brownian motion as above with infinitesimal mean $a = 0$ and infinitesimal variance $\mu = 1$. Then Y_t may be written

$$Y_t = y_0 + \sigma W_t + (r - \frac{1}{2}\sigma^2)t \ . \tag{8}$$

You can check this, since $\mathrm{var}(dY) = \sigma^2 \mathrm{var}(dW) = \sigma^2 dt$, and $\mathrm{E}[dY] = \mathrm{E}[dW](r - \frac{1}{2}\sigma^2)dt$. A geometric series is the exponential of an arithmetic series. If we think

of W_t as "arithmetic" Brownian motion, then *geometric Brownian motion* is a fitting way to describe the process

$$X_t = X_0 e^{\sigma W_t + \left(r - \frac{1}{2}\sigma^2\right)t} \ . \tag{9}$$

This formula expresses a diffusion process X_t in terms of a Brownian motion. The classes of weeks 3 and 4 describe the *stochastic differential equation* approach to stochastic modeling that relates any diffusion to a corresponding Brownian motion.

Basic probability uses probability densities to find expected values of functions of a random object. Week 3 introduces methods for getting expected values related to diffusion processes that do not use probability densities. The expected values are *value functions*, which depend on parameters x and t. Various value functions satisfy partial differential equations (*PDEs*) in x and t, which are in the family of *backward equations*. The relationship between diffusions and and PDEs can be used in the other direction. Associating a diffusion process to a PDE allows you evaluate the PDE solution by simulating the diffusion.

Week 4 explains the *Stochastic differential equation* (*SDE*) approach to describing a diffusion model of a stochastic process. The SDE approach relates a diffusion process to Brownian motion. This is another reason why Brownian motion is central to stochastic calculus. There is an Ito's lemma for a general diffusion process. The SDE and Ito's lemma form a stochastic calculus version of the chain rule from ordinary calculus.

Engineers and finance professionals use diffusion models to find and optimize control and decision strategies for stochastic processes. The Ito integral may be thought of in terms of dynamic trading strategies. Many stochastic control and decision problems have value functions that satisfy PDEs. These are distinguished from "passive" value functions by being (usually) nonlinear. Week 5 discusses these issues in the context of trading and hedging strategies on an asset whose price is a geometric Brownian motion. This includes the Black Scholes theory of option pricing and Merton's theory of optimal asset allocation.

Stochastic calculus needs a new way to specify the probability distribution of random paths. In basic probability, you learn that if a random object is specified by finitely many continuous components, you can specify its distribution using a probability density function. The distribution of a discrete object can be specified by giving the probabilities of every possible outcome. Neither of these approaches works for random paths because a path is not a discrete object, nor is is specified any finite collection of parameters. The number of parameters needed to describe an object is the *dimension*. If no finite set of parameters can do it, the object is called *infinite dimensional*. The random paths of stochastic calculus are infinite dimensional objects.

The theory of diffusion processes uses abstract *probability measures* to specify the distribution of random paths. Along with probability measures, there is an abstract form of integration that is related to the Lebesgue integral you might see at the end of a course on mathematical analysis. Expected values of functions of a random path are defined in this way. This is the infinite dimensional substitute

for integrating over parameters using a probability density to get the expected value of a function of a finite dimensional random object. This machinery is hard to use in specific applications, which is why the value functions of Week 3 are important. This course gives at best a very superficial description of probability measures.

The exception to this "no measures" rule comes in Week 6 with *Girsanov's change of measure theorem*. Two probability distributions on the same set of paths may be related by a *likelihood ratio*. The likelihood ratio is the ratio of probability densities when there are probability densities. But a likelihood ratio sometimes can relate probability distributions for two diffusion processes. *Girsanov's formula* is a formula for the likelihood ratio that relates two diffusion processes with the same infinitesimal variance but different infinitesimal mean. One example is the Black Scholes theory of option pricing, which has the "historical" process with a historical rate of return and volatility, and a *risk neutral* process with the same volatility but different rate of return. Another example is in the theory of fixed income assets where Girsanov's formula gives a change of *numeraire*.

2 Brownian motion paths, scaling

Brownian motion is a specific diffusion process that plays a special role in stochastic calculus. A Brownian motion path is a function of a time variable t written X_t. We assume the starting value $X_0 = 0$ and that X_t is defined for $t \geq 0$. In the language of the introduction, Brownian motion has infinitesimal mean $a = 0$ and infinitesimal variance $\mu = 1$. This section describes those properties a little more deeply. It also explains the *independent increments* property, which should be called the *i.i.d. increments* property, but isn't. *Scalings* describe how big some quantity is relative to another. If a quantity is random, you don't say exactly big it is, but roughly how big it typically is. The scaling in this section concerns how big increments of Brownian motion are over a specific amount of time.

Let us start with a reminder of conditional probability and independence. Conditional probability and conditional expectation are ways to express how probabilities change when you acquire new information. For example, suppose $Y = (Y_1, Y_2)$ is a two component random variable with probability density $u(y_1, y_2)$. The probability density of Y_2 alone is the marginal

$$v(y_2) = \int u(y_1, y_2)\, dy_1 \ . \tag{10}$$

This is the probability density of Y_2 if you have no information about Y_1. It is the *unconditional* density of Y_2. The conditional distribution of Y_2, if you know the value of Y_1 is

$$u(y_2|y_1) = \frac{u(y_1, y_2)}{\int u(y_1, y_2)\, dy_2} \ . \tag{11}$$

One obvious difference between the unconditional density (10) and the conditional density (11) is that the conditional density is a function of y_1. The conditional density or the conditional probability distribution is a function of the information you have. As long as Y_1 and Y_2 are related, the conditional distribution of Y_2 depends on Y_1. If the conditional density (11) does not depend on Y_1, then Y_1 and Y_2 are independent random variables. You can think of $Y = (Y_1, Y_2)$ as a random object. Knowing Y_1 is some information about this random object. Conditional probability and independence are about how this information changes or doesn't change the distribution of another part of the random object, Y_2.

We next describe some properties of Brownian motion without saying why they are true. You might say that they are the definition of Brownian motion. But this raises the question of whether these properties are possible. Is there a probability distribution for random paths with these properties? More importantly, do these properties describe something that might arise naturally as a model of a physical process? Section 3 gives answers to these questions. Brownian motion is a limit of *random walk*, which obviously exists and is a natural stochastic model. Brownian motion is a simple approximate way to describe many simple stochastic processes.

We write X to represent the whole path with values X_t. This is an infinite dimensional random object. The value of X at a specific time t_1 is partial information about X. Going further, let $X_{[t_1,t_2]}$ denote the piece of the path with $t \in [t_1, t_2]$, which is determined by all the values X_t for $t \in [t_1, t_2]$. This is more information, but still partial information about the whole path X. The *increment process* or *increment path* over $[t_1, t_2]$ will be written $Y_{[t_1,t_2]}$. Let s be a positive time increment with $s \leq t_2 - t_1$. The corresponding increment of X is

$$Y_{[t_1,t_2],s} = X_{t_1+s} - X_{t_1} \ . \tag{12}$$

We choose not to consider $Y_{[t_1,t_2],s}$ for $s < 0$ or $s > t_2 - t_1$, because then $Y_{[t_1,t_2]}$ would depend on X values outside the interval $[t_1, t_2]$, and this would mess up the upcoming definition of the independent increments property.

Brownian motion has the *independent increments property*, which is the property that increments over disjoint time intervals are independent. If $t_1 \leq t_2 \leq t_3 \leq t_4$, then $[t_1, t_2]$ and $[t_3, t_4]$ are disjoint time intervals. [This might not be exactly true if $t_2 = t_3$, but that case is also allowed in the independent increments property.] The property is that the increment paths $Y_{[t_1,t_2]}$ and $Y_{[t_3,t_4]}$ are independent. The distribution of the random increment $Y_{[t_3,t_4]}$, as a path, is independent of the increment $Y_{[t_1,t_2]}$. Remember that $Y_{[t_1,t_2]}$ and $Y_{[t_3,t_4]}$ are infinite dimensional random objects.

The independent property can be repeated to get more independent increments. For example, suppose we have four intervals with eight endpoints $[t_1, t_2]$, $[t_3, t_4]$, $[t_5.t_6]$, $[t_7, t_8]$, with $t_1 \leq t_2 \leq t_3 \leq t_4 \leq t_5 \leq t_6 \leq t_7 \leq t_8$. The first two corresponding increment processes are independent of each other, because the first two intervals are disjoint. Also, the first two increment processes are part of the increment process $Y_{[t_1,t_4]}$, which contains both of them. This makes them

independent of anything in $Y_{[t_5,4_8]}$, including the increment processes for $[t_5 t_6]$ and $[t_7, t_8]$. Exercise 2 shows that random objects can be pairwise independent without forming an independent family. Increment paths over any collection of disjoint intervals form an independent family.

For Brownian motion, increments over the same amount of time have the same probability distribution. If $t_4 - t_3 = t_2 - t_1$, then the random objects $Y_{[t_1,t_2]}$ and $Y_{[t_3,t_4]}$ have the same statistical properties. Increments are identically distributed as well as independent. This "same distribution" property is because the same the same random noise process is running the whole time This is explained in Section 3, just after the formula (22).

Let us look at a Brownian motion path from the point of view of integral calculus. Break a path into many small pieces and ask what happens when you put the pieces together. Choose a "final time" T and a number of increments n, then define the corresponding increment of time $\Delta t = T/n$. We keep T fixed, so Δt goes to zero as n goes to infinity. The n equal length time intervals have the form $[t_k, t_{k+1}]$, with $t_k = k\Delta t$, and $k = 0, 1, \ldots, n-1$. We write $\Delta X_k = X_{t_k+\Delta t} - X_{t_k}$ for the corresponding increments of X. The i.i.d. property implies that these ΔX_k are i.i.d. random variables.

The overall increment $X_T - X_0 = X_T$ is the sum of the small increments

$$X_T = \sum_{k=0}^{n-1} \Delta X_k \ . \tag{13}$$

The expected value of X_T is the sum of the expected values of ΔX_k. These are all the same (the *i.d.* of *i.i.d.*), so we drop the index k and just write

$$\mathrm{E}[\, X_T] = n\, \mathrm{E}[\, \Delta X] \ .$$

The variance of X_T is the sum of the variances (the first *i.* of *i.i.d.*), so we also have

$$\mathrm{var}(X_T) = n\, \mathrm{var}(\Delta X) \ .$$

We assumed that the infinitesimal mean of Brownian motion, which is the $a(x)$ in (2), is zero. When Δt is small, but not the "infinitely small" dt, we write the infinitesimal mean axiom as

$$\mathrm{E}[\, \Delta X] = o(\Delta t).$$

This means that

$$\frac{\mathrm{E}[\Delta X]}{\Delta t} \ \to \ 0 \ , \quad \text{as } \Delta t \to 0 \ .$$

But $\frac{1}{\Delta t} = \frac{n}{T}$, so we may do the following possibly silly calculation

$$\begin{aligned} \mathrm{E}[\, X_T] &= n\, \mathrm{E}[\, \Delta X] \\ &= \frac{T}{\Delta t} \mathrm{E}[\, \Delta X] \\ &= T\, \frac{\mathrm{E}[\, \Delta X]}{\Delta t} \\ &\to 0 \ \text{ as } \Delta t \to 0 \ . \end{aligned}$$

[This reasoning used the "little oh" symbol. This course will also use its relative, the "big Oh". You should look this up if it's unfamiliar.] The "silly" thing is that $\mathrm{E}[X_T]$ does not depend on n or Δt. Therefore, the only way $\mathrm{E}[X_T] \to 0$ as $n \to \infty$ is $\mathrm{E}[X_T] = 0$.

The variance calculation is similar. The infinitesimal variance formula (3) with $\mu(x) = 1$ for all x may be written in little oh notation as (with $\mu(x)\Delta t = \Delta t$)

$$\mathrm{var}(\Delta X) = \Delta t + o(\Delta t) \ .$$

Therefore, reasoning as before

$$\begin{aligned} \mathrm{var}(X_T) &= n \, \mathrm{var}(\Delta X) \\ &= \frac{T}{\Delta t} \mathrm{var}(\Delta X) \\ &= \frac{T}{\Delta t} \left(\Delta t + o(\Delta t) \right) \\ &= T + T \, \frac{o(\Delta t)}{\Delta t} \\ &\to T \ \text{ as } \Delta t \to 0 \ . \end{aligned}$$

As before, $\mathrm{var}(\, X_T)$ does not depend on Δt, so we learn that

$$\mathrm{E}[X_T] = 0 \tag{14}$$

$$\mathrm{var}(X_T) = T \ . \tag{15}$$

In the case of Brownian motion, the infinitesimal mean being zero implies that the mean itself is exactly zero. The infinitesimal variance $\mu = 1$ implies that the variance is equal to T.

We want to go beyond the mean and variance calculations to describe the probability distribution of X_T. But this requires another hypothesis. The reasoning that we used to derive (14) and (15) applies to the increments ΔX_k too. They all have the same distribution, so we drop the subscript and write ΔX for a random variable with the same distribution. The mean of ΔX is zero and the variance is equal to the time, which is Δt,

$$\left. \begin{aligned} \mathrm{E}[\Delta X] \ &= \ 0 \\ \mathrm{var}(\Delta X) \ &= \Delta t \ . \end{aligned} \right\} \tag{16}$$

But what does ΔX look like as a random variable? One simple possibility is that ΔX is a *scaled* version of a random variable, Z whose distribution does not depend on Δt. Suppose Z is a random variable with

$$\begin{aligned} \mathrm{E}[Z] &= 0 \\ \mathrm{var}(Z) &= 1 \ . \end{aligned}$$

The following scaling hypothesis is consistent with the ΔX moments (16):

$$\Delta X = \sqrt{\Delta t} \, Z \ . \tag{17}$$

I want to emphasize that this hypothesis is just a guess. It isn't required by the mean/variance relations (16) or the independent increments property. In fact, the *centered Poisson process* has independent and identically distributed increments and the mean variance formulas (16), but does not have a scaling property of the form (17). If we make the scaling hypothesis, then $\Delta X_k = \sqrt{\Delta t} Z_k$, with i.i.d. Z_k, so the sum formula (13) may be written

$$X_T = \sqrt{\Delta t} \sum_{k=0}^{n-1} Z_k \ .$$

The central limit theorem applies to this sum because the the distribution of the Z_k is independent of n or Δt. The central limit theorem implies that X_T is approximately Gaussian. This approximation gets better and better as $n \to \infty$. But (as we already said), the distribution of X_T does not depend on n or Δt. This implies that X_T is Gaussian, exactly.

Any information about X_T also applies to any increment of the process X over the time period T. The intervals $[t_k, t_{k+1}]$ are disjoint, so the ΔX_k are independent. We just saw that the first increment $\Delta X_0 = X_{\Delta t}$ is Gaussian with mean zero and variance Δt. Therefore, for all k,

$$\Delta X_k \sim \mathcal{N}(0, \Delta t) \ , \ \textit{i.i.d.} \tag{18}$$

The increments for other diffusion processes, not simple Brownian motion, are approximately Gaussian for small Δt, but they are not exactly Gaussian in general.

The scaling hypothesis (17) turned out to be *self consistent*, which means consistent with itself. We assumed only the scaling, then applied the central limit theorem to deduce that the increments are Gaussian with distribution (18). But mean zero Gaussians are all scalings of each other. In particular, if Z is a standard normal $Z \sim \mathcal{N}(0, 1)$, then ΔX has the same distribution as $\sqrt{\Delta t} Z$, which is (17). What we did is like a "guess and check" strategy from elementary math. We guessed something about the answer and then verified that we were right. Well, not quite, but since the guess was self consistent, at least we did not show that we were wrong.

3 Random walk and convergence in distribution

Brownian motion is a "coarse grained" approximation to random walk, and random walk is a "find grained" model that leads to Brownian motion. To define *random walk*, let U_k be an i.i.d. family of random variables with

$$\begin{aligned} \mathrm{E}[U] &= 0 \\ \mathrm{var}(U) &= \sigma^2 \ . \end{aligned}$$

The random walk process is

$$S_n = \sum_{k=0}^{n-1} U_k \ .$$

The coarse grained description is a simple but approximate description of S_n when n is large. The central limit theorem gives a simple approximate description of the distribution of S_n, it is approximately Gaussian. But the simple version of the central limit theorem does not describe S_n as a process.

The "walk" part of *random walk* comes from the case when $U = \pm 1$ with equal probability. The position of the walker after n steps is S_n. The walker starts at $S_0 = 0$. A step in the random walk is $S_{n+1} = S_n + U_n$. The step goes to the right if $S_{n+1} = S_n + 1$ and to the left otherwise. Left and right have equal probability $\Pr(U_n = 1) = \Pr(U_n = -1) = \frac{1}{2}$. A "coarse grained" description means a description of features you can "see from far away", which are the big (coarse) features. For example, a typical size of S_n is (see Exercise 3) is on the order of $\sqrt{n}$. The central limit theorem says that

$$\Pr(a\sqrt{n} < S_n < b\sqrt{n}) \approx \frac{1}{\sqrt{2\pi}} \int_{a/\sigma}^{b/\sigma} e^{-\frac{1}{2}z^2} \, dz \; . \tag{19}$$

This is a "coarse grained" statement because $a\sqrt{n}$ and $b\sqrt{n}$ are far apart. The Gaussian approximation can be wrong on a fine scale. For example, because S_n is always an integer,

$$\Pr\left(\frac{1}{4} \leq S_n \leq \frac{3}{4} \right) = 0 \; .$$

The endpoints $\frac{1}{4}$ and $\frac{3}{4}$ are too close together for a coarse grained approximation to apply.

How do you decide which questions are "fine grained" or "coarse grained"? One criterion is whether the question "disappears under scaling". This is related to mathematical statements of the central limit theorem that use "scalings". We will use the following scaling idea. Choose a small Δt and arbitrarily decide that one step in the random walk happens in every Δt interval of time. This means that step $S_n \to S_{n+1} = S_n + U_n$ happens at time $t_n = n\Delta t$. The value of the random walk at time T is S_n. This is likely to be a large number because $\mathrm{var}(S_n) = n\sigma$ and $n \to \infty$ as $\Delta t \to 0$. We use *Brownian motion scaling* to define a *rescaled* process $X^{\Delta t}$ that has the same variance as Brownian motion. This involves only multiplying the random walk by the appropriate *scaling factor*, as

$$X^{\Delta t}_{t_n} = \frac{\sqrt{\Delta t}}{\sigma} S_n \; .$$

[The notation $X^{\Delta t}_{t_n}$ is not great, but the other notations I tried were worse.] In this way, the random walk process S_n is scaled to the random path $X^{\Delta t}_{t_n}$.

This scaled path is only defined for times t_n that are multiples of Δt. The definition is extended to all $t \geq 0$ by "connecting the dots", which means linear interpolation. If t is not one of the times t_n, we choose n so that $t_n < t < t_{n+1}$. Then choose α so that $t = \alpha t_n + (1 - \alpha) t_{n+1}$. The interpolated path value is

$$X^{\Delta t}_t = \alpha X^{\Delta t}_{t_n} + (1 - \alpha) X^{\Delta t}_{t_{n+1}} \; .$$

If you set $\alpha = 0$ in this formula, both sides evaluate to $X_{t_n}^{\Delta t}$. If you set $\alpha = 1$, then both sides are equal to $X_{t_{n+1}}^{\Delta t}$. This is the continuous and piecewise linear function that has the given values at the times t_n.

If you fix any t and take the limit $\Delta t \to 0$, then the family of random variables $X_t^{\Delta t}$ has mean zero and variance that converges to t. The variance is equal to t if t is one of the times t_n. Otherwise it is within Δt of t. The distribution of $X_t^{\Delta t}$ converges to normal mean zero variance t, by the ordinary central limit theorem. This means that if $V(x)$ is a continuous and bounded function of x, then

$$\mathrm{E}\big[\, V(X_t^{\Delta t})\big] \;\to\; \mathrm{E}_{\mathcal{N}(0,t)}[\, V(Y)] \quad \text{as } \Delta t \to 0 \,. \tag{20}$$

The notation $\mathrm{E}_{**}[V(Y)]$ means the expectation, assuming that Y has the distribution $**$. In this case, that distribution is normal mean zero and variance t. This may be written more explicitly as

$$\mathrm{E}_{\mathcal{N}(0,t)}[\, V(Y)] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} V(y) e^{-\frac{y^2}{2t}} \,.$$

This limit defines what it means for $X_t^{\Delta t}$ to converge to $\mathcal{N}(0,t)$, *in distribution*. The central limit theorem is about convergence in distribution.

Convergence in distribution does not mean that the numbers $X_t^{\Delta t}$ converge to Y as $\Delta t \to 0$. The numbers $X_t^{\Delta t}$ probably don't converge at all as $\Delta t \to 0$. This is because $n \to \infty$ as $\Delta t \to 0$ for fixed t. When you go from Δt to $\frac{1}{2}\Delta t$, you use twice as many numbers U_k. That means that the new sum with $2n$ terms is likely to be much different from the old sum with n terms. But even if the numbers $X_t^{\Delta t}$ did have a limit as $\Delta t \to 0$, there would be no reason for this limit to be Y. In fact, it's a "dumb question" because there is no relation between $X_t^{\Delta t}$ and Y. Even if they had the same distribution, there would be nothing connecting them. Two random variables with the same distribution do not have to be equal or even related to each other. To make this point in a different way, the distribution limit formula (20) would be true if we had $-Y$ instead of Y on the right side. If $X_t^{\Delta t}$ converges to Y, then it does not converge to $-Y$.

The point of this section is that the distribution of the scaled path $X^{\Delta t}$ converges to the distribution of a Brownian motion path. This theorem is called *Donsker's invariance principle* and was first proved by Courant Institute mathematician Monroe Donsker. If $V(X)$ is a bounded continuous function of the path X, then

$$\mathrm{E}\big[\, V(X^{\Delta t})\big] \;\to\; \mathrm{E}_{\mathrm{BM}}[\, V(X)] \quad \text{as } \Delta t \to 0 \,. \tag{21}$$

This theorem includes the central limit theorem, because $V(X) =$ function of X_t is allowed. But there are many path functions that are functions of the path as a whole and not on X_t for just one t. These are called *path dependent* in finance. Here are some examples

Extreme values: The maximum over a fixed time range is

$$V(X) = M_T = \max_{0 \le t \le T} X_t \,.$$

This is path dependent because different paths achieve their maximum at different times. We will find a simple formula for PDF of M_T for Brownian motion. There is no exact formula for the maximum of a random walk, but it is approximately the same as for Brownian motion if n is large – after rescaling of course.

Integrals: Economists consider integrals of the form

$$\int_0^T e^{-\rho(T-t)} U(X_t)\, dt.$$

You can imagine many other path functions that involve integrals involving the path.

Brownian motion, summary

Relation to random walk

The increment process of a random walk would be defined as

$$S_{[n_1,n_2],m} = S_{n_1+m} - S_{n_1} \ .$$

This is like the increment of Brownian motion (12), except that the endpoints n_1 and n_2, and the increment m are integers. The increment process of random walk is determined by the steps in that interval. The increment is zero if $m = n_1$. For $m > n_1$ but $n \leq n_2$, the increment includes all the steps starting after n_1 and going up to m:

$$S_{[n_1,n_2],m} = \sum_{k=n_1+1}^{n_1+m} U_k \ . \tag{22}$$

Random walk increment processes over disjoint intervals are independent because they are determined by independent random steps. The steeps are i.i.d, so increment processes of the same length are identically distributed.

If you think of the steps U_k as "noise" or "shocks", then the increment random walk process $S_{[n_1,n_2],m}$ represents noise that arrive after step n_1. In the Brownian motion scaling limit, this noise arrives all the time rather than at specific times. The Brownian motion increment (12) represents the noise that arrives after time t_1. This noise is called *white noise* and is more abstract than makes sense for these six classes. Like the random steps U_k a white noise process in disjoint intervals is independent. Increment processes, which are sums or integrals of random steps or white noise, are independent for disjoint intervals of time. The variance of a random walk increment is proportional to the number of steps. The variance of a Brownian motion increment is proportional to the amount of white noise, which is proportional to the amount of time: $\text{var}(Y_{[t_1,t_2],s}) \sim s$. We choose Brownian motion scalings so that tbe proportionality constant is equal to 1. Therefore

$$\text{var}(X_{t_1+s} - X_{t_1}) = s \ .$$

This simple formula which holds for all positive numbers t_1 and s is the limit of approximate formulas for $X^{\Delta t}$ that only hold at the times t_k.

Brownian motion is Gaussian

We start with a review of the multi-variate normal distribution. This material should be familiar, though the notation may be different from what you've seen. Otherwise, you should stop and look up the multivariate distribution.

Let $Y = (Y_1, \ldots, Y_n)$ be a multivariate random (Gaussian) variable. The PDF of Y is Gaussian if it involves a "quadratic exponential". A quadratic function of $y \in \mathbb{R}^n$ has the form

$$\phi(y) = \frac{1}{2} y^t H y - \xi^t y + C \; . \tag{23}$$

The $\frac{1}{2}$ in the first term and "$-$" sign in the second term will simplify things later. In this formula, y and ξ are column vectors, and H is a symmetric $n \times n$ positive definite matrix, which is called the *information matrix*. This is the multivariate version of the simple quadratic function $ay^2 + by + c$, except that a is replaced by $\frac{1}{2}H$ and b is replaced by $-\xi^t$. You can "complete the square", as you can for simple quadratics. One way to do this is to calculate [This formula is true only if $H^t = H$.]

$$\frac{1}{2} (y - \mu)^t H (y - \mu) = \frac{1}{2} y^t H y - \mu^t H y + \frac{1}{2} \mu^t H \mu \; .$$

This matches the form of ϕ if $\xi = H\mu$, which is the same as $\mu = H^{-1}\xi$. The inverse matrix exists because H is positive definite. Therefore, ϕ may be expressed as

$$\phi(y) = \frac{1}{2} (y - \mu)^t H (y - \mu) + \widetilde{C} \; .$$

The new constant $\widetilde{C}$ depends on C and μ, but the exact value will be determined later. The Gaussian probability density is

$$u(y) = \frac{1}{\widetilde{Z}} e^{-\phi(y)} = \frac{1}{Z} e^{-\frac{1}{2}(y-\mu)^t H (y-\mu)} \; . \tag{24}$$

This is multi-variate normal with mean μ and covariance $C = H^{-1}$. The *normalization constant* Z is "just some number". The formula involves π, square roots, and $\det(H)$.

The Brownian motion path is Gaussian in a more abstract sense. The probability density formula (24) does not directly apply to "infinite dimensional" objects such as paths. But multivariate Gaussians have the property that any linear function of a multivariate Gaussian also is Gaussian. If X is Gaussian with n components and $Y = AX$ for some $d \times n$ matrix A, then Y is a $d-$component Gaussian. The mean and covariance matrix of Y can be calculated from the mean and covariance of X. A version of this applies when X is a Brownian motion path.

Although there is no Brownian motion probability density, there is a probability density for the values of the path at a finite set of times. Brownian motion starts at $t_0 = 0$. Consider an increasing sequence of times $t_0 < t_1 < \cdots < t_n$ and

consider the corresponding values $X_j = X_{t_j}$. [I hope this conflict of notation is not confusing.] We seek a formula for PDF $(X_1, \ldots, X_n) \sim u_n(x_1, \ldots, x_n)$. This should take the form

$$u_n(x_1 m \ldots, x_n) = \frac{1}{Z} e^{-\phi(x_1, \ldots, x_n)} \ ,$$

where ϕ is a quadratic function of $x_1, \ldots, x_n$. We will find a formula for ϕ, which will turn out to be quadratic.

We build the formula for ϕ in a sequence of steps that uses the fact that increments are independent Gaussians. To start, $X_1 \sim \mathcal{N}(0, t_1)$. The PDF of X_1 is

$$u_1(x_1) = \frac{1}{Z} e^{-\frac{x_1^2}{2t_1}} \ .$$

[We write Z for any normalization factor. Numbers Z in different formulas are not the same. This saves us from the distraction of writing formulas normalization constants that are not helpful.] Next, the increment $X_2 - X_1$ is independent of X_1 and has mean zero and variance $t_2 - t_1$. The conditional distribution of X_2, given X_1, is normal with mean X_1 and variance $t_2 - t_1$. The conditional density is

$$u_2(x_2|x_1) = \frac{1}{Z} e^{-\frac{(x_2 - x_1)^2}{t_2 - t_1}} \ .$$

The joint density is given by Bayes' rule as the product of the marginal of X_1 and the conditional of X_2 given X_1. That is

$$u_2(x_1, x_2) = u_2(x_2|x_1)\, u_1(x_1) = \frac{1}{Z} e^{-\frac{1}{2}\left[\frac{(x_2 - x_1)^2}{t_2 - t_1} + \frac{x_1^2}{t_1}\right]} \ . \tag{25}$$

As a reminder, the Z values in the last 3 formulas are all different. The density formula (25) is a multivariate Gaussian of the form (24) with

$$\phi(x_1, x_2) = \frac{1}{2}\left[\frac{(x_2 - x_1)^2}{t_2 - t_1} + \frac{x_1^2}{t_1}\right] \ .$$

This is a quadratic function of (x_1, x_2), but you would have to do some algebra to identify the entries of the precision matrix H. We see that $\mu = 0$, because $\nabla\phi(0) = 0$. We use the independent increments property to go further. The increment $X_3 - X_2$ is independent of $X_{[0,t_2]}$, and it is normal with mean zero and variance $t_3 - t_2$. Therefore, the conditional density of X_3 depends only on X_2 and is given by

$$u_3(x_3|x_1, x_2) = \frac{1}{Z} e^{-\frac{(x_3 - x_2)^2}{2(t_3 - t_2)}} \ .$$

The joint density of X_3 with (X_1, X_2) is given by the same Bayes' rule

$$u_3(x_1, x_2, x_3) = \frac{1}{Z} e^{-\frac{1}{2}\left[\frac{(x_3 - x_2)^2}{t_3 - t_2} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \frac{x_1^2}{t_1}\right]} \ .$$

The pattern of the general formula should now be clear. The formula is simpler if we define $x_0 = 0$.

$$u_n(x_1, \ldots, x_n) = \frac{1}{Z} e^{-\frac{1}{2}\sum_{j=0}^{n-1} \frac{(x_{j+1} - x_j)^2}{t_{j+1} - t_j}} \ . \tag{26}$$

Brownian motion is self similar

There is a sense in which Brownian motion is the same on every scale. Imagine watching the process X_t but with time running faster or slower than t by a factor of R. The variance of X_{Rt} is Rt. This is rescaled to have variance equal to t by by multiplying by $\frac{1}{\sqrt{R}}$. The rescaled process is

$$X'_t = \frac{1}{\sqrt{R}} X_{Rt} \; . \tag{27}$$

This has the $\mathrm{E}[\, X'] = 0$ and $\mathrm{var}(\, X') = t$. It turns out that the process X', as a random path, has the same probability distribution as X. This is the self-similarity property. If you rescale time (t) and space (X) by the appropriate powers of the scaling parameter R, the rescaled process is identical to the original process. Identical means identical in the statistical sense. The scaling formula (27) is a *similarity transformation*. Getting the same process makes the process self-similar.

To see what that means, imagine R being large. If $t \in [0, 1]$ then Rt ranges from 0 to R, which is a larger range. After rescaling space by the right factor, the behavior on scale R is the same as the behavior on scale 1.

4 Exercises

1. The *correlation coefficient* between random variables Y_1 and Y_2 is a measure of the relation between them

$$\rho_{12} = \frac{\mathrm{cov}(Y_1, Y_2)}{\sqrt{\mathrm{var}(Y_1)\,\mathrm{var}(Y_2)}} \; .$$

   Random variables are *uncorrelated* if $\rho_{12} = 0$. If $\rho_{12} \neq 0$ then Y_1 and Y_2 are not independent. Suppose $Z \sim \mathcal{N}(0, 1)$. This means that the distribution of Z is Gaussian, or *normal*, mean zero variance 1. Suppose Y_2 is any random variable $\mathrm{var}(Y_2) > 0$, and $Y_1 = ZY_2$. Show that Y_1 and Y_2 are uncorrelated but not independent.

2. Let Y_1, Y_2, and Y_3 be three binary random variables, which means that they take values $Y = 0$ and $Y = 1$ only. They are *pairwise independent* if Y_1 is independent of Y_2, Y_1 is independent of Y_3 and Y_2 is independent of Y_3. Give an example in which they are pairwise independent and "fair" ($\Pr(Y_j = 0) = \frac{1}{2}$ for each j), but not independent. If they are an independent family of random variables, then $\Pr(Y_1 = 0, Y_2 = 0, Y_3 = 0) = \frac{1}{8}$. Find a family that is pairwise independent (so $\Pr(Y_1 = 0, Y_2 = 0,) = \frac{1}{4}$, etc.) but not independent.

3. Suppose $X \sim \mathcal{N}(0, \sigma^2)$. Show that

$$\mathrm{E}[|X|] = \sqrt{\frac{2}{\pi}}\, \sigma \; .$$

Show that if S_n is a random walk with $\text{var}(U) = \sigma^2$, then

$$\text{E}[|S_n|] \approx \sqrt{\frac{2}{\pi}}\, \sigma n^{\frac{1}{2}} \ .$$

Verify the approximate formula (19).

4. *Interpolation* means finding a value for a function at value of its argument when values on either side are known. Applied to Brownian motion, this would mean finding a value for X_t if $t_1 < t < t_2$ and X_{t_1} and X_{t_2} are given. This is a random process, so the value of X_t is not determined by X_{t_1} and X_{t_2}. Instead, we seek the conditional probability distribution

$$X_t \sim u(\, x \mid X_{t_1} = x_1, X_{t_2} = x_2) \ .$$

This is a question about the increment $Y_{[t_1,t_2],t-t_1}$. Simplify notation by defining values of the increment at t_1, $t = 2_2$, and t

$$Y_1 = Y_{[t_1,t_2],t_1-t_1} \ , \ \ Y_2 = Y_{[t_1,t_2],t_2-t_1} \ , \ \ Y = Y_{[t_1,t_2],t-t_1} \ .$$

Of course, $Y_1 = 0$. The remaining values (Y, Y_2) are a two dimensional Gaussian. Write the PDF for the joint density $u(y, y_2) = * * *$, then use this to to find the conditional density of Y given Y_2. Unwind this to find the conditional distribution

$$X_t \sim \mathcal{N}(\mu_t, \sigma_t^2) \ , \ \ \text{conditional on } X_{t_1} = x_1,\, X_{t_2} = x_2 \ .$$

Show that μ_t is linear interpolation of the values x_1, x_2. Show that σ_t^2 goes to zero as $t \to t_1$ or $t \to t_2$ and explain why this is natural. Show that σ_t^2 is maximized at the midpoint $\frac{1}{2}(t_1 + t_2)$. If the value of X_{t_2} were not given, then $\sigma_t^2 = t - t_1$. Explain why σ_t^2 is smaller than this.

5. The *exponential* random variable is a positive random number with PDF $u(t) = \lambda e^{-\lambda t}$ if $t > 0$ and $u(t) = 0$ if $t < 0$. Call this distribution $\text{Exp}(\lambda)$. Download and run the code `ExponentialSampler.py`. This is Python 3. It uses Numpy version 18 or later. If you get an error message because of the random number generator, you may need to update your Numpy. I strongly advise using Python 3 in command line mode rather than in an IDE. That's what most serious developers do. You should get some output at the terminal, a plot should pop up, and it should write a plotfile `ExponentialHistogram.pdf`. You have to close the popup file manually each time you run the code.

   (a) Show that $\text{E}[T] = \frac{1}{\lambda}$.

   (b) Show that if U is a standard uniform ($U \in [0, 1]$ uniformly distributed), then $T = -\frac{1}{\lambda}\log(U)$ has $T \sim \text{Exp}(\lambda)$. The random number generator `rg.random()` returns a standard uniform. This explains line 32 in `ExponentialSampler.py`.

(c) Consider the conditional distribution of $T \sim \text{Exp}(\lambda)$ conditional on $T > t_s$. Show that this conditional distribution is $t_s + \text{Exp}(\lambda)$. This means that conditional on $T > T_s$, the extra time $T - t_s$ is also exponential with the same rate. The exponential distribution is used to model how long it takes a new lightbulb (or any device) to break. Explain that in the exponential model, a lightbulb that has not broken yet is as good as new.

(d) Explain why the procedure `sim(..)` produces T that with the conditional distribution of part (c).

(e) Let T_k be an independent sequence $T_k \sim \text{Exp}(\lambda)$. Let $N = \min\{k | T_k > t_s\}$. Calculate $\text{E}[N]$ as a function of λ and t_s. Show that the number produced by the code `ExponentialSampler.py` agrees with this.

6. Modify the code from Exercise 5 to study the hitting time related to exponential random variables. Define

$$R_n = \sum_{k=1}^{n} T_k \ , \ \ T_k \sim \text{Exp}(\lambda) \ \text{ i.i.d.}$$

From Exercise 5 part (a), we know $\text{E}[R_n] = \frac{n}{\lambda}$. Consider a positive "gap" g, and consider

$$N = \min\left\{ n \mid R_n > \frac{n}{\lambda} + g \right\} \ .$$

Consider the random variables $Y_k = T_k - \frac{1}{\lambda}$ and $S_n = R_n - \frac{n}{\lambda}$. let X_t be a standard Brownian motion with $X_0 = 0$ and $\text{var}(X_t) = t$. The first *hitting time* (also called *first passage time*) at $M > 0$ is

$$T = \min\{t \mid X_t \geq M\} \ .$$

We will see (Week 3) that T has PDF

$$u(t) = \frac{M}{\sqrt{2\pi t^3}} e^{-\frac{M^2}{2t}} \ .$$

Use this density and the Brownian motion scaling of the random walk S_n to estimate the distribution of N when g is large, with λ fixed. You may assume that the hitting time N of the random walk, properly scaled, is approximately related to the corresponding hitting time for Brownian motion.

Modify `ExponentialSampler.py` from Exercise 5 to test this theoretical prediction. Modify the function `sim` to simulate a first hitting time for the random walk. Make a histogram of the random times N. Make bins of width L, so that the *bin counts* are $B_j = \#\{k \mid jL \leq N_k < (j+1)L\}$. Make a plot showing the empirical bin counts and the theoretical prediction. Make a table of this information. The Brownian motion approximation is valid when g is large, because that forces a large number of steps

before $R_n \geq \frac{1}{\lambda} + g$. The histogram will be too noisy to be useful if L is too small. If the bins are too large, then there will not be enough bins to be an interesting test of the theory. Once your code is running, experiment with parameters to get as good a agreement with theory as you can.

Your code must follow style rules followed by `ExponentialSampler.py`. These include: all floating point number output must be formatted. Never use [str(x) if x is a floating point number in the code you upload. You are free to do that while debugging. Comment a lot. Make the comments useful. Use white space to make things line up vertically as much as possible to make the code easy to read, Use a docstring for any function. Put your name and contact information at the top, along with when and why you wrote the code. Put relevant numbers in the plot title and legends. Do not "hard wire" code parameters. Every code parameter should have a variable name and an assignment statement with a comment. Tabular output should be aligned under table headings.