Stochastic Calculus Notes, Lecture 5

Last modified October 21, 2004

1 Brownian Motion

1.1. Introduction: Brownian motion is the simplest of the stochastic processes called diffusion processes. It is helpful to see many of the properties of general diffusions appear explicitly in Brownian motion. In fact, the Ito calculus makes it possible to describea any other diffusion process may be described in terms of Brownian motion. Furthermore, Brownian motion arises as a limit or many discrete stochastic processes in much the same way that Gaussian random variables appear as a limit of other random variables throught the central limit theorem. Finally, the solutions to many other mathematical problems, particilarly various common partial differential equations, may be expressed in terms of Brownian motion. For all these reasons, Brownian motion is a central object to study.

1.2. History: Late in the 18^{th} century, an English botanist named Brown looked at pollen grains in water under a microscope. To his amazement, they were moving randomly. He had no explination for supposedly inert pollen grains, and later inorganic dust, seeming to swim as though alive. In 1905, Einstein proposed the explination that the observed "Brownian" motion was caused by individual water molecules hitting the pollen or dust particles. This allowed him to estimate, for the first time, the weight of a water molecule and won him the Nobel prize (relativity and quantum mechanics being too controversial at the time). This is the modern view, that the observed random motion of pollen grains is the result of a huge number of independent and random collisions with tiny water molecules.

1.3. Basics: The mathematical description of *Brownian motion* involves a random but continuous function on time, $X(t)$. The *standard* Brownian motion starts at $x = 0$ at time $t = 0$: $X(0) = 0$. The displacement, or *increment* between time $t_1 > 0$ and time $t_2 > t_1$, $Y = X(t_2) - X(t_1)$, is the sum of a large number of i.i.d. mean zero random variables, (each modeling the result of one water molecule collision). It is natural to suppose that the number of such collisions is proportional to the time increment. This implies, throught the central limit theorem, that Y should be a Gaussian random variable with variance proportional to $t_2 - t_1$. The *standard* Brownian motion has X normalized so that the variance is equal to $t_2 - t_1$. The random "shocks" (a term used in finance for any change, no matter how small) in disjoint time intervals should be independent. If $t_3 > t_2$ and $Y_2 = X(t_3) - X(t_2)$, $Y_1 = X(t_2) - Xt_1)$, then Y_2 and Y_1 should be independent, with variances $t_3 - t_2$ and $t_2 - t_1$ respectively. This makes the increments Y_2 and Y_1 a two dimensional multivariate normal.

1.4. Wiener measure: The probability space for standard Brownian motion is $C_0([0,T],R)$. As we said before, this consists of continuous functions, $X(t)$, defined for t in the range $0 \leq t \leq T$. The notation C_0 means[1] that $X(0)=0$. The $\sigma-$algebra representing full information is the Borel algebra. The infinite dimensional Gaussian probability measure on $C_0([0,T],R)$ that represents Brownian motion is called *Wiener measure*[2].

This measure is uniquely specified by requiring that for any times $0=t_0 < t_1 < \cdots < t_n \leq T$, the increments $Y_k = X(t_{k+1}) - X(t_k)$ are independent Gaussian random variables with $\text{var}(Y_k) = t_{k+1} - t_k$. The proof (which we omit) has two parts. First, it is shown that there indeed is such a measure. Second, it is shown that there is only one such. All the information we need is contained in the joint distribution of the increments. The fact that increments from disjoint time intervals are independent is the *independent increments* property. It also is possible to consider Brownian motion on an infinite time horizon with probability space $C_0([0,\infty),R)$.

1.5. Technical aside: There is a different descripton of the Borel $\sigma-$algebra on $C_0([0,T],R)$. Rather than using balls in the sup norm, one can use sets more closely related to the definition of Wiener measure through the joint distribution of increments. Choose times $0=t_0 < t_1 < \cdots t_n$, and for each t_k a Borel set, $I_k \subseteq R$ (thought of as "intervals" though they may not be). Let A be the event $\{X(t_k) \in I_k \text{ for all } k\}$. The set of such events forms an algebra (check this), though not a $\sigma-$algebra. The probabilities $P(A)$ are determined by the joint distributions of the increments. The Borel algebra on $C_0([0,T],R)$ is generated by this algebra (proof ommitted), so Wiener measure (if it exists) is determined by these probabilities.

1.6. Transition probabilities: The *transition probability density* for Brownian motion is the probability density for $X(t+s)$ given that $X(t)=y$. We denote this by $G(y,x,s)$, the "G" standing for *Green's function*. It is much like the Markov chain transition probabilities $P^t_{y,x}$ except that (i) G is a probability density as a function of x, not a probability, and (ii) tr is continuous, not discrete. In our case, the increment $X(t+s)-X(t)$, is Gaussina with variance s. If we learn that $X(t)=y$, then y becomes the expected value of $X(t+s)$. Therefore,

$$G(y,x,s) = \frac{1}{\sqrt{2\pi s}} e^{(x-y)^2/2s} \; . \tag{1}$$

1.7. Functionals: An element of $\Omega = C_0([0,T],R)$ is called X. We denote by $F(X)$ a real valued function of X. In this context, such a function is often called a *functional*, to keep from confusing it with $X(t)$, which

[1] In other contexts, people use C_0 to indicate functions with "compact support" (whatever that means) or functions that tend to zero as $t \to \infty$, but not here.

[2] The American mathematician and MIT professor Norbert Wiener was equally brilliant and inarticulate.

is a random function of t. This functional is just what we called a "function of a random variable" (the path X palying the role of the abstract random outcome ω). The simplest example of a functional is just a function of $X(T)$: $F(X) = V(X(T))$. More complicated functionals are integrals: $F(X) = \int_0^T V(X(t))dt$. extrema: $F(X) = \max_{t\leq T} X(t)$, or stopping times such as $F(X) = \min\left\{t \text{ such that } \int_0^t X(s)dx \leq 1\right\}$. Stochastic calculus provides tools for computing the expected values of many such functionals, often through solutions of partial differential equations. Computing expected values of functionals is our main way to understand the behavior of Brownian motion (or any other stochastic process).

1.8. Markov property: The independent increments property makes Brownian motion a Markov process. Let $\mathcal{F}_t$ be the $\sigma-$algebra generated by the path up to time t. This may be characterized as the $\sigma-$algebra generated by all the random variables $X(s)$ for $s \leq t$, which is the smallest $\sigma-$algebra in which all the functions $X(s)$ are measurable. It also may be characterized as the $\sigma-$algebra generated by events of the form A above ("Tehcnical aside") with $t_n \leq t$ (proof ommitted). We also have the $\sigma-$algebra $\mathcal{G}_t$ generated by the present only. That is, $\mathcal{G}_t$ is generated by the single random variable $X(t)$; it is the smallest $\sigma-$algebra in which $X(t)$ is measurable. Finally, we let $\mathcal{H}_t$ denote the $\sigma-$algebra that depends only on future values $X(s)$ for $s \geq t$. The Markov property states that if $F(X)$ is any functional measurable with respect to $\mathcal{H}_t$ (i.e. depending only on the future of t), then $E[F \mid \mathcal{F}_t] = E[F \mid \mathcal{G}_t]$.

Here is a quick sketch of the proof. If $F(X)$ is a function of finitely many values, $X(t_k)$, with $t_k \geq t$, then then $E[F \mid \mathcal{F}_t] = E[F \mid \mathcal{G}_t]$ follows from the independent increments property. It is possible (though tedious) to show that any F measurable with respect to $\mathcal{H}_t$ may be approximated by a functional depending on finitely many future times. This extends $E[F \mid \mathcal{F}_t] = E[F \mid \mathcal{G}_t]$ to all F measurable in $\mathcal{H}_t$.

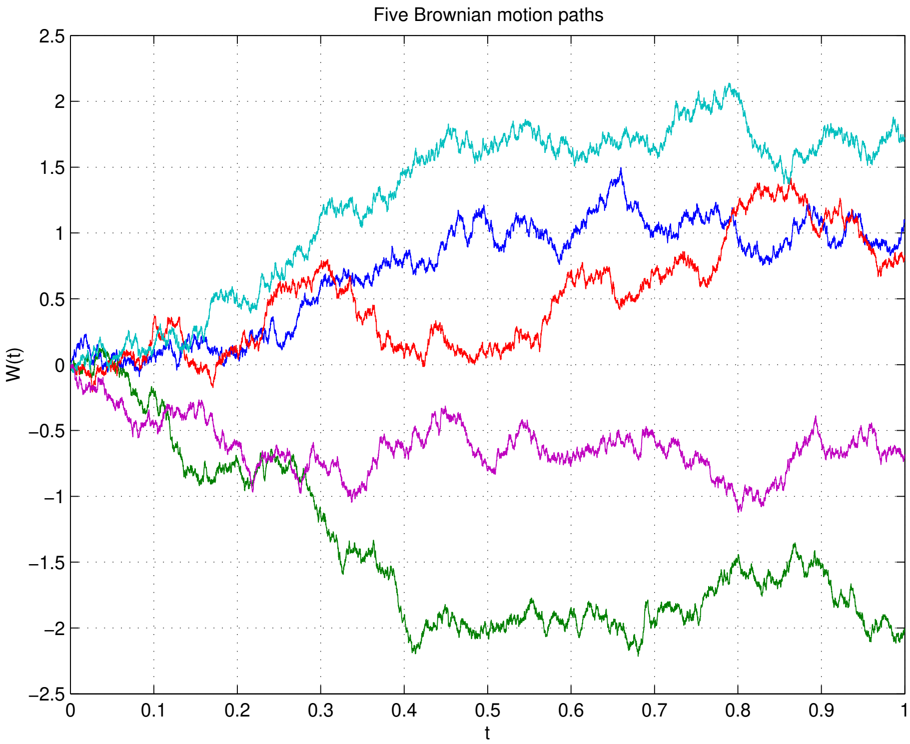
1.9. Path probabilities: For discrete Markov chains, as here, the individual outcomes are paths, X. For Markov chains one can compute the probability of an individual path by multiplying the transition probabilities. The situation is different Brownian motion, where each individual path has probability zero. We will make much use of the following partial substitute. Again choose times $t_0 = 0 < t_1 < \cdots < t_n \leq T$, let $\vec{t} = (t_1, \ldots, t_n)$ be the vector of these times, and let $\vec{X} = (X(t_1), \ldots, X(t_n))$ be the vector of the corresponding *observations* of X. We write $U^{(n)}(\vec{x}, \vec{t})$ for the joint probability density for the n observations, which is found by multiplying together the transition probability densities (1) (and using properties of exponentials):

$$U^{(n)}(\vec{x}, \vec{t}) \quad = \quad \prod_{k=0}^{n-1} G(x_k, x_{k+1}, t_{k+1} - t_k)$$

$$= \frac{1}{(2\pi)^{n/2}} \prod_{k=0}^{n-1} \frac{1}{\sqrt{t_{k+1}-t_k}} \exp\left(\frac{-1}{2} \sum_{k=0}^{n-1} \frac{(x_{k+1}-x_k)^2}{t_{k+1}-t_k} \right) . \quad (2)$$

The formula (2) is a concrete summary of the defining properties of the probability measure for Brownian motion, Wiener measure: the independent increments property, the Gaussian distribution of the increments, the variance being proportional to the time differences, and the increments having mean zero. It also makes clear that each finite collection of observations forms a multivariate normal. For any of the events A as in "Technical aside", we have

$$P(A) = \int_{x_1 \in I_1} \cdots \int_{x_n \in I_n} U^{(n)}(x_1, \ldots, x_n, \vec{t}) dx_1 \cdots dx_n .$$



1.10. Consistency: You cannot give just any old probability densities to replace the joint densities (2). They must satisfy a simple *consistency* condition. Having given the joint density for n observations, you also have given the joint density for a subset of these observations. For example, the joint density for $X(t_1)$ and $X(t_3)$ must be the marginal of the joint density of $X((t_1)$, $X(t_2)$, and $X(t_3)$:

$$U^{(2)}(x_1, x_3, t_1, t_3) = \int_{x_2=-\infty}^{\infty} U^{(3)}(x_1, x_2, x_3, t_1, t_2, t_3) dx_2 .$$

It is possible to verify these consistency conditions by direct calculation with the Gaussian integrals. A more abstract way is to understand the consistency conditions as adding random increments. The $U^{(2)}$ density says that we get $X(t_3)$ from $X(t_1)$ by adding an increment that is Gaussian with mean zero and variance $t_3 - t_1$. The $U^{(2)}$ says that we get $X(t_3)$ from $X(t_2)$ by adding a Gaussian with mean zero and variance $t_3 - t_2$. In turn, we get $X(t_2)$ from $X(t_1)$ by adding an increment having mean zero and variance $t_2 - t_1$. Since the smaller time increments are Gaussian and independent of each other, their sum is also Gaussian, with mean zero and variance $(t_3 - t_2) + (t_2 - t_1)$, which is the same as the variance in going from $X(t_1)$ to $X(t_3)$ directly.

1.11. Rough paths: The above picture shows 5 Brownian motion paths. They are random and differ in gross features (some go up, others go down), but the fine scale structure of the paths is the same. They are not smooth, or even differentiable functions of t. If $X(t)$ is a differentiable function of t, then for small Δt its increments are roughly proportional to Δt:

$$\Delta X = X(t + \Delta t) - X(t) \approx \frac{dX}{dt} \Delta t l.$$

For Brownian motion, the expected value of the *square* of ΔX (the variance of ΔX) is proportional to Δt. This suggests that typical values of ΔX will be on the order of $\sqrt{\Delta t}$. In fact, an easy calculation gives

$$E[|\Delta X|] = \frac{\sqrt{\Delta t}}{2\pi} \; .$$

This would be impossible if successive increments of Brownian motion were all in the same direction (see "Total variation" below). Instead, Brownian motion paths are constantly changing direction. They go nowhere (or not very far) fast.

1.12. Total variation: One quantitative sense of path roughness is the vact that Brownian motion paths have infinite total variation. The *total variation* of a function $X(t)$ measures the total distance it moves, counting both ups and downs. For a differentiable function, this would be

$$\text{TV}(X) = \int_0^T \left| \frac{dX}{dt} \right| dt l. \tag{3}$$

If $X(t)$ has simple jump discontinuities, we add the sizes of the jumps to (3). For general functions, the total variation is

$$\text{TV}(X) = \sup \sum_{k=0}^{n-1} |X(t_{k+1}) - X(t_k)| \; , \tag{4}$$

where the supremum as over all positive n and all sequences $t_0 = 0 < t_1 < \cdots < t_n \leq T$.

Suppose $X(t)$ has finitely many local maxima or minima, such as t_0 = local max, t_1 = local min, etc. Then taking these t values in (4) gives the exact total variation (further subdivision does not increase the left side). This is one way to relate the general definition (4) to the definition for differentiable functions (**??**). This does not help for Brownian motion paths, which have infinitely many local maxima and minima.

1.13. Almost surely: Let $A \in \mathcal{F}$ be a measurable event. We say A happens *almost surely* if $P(A) = 1$. This allows us to establish properties of random objects by doing calculations (stochastic calculus). For example, we will show that Brownian motions paths have infinite total variation almost surely by showing that for any (small) $\epsilon > 0$ and any (large) N,

$$P(\mathrm{TV}(X) < N) < \epsilon \ . \tag{5}$$

Let $B \subset C_0([0,t], R)$ be the set of paths with finite total variation.. This is a countable union

$$B = \bigcup_{N>0} \{\mathrm{TV}(X) < N\} = \bigcup_{N>0} B_N \ .$$

Since $P(B_N) < \epsilon)$ for any $\epsilon > 0$, we must have $P(B_N) = 0$. Countable additivity then implies that $P(B) = 0$, which means that $P(\mathrm{TV} = \infty) = 1$.

There is a distinction between outcomes that do not exist and events that never happen because they have probability zero. For example, if Z is a one dimensional Gaussian random variable, the outcome $Z = 0$ does exist, but the event $\{Z = 0\}$ is impossible (never will be observed). This is what we mean when we say "a Gaussian random variable never is zero", or "every Brownian motion path has invinite total variation".

1.14. The TV of BM: The heart of the matter is tha actual calculation behind the inequality (5). We choose an $n > 0$ and define (not for the last time) $\Delta t = T/n$ and $t_k = k\Delta t$. Let Y be the random variable

$$Y = \sum_{k=0}^{n-1} |X(t_{k+1}) - X(t_k)| \ .$$

Remember that Y is one of the candidates we must use in the supremem (4) that defines the total variation. If Y is large, then the total variation is at least as large. Because $E[|\Delta X|] = \sqrt{\frac{2}{\pi}}\sqrt{\Delta t}$, we have $E[Y] = \sqrt{\frac{2}{\pi}}\sqrt{T}\sqrt{n}$. A calculation using the independent increments property shows that

$$\mathrm{var}(Y) = \left(1 - \frac{2}{\pi}\right) T$$

for any n. Tchebychev's inequality[3] implies that

$$P\left(Y < \left(\sqrt{\frac{2}{\pi}}\sqrt{n} - k\sqrt{1-\frac{2}{\pi}}\right)\sqrt{T}\right) \leq \frac{1}{k^2} \; .$$

If we take very large n and medium large k, this inequality says that it is very unlikely for Y (or total variation of X) to be much less than $const\sqrt{n}$. Our inequality (5) follows from this whth a suitable choice of n and k.

1.15. Structure of BM paths: For any function $X(t)$, we can define the total variation on the interval $[t_1, t_2]$ in an obvious way. The odometer of a car records the distance travelled regardless of the direction. For $X(t)$, the total variation on the interval $[0, t]$ plays a similar role. Clearly, X is monotone on the interval $[t_1, t_2]$ if and only if $\mathrm{TV}(X, t_1, t_2) = |X(t_2) - X(t_1)|$. Otherwise, X has at least one local min or max within $[t_1, t_2]$. Now, Brownian motion paths have infinite total variation on any interval (the proof above implies this). Therefore, a Brownian motion path has a local max or min within any interval. This means that (like the rational numbers, for example) the set of local maxima and minima is *dense*: There is a local max or min arbitrarily close to any given number.

1.16. Dynamic trading: The infinite total variation of Brownian motion has a consequence for dynamic trading strategies. Some of the simplest dynamic trading strategies, Black-Scholes hedging, and Merton half stock/half cash trading, call for trades that are proportional to the change in the stock price. If the stock price is a diffusion process and there are transaction costs proportional to the size of the trade, then the total transaction costs will either be infinite (in the idealized continuous trading limit) or very large (if we trade as often as possible). It turns out that dynamic trading strategies that take trading costs into account can approach the idealized zero cost strategies when trading costs are small. Next term you will learn how this is done.

1.17. Quadratic variation: A more useful measure of roughness of Brownian motion paths and other diffusion processes is *quadratic variation.* Using previous notations: $\Delta t = T/n$, $t_k = k\Delta t$, the definition is[4] (where $n \to \infty$ as $\Delta t \to 0$ with $t = n\Delta t$ fixed)

$$Q(X) = \lim_{\Delta t \to 0} Q_n(X) = \lim_{\Delta t \to 0} \sum_{k=0}^{n-1} \left(X(t_{k+1} - X(t_k)\right)^2 \; . \tag{6}$$

[3] If $E[Y] = \mu$ and $\mathrm{var}(Y) = \sigma^2$, then $P(|Y - \mu| > k\sigma) < \frac{1}{k^2}$. The proof and more examples are in any good basic probability book.

[4] It is possible, though not customary, to define $\mathrm{TV}(X)$ using evenly spaced points. In the limit $\Delta t \to 0$, we would get the same answer for continuous paths or paths with $\mathrm{TV}(X) < \infty$. You don't have to use uniformly spaced times in the definition of $Q(X)$, but I think you get a different answer if you let the times depend on X as they might in the definition of total variation.

If X is a differentiable function of t, then its quadratic variation is zero (Q_n is the sum of n terms each of order $1/n^2$). For Brownian motion, $Q(T) = T$ (almost surely). Clearly $E[Q_n] = T$ for any n (independent increments, Gaussian increments with variance Δt). The independent increments property also lets us evaluate $\text{var}(Q_n) = 3T^2/n$ (the sum of n terms each equal to $3\Delta t^2 = 3T^2/n^2$). Thus, Q_n must be increasingly close to T as n gets larger[5]

1.18. Trading volatility: The quadratic variation of a stock price (or a similar quantity) is called it's "realized volatility". The fact that it is possible to buy and sell realized volatility says that the (geometric) Brownian motion model of stock price movement is not completely realistic. That model predicts that realized volatility is a constant, which is nothing to bet on.

1.19. Brownian bridge construction:

1.20. Continuous time stochastic process: The general abstract definition of a continuous time stochastic process is just a probability space, Ω, and, for each $t > 0$, a $\sigma-$algebra $\mathcal{F}_t$. These algebras should form a filtration (corresponding to increase of information): $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}$ if $t_1 \leq t_2$. There should also be a family of random variables $Y_t(\omega)$, with Y_t measurable in $\mathcal{F}_t$ (i.e. having a value known at time t). This explains why probabilists often write X_t instead of $X(t)$ for Brownian motion and other diffusion processes. For each t, we think of X_t as a function of ω with t simply being a parameter. Our choice of probability space $\Omega = C_0([0,T], R)$ implies that for each ω, $X_t(\omega)$ is a continuous function of t. (Actually, for simple Brownian motion, the path X plays the role of the abstract outcome ω, though we never write $X_t(X)$.) Other stochastic processes, such as the Poisson jump process, do not have continuous sample paths.

1.21. Continuous time martingales: A stochastic process F_t (with Ω and the $\mathcal{F}_t$) is a martingale if $E[F_s \mid \mathcal{F}_t] = F_t$ for $s > t$. Brownian motion forms the first example of a continuous time martingale. Another famous martingale related to Brownian motion is $F_t = X_t^2 - t$ (the reader should check this). As in discrete time, any random variable, Y, defines a continuous time martingale through conditional expectations: $Y_t = E[Y \mid \mathcal{F}_t]$. The Ito calculus is based on the idea that a stochastic integral with respect to X should produce a martingale.

2 Brownian motion and the heat equation

2.1. Introduction: Forward and backward equations are tools for calculating probabilities and expected values related to Brownian motion, as they are for Markov chains and stochastic processes more generally. The probability density of $X(t)$ satisfies a forward equation. The conditional expectations $E[V \mid \mathcal{F}_t]$

[5]Thes does not quite prove that (almost surely) $Q_n \to T$ as $n \to \infty$. We will come back to this point in later lectures.

satisfy backward equations for a variety of functionals V. For Brownian motion, the forward and backward equations are partial differential equations, either the *heat equation* or a close relative. We will see that the theory of partial differential equations of diffusion type (the heat equation being the a prime example) and the theory of diffusion processes (Brownian motion being a prime example) each draw from the other.

2.2. Forward equation for the probability density: If $X(t)$ is a standard Brownian motion with $X(0) = 0$, then $X(t) \sim \mathcal{N}(0,t)$, so its probability density is (see (1))

$$u(x,t) = G(0,x,t) = \frac{1}{\sqrt{2\pi t}} e^{x^2/2t} \; .$$

Directly calculating partial derivatives, we can verify that

$$\partial_t G = \frac{1}{2}\partial_x^2 G \; . \tag{7}$$

We also could consider a Brownian motion with a more general initial density $X(0) \sim u_0(x)$. Then $X(t)$ is the sum of independent random variables $X(0)$ and an $\mathcal{N}(0,t)$. Therefore, the probability density for $X(t)$ is

$$u(x,t) = \int_{y=-\infty}^{\infty} G(y,x,t) u_0(y) dy = \int_{y=-\infty}^{\infty} G(0,x-y,t) u_0(y) dy \; . \tag{8}$$

Again, direct calculation (differentiating (8), x and t derivatives land on G) shows that u satisfies

$$\partial_t u = \frac{1}{2}\partial_x^2 u \; . \tag{9}$$

This is the *heat equation*, also called *diffusion equation*. The equation is used in two ways. First, we can compute probabilities by solving the partial differential equation. Second, we can use known probability densities as solutions of the partial differential equation.

2.3. Heat equation via Taylor series: The above is not so much a derivation of the heat equation as a verification. We are told that $u(x,t)$ (the probability density of X_t) satisfies the heat equation and we verify that fact. Here is a method for deriving a forward equation without knowing it in advance. We assume that $u(x,t)$ is smooth enough as a function of x and t that we may expand it to to second order in Taylor series, do the expansion, then take the conditional expectation of the terms. Variations of this idea lead to the backward equations and to major parts of the Ito calculus.

Let us fix two times separated by a small Δt: $t' = t + \Delta t$. The rules of conditional probability allow us to compute the density of $X = X(t')$ in terms of the density of $Y = X(t)$ and the transition probabilit density (1):

$$u(x, t+\Delta t) = \int_{y=-\infty}^{\infty} G(y,x,\Delta t) u(y,t) dy \; . \tag{10}$$

The main idea is that for small Δt, $X(t + \Delta t)$ will be close to $X(t)$. This is expressed in G being small unless y is close to x, which is evident in (1). In the integral, x is a constant and y is the variable of integration. If we would approximate $u(y,t)$ by $u(x,t)$, the value of the integral just would be $u(x,t)$. This would give the true but not very useful approximation $u(x, t + \Delta t) \approx u(x,t)$ for small Δt. Adding the next Taylor series term (writing u_x for $\partial_x u$): $u(y,t) \approx u(x,t) + u_x(x,t)(y-x)$, the integral does not change the result because $\int G(y, x, \Delta t)(y - x)dy = 0$. Adding the next term:

$$u(y,t) \approx u(x,t) + u_x(x,t)(y-x) + \frac{1}{2}u_{xx}(x,t)(y-x)^2 \ ,$$

gives (because $E[(Y - X)^2] = \Delta t$)

$$u(x, t + \Delta t) \approx u(x,t) + \frac{1}{2}u_{xx}(x,t)\Delta t \ .$$

To derive a partial differential equation, we expand the left side as $u(x, t+\Delta t) = u(x,t) + u_t(x,t)\Delta t + O(\Delta t^2)$. On the right, we use

$$\int G(y, x, \Delta t) \left|y - x\right|^3 dy = O(\Delta t^{3/2}) \ .$$

Altogether, this gives

$$u(x,t) + u_t(x,t)\Delta t = u(x,t) + u_{xx}(x,t)\Delta t + O(\Delta t^{3/2}) \ \ .$$

If we cancel the common $u(x,t)$ then cancel the common factor Δt and let $\Delta t \to 0$, we get the desired heat equation (9).

2.4. The initial value problem: The heat equation (9) is the Brownian motion anologue of the forward equation for Markov chains. If we know the time 0 density $u(x, 0) = u_0(x)$ and the evolution equation (9), the values of $u(x,t)$ are completely and uniquely determined (ignoring mathematical technicalities that would be unlikely to trouble a practical person). The task of finding $u(x,t)$ for $t > 0$ from $u_0(x)$ and (9) is called the "initial value problem", with $u_0(x)$ being the "initial value" (or "values"??). This initial value problem is "well posed", which means that the solution, $u(x,t)$, exists and depends continuously on the initial data, u_0. If you want a proof that the solution exists, just use the integral formula for the solution (8). Given u_0, the integral (8) exists, satisfies the heat equation, and is a continuous function of u_0. The proof that u is unique is more technical, partly because it rests on more technical assumptions.

2.5. Ill posed problems: In some situations, the problem of finding a function u from a partial differential equation and other data may be "ill posed", useless for practical purposes. A problem is ill posed if it is not well posed. This means either that the solution does not exist, or that it does not depend continuously on the data, or that it is not unique. For example, if I try to find $u(x,t)$ for positive t knowing only $u_0(x)$ for $x > 0$, I must fail. A mathematician would say that the solution, while it exists, is not unique, there being many different ways to give $u_0(x)$ for $x > 0$, each leading to a different u. A more subtle situation arises, for example, if we give $u(x,T)$ for all x and wish to determine $u(x,t)$ for $0 \leq t < T$. For example, if $u(x,T) = \mathbf{1}_{[0,1]}(x)$, there is no solution (trust me). Even if there is a solution, for example given by (8), is does not depend continuously on the values of $u(x,T)$ for $T > t$ (trust me).

The heat equation (9) relates values of u at one time to values at another time. However, it is "well posed" only for determining u at future times from u at earlier times. This "forward equation" is well posed only for moving forward in time.

2.6. Conditional expectations: We saw already for Markov chains that certain conditional expected values can be calculated by working backwards in time with the backward equation. The Brownian motion version of this uses the conditional expectation

$$f(x,t) = E[V(X_T) \mid X_t = x] \ . \tag{11}$$

One "modern" formulation of this defines $F_t = E[V(X_t) \mid \mathcal{F}_t]$. The Markov property implies that F_t is measurable in $\mathcal{G}_t$, which makes it a function of X_t. We write this as $F_t = f(X_t, t)$. Of course, these definitions mean the same thing and yield the same f. The definition is also sometimes written as $f(x,t) = E_{x,t}[V(X_T)]$. In general if we have a parametrized family of probability measures, P_α, we write the expected value with respect to P_α as $E_\alpha[\cdot]$. Here, the probability measure $P_{x,t}$ is the Wiener measure describing Brownian motion paths that start from x at time t, which is defined by the densities of increments for times larger than t as before.

2.7. Backward equation by direct verification: Given that $X_t = x$, the conditional density for X_T is same transition density (1). The expectation (11) is given by the integral $f(x,t)$ as an integral, we get

$$f(x,t) = \int_{-\infty}^{\infty} G(x,y,T-t)V(y)dy \ . \tag{12}$$

We can verify by explicit differentiation (x and t derivatives act on G) that

$$\partial_t f + \frac{1}{2}\partial_x^2 f = 0 \ . \tag{13}$$

Note that the sign of ∂_t here is not what it was in (9), which is because we are calculating $\partial_t G(T-t)$ rather than $\partial_t G(t)$. This (13) is the *backward equation.*

2.8. Backward equation by Taylor series: As with the forward equation (9), we can find the backward equation by Taylor series expansions. We start by choosing a small Δt and expressing $f(x,t)$ in terms of[6] $f(\cdot, t+\Delta t)$. As before, define $F_t = E[V(X_T) \mid \mathcal{F}_t] = f(X_t, t)$. Since $\mathcal{F}_t \subset \mathcal{F}_{t+\Delta t}$, the tower property implies that $F_t = E[F_{t+\Delta t} \mid \mathcal{F}_t]$.

$$\begin{aligned} f(x,t) &= E_{x,t}[f(X_{t+\Delta t})] \\ &= \int_{y=-\infty}^{\infty} f(y, t+\Delta t) G(x,y,\Delta t) dy \; . \end{aligned} \tag{14}$$

As before, we expand $f(y, t+\Delta t)$ about x, t dropping terms that contribute less than $O(\Delta t)$:

$$\begin{aligned} & f(y, t+\Delta t) \\ & \quad = f(x,t) + f_x(x,t)(y-x) + \frac{1}{2} f_{xx}(x,t)(y-x)^2 + f_t(x,t)\Delta t \\ & \quad + O(|y-x|^3) + O(\Delta t^2) \; . \end{aligned}$$

Substituting this into (14) and integrating each term leads to

$$f(x,t) = f(x,t) + 0 + \frac{1}{2} f_{xx}(x,t)\Delta t + f_t(x,t)\Delta t + O(\Delta t^{3/2}) + O(\Delta t^2) \; .$$

A bit of algebra and $\Delta t \to 0$ then gives (13).

For future reference, we pause to note the differences between this derivation of (13) and the related derivation of (9). Here, we integrated G with respect to its second argument, while earlier we integrated with respect to the first argument. This does not matter for the special case of Brownian motion and the heat equation because $G(x,y,t) = G(y,x,t)$. When we apply this reasoning to other diffusion processes, $G(x,y,t)$ will be a probability density as a function of y for every x, but it need not be a probability density as a function of x for given y. This is an anologue of the fact in Markov chains that the transition matrix P acts from the left on column vectors f (summing P_{jk} over k) but from the right on row vectors u (summing P_{jk} over j). For each j, $\sum_k P_{jk} = 1$ but the column sums $\sum_j P_{jk}$ may not equal one. Of course, the sign of the ∂_t term is different in the two cases because we did the t Taylor series on the right side of (14) but on the left side of (10).

2.9. The final value problem: The *final values* $f(x,T) = V(x)$, together with the backward evolution equation (13) allow us to determine the values $f(\cdot, t)$ for $t < T$. The definition (11) makes this obvious. This means that the *final value problem* for the backward heat equation is a well posed problem.

On the other hand, the initial value problem for the backward heat equation is not a well posed problem. If we have a $f(x,0)$ and we want a $V(x)$ that leads to it, we are probably out of luck.

[6]The notation $f(\cdot, t+\Delta t)$ is to avoid writing $f(x, t+\Delta t)$ which might imply that the value $f(x,t)$ depends only on f at time $t+\Delta t$ for the same x value. Instead, it depends on all the values $f(y, t+\Delta t)$.

2.10. Duality: As for Markov chains, we can express the expected value of $V(X_T)$ in terms of the probability density at any earlier time $t \leq T$

$$E[V(X_T)] = \int u(x,t)f(x,t)dx \; .$$

This again implies that the right side is independent of t, which in turn allows us to derive the forward equation (9) from the backward equation (13) or conversely. For example, differentiating and using (13) gives

$$\begin{aligned} 0 &= \frac{d}{dt} \\ &= \int u_t(x,t)f(x,t)dx + \int u(x,t)f_t(x,t)dx \\ &= \int u_t(x,t)f(x,t)dx - \int u(x,t)\tfrac{1}{2}f_{xx}(x,t)dx \; . \end{aligned}$$

To derive an equation involving only u derivatives, we want to integrate the last integral by parts to move the x derivatives from f to u. In this formal derivation, we will assume that the probability density $u(x,t)$ decays to zero fast enough as $|x| \to \infty$ that we can neglect possible boundary terms at $x = \pm\infty$. This gives

$$\int \left(u_t(x,t) - \tfrac{1}{2}u_{xx}(x,t)\right) f(x,t)dx = 0 \; .$$

If this relation holds for a sufficiently rich family of functions f, we can only conclude that $u_t - \frac{1}{2}u_{xx}$ is identically zero, which is the forward equation (9).

2.11. The smoothing property, regularity: Solutions of the forward or backward heat equation become smooth functions of x and t even if the initial data (for the forward equation) or final data (for the backward equation) are not smooth. For u, this is clear from the integral formula (8). If we differentiate with respect to x, this derivative passes under the integral and onto the G factor. This applies also to x or t derivatives of any order, since the corresponding derivatives of G are still smooth integrable functions of x. The same can be said for f using (12); as long as $t < T$, any derivatives of f with respect to x and/or t are bounded. A function that has all partial derivatives of any order bounded is called "smooth". (Warning, this term is not used consistently. Some people say "smoooth" to mean, for example, merely having derivatives up to second order bounded.) Solutions of more general forward and backward equations often, but not always, have the smoothing property.

2.12. Rate of smoothing: Suppose the payout (and final value) function, $V(x)$, is a discontinuous function such as $V(x) = \mathbf{1}_{x<0}(x)$ (a "digital" option in finance). The solution to the backward equation can be expressed in terms of the *cumulative normal* (with $Z \sim \mathcal{N}(0,1)$)

$$N(x) = P(Z < x) = \frac{1}{\sqrt{2\pi}} \int_{z=-\infty}^{x} e^{-z^2/2} dz \; .$$

Then we have

$$\begin{aligned} f(x,t) &= \int_{y=-\infty}^{0} G(x,y,T-t)dy \\ &= \frac{1}{\sqrt{2\pi(T-t)}}\int_{y=-\infty}^{0} e^{-(x-y)^2/2(t-t)}dy \\ f(x,t) &= N(x/\sqrt{T-t})\ . \end{aligned} \tag{15}$$

From this it is clear that f is differentiable when $t < T$, but the first x derivative is as large as $1/\sqrt{T-t}$, the second as large as $1/(T-t)$, etc. All derivatives *blow up* as $t \to T$ with higher derivatives blowing up faster. This can make numerical solution of the backward equation difficult and inaccurate when the final data $V(x)$ is not smooth.

The formula (15) can be derived without integration. One way is to note that $f(x,t) = P(X_T < 0 \mid X_t = x)$ and $X_T \sim x + \sqrt{T-t}Z$, (Gaussian increments) so that $X_T < 0$ is the same as $Z < x/\sqrt{T-t}$. Even without the normal probability, a physicist would tell you that $\Delta X \sim \sqrt{\Delta t}$, so the hitting probability starting from x at time t has to be some function of $x/\sqrt{T-t}$.

2.13. Diffusion: If you put a drop of ink into a glass of still water, you will see the ink slowly diffuse through the water. This is modelled as a vast number of tiny ink particles each preforming an independent Brownian motion in the water. Let $u(x,t)$ represent the density of particles about x at time t (say, particles per cubic millemeter). This u satisfies the heat equation but not the requirement that $\int u(x,t)dx = 1$. If ink has been diffusing through water for some time, there will be dark regions with a high density of particles (large u) and lighter regions with smaller u. In the absence of boundaries (sides of the class and the top of the water), the ink distribution would be Gaussian.

2.14. Heat: Heat also can diffuse through a medium, as happens when we put a thick metal pan over a flame and wait for the other side to heat up. We can think of $u(x,t)$ as representing the temperature in a metal at location x at time t. This helps us interpret solutions of the heat equation (9) when u is not necessarily positive. In particular, it helps us imagine the *cancellation* that can occur when regions of positive and negative u are close to each other. Heat flows from the high temperature regions to low or negative temperature regions in a way that makes the temperature distribution a more uniform. A physical argument that heat (temperature) flowing through a metal should satisfy the heat equation was given by the French mathematical phycisist, friend of Napoleon, and founder of Ecole Polytechnique, Joseph Fourier.

2.15. Hitting times: A stopping time, τ, is any time that depends on the Brownian motion path X so that the event $\tau \leq t$ is measurable with respect to $\mathcal{F}_t$. This is the same as saying that for each t there is some process that has as input the values X_s for $0 \leq s \leq t$ and as output a decision $\tau \leq t$ or $\tau > t$. One kind of stopping time is a hitting time:

$$\tau_a = \min\left(t \mid X_t = a\right) \ .$$

More generally (particularly for Brownian motion in more than one dimension) if A is a closed set, we may consider $\tau_A = \min(t \mid X_t \in A)$. It is useful to define a Brownian motion that stops at time τ: $\tilde{X}_t = X_t$ if $t \leq \tau$, $\tilde{X}_t = X_\tau$ if $t \geq \tau$.

2.16. Probabilities for stopped Brownian motion: Suppose X_t is Brownian motion starting at $X_0 = 1$ and $\tilde{X}$ is the Brownian motion stopped at time τ_0, the first time $X_t = 0$. The probability measure, P_t, for $\tilde{X}_t$ may be written as the sum of two terms, $P_t = P_t^s + P_t^{ac}$. (Since $\tilde{X}_t$ is a single number, the probability space is $\Omega = R$, and the $\sigma-$algebra is the Borel algebra.) The "singular" part, P_t^s, corresponds to the paths that have been stopped. If $p(t)$ is the probability that $\tau \leq t$, then $P_t^s = p(t)\delta(x)$, which means that for any Borel set, $A \subseteq R$, $P_t^s(A) = p(t)$ if $0 \in A$ and $P_t^s(A) = 0$ if $0 \notin A$. This δ is called the "delta function" or "delta mass"; it puts weight one on the point zero and no weight anywhere else. Probabilists sometimes write δ_{x_0} for the measure that puts weight one on the point x_0. Phycisists write $\delta_{x_0}(x) =$ '$delta(x = x_0)$. The "absolutely continuous" part, P_t^{ac}, is given by a density, $u(x,t)$. This means that $P_t^{ac}(A) = \int_A u(x,t)dx$. Because $\int_R u(x,t)dx = 1 - p(t) < 1$, u, while being a density, is not a probability density.

This decomposition of a measure (P) as a sum of a singular part and absolutely continuous part is a special case of the Radon Nikodym theorem. We will see the same idea in other contexts later.

2.17. Forward equation for u: The density for the absolutely continuous part, $u(x,t)$, is the density for paths that have not touched $X = a$. In the diffusion interpretation, think of a tiny ink particle diffusing as before but being absorbed if it ever touches a. It is natural to expect that when $x \neq a$, the density satisfies the heat equation (9). u "knows about" the boundary condition because of the "boundary condition" $u(a,t) = 0$. This says that the density of particles approaches zero near the absorbing boundary. By the end of the course, we will have several ways to prove this. For now, think of a diffusing particle, a Brownian motion path, as being hyperactive; it moves so fast that it has already visited a neighborhood of its current location. In particluar, if X_t is close to a, then very likely $X_s = a$ for some $s < t$. Only a small minority of the particles at x near a, with small density $u(x,t) \to 0$ as $x \to a$ have not touched a.

2.18. Probability flux: Suppose a Brownian motion starts at a random point $X_0 > 0$ with probability density $u_0(x)$ and we take the absorbing boundary at $a = 0$. Clearly, $u(x,t) = 0$ for $x < 0$ because a particle cannot cross from positive to negative without crossing zero, the Brownian motion paths being continuous. The probability of not being absorbed before time t is given by

$$1 - p(t) = \int_{x>0} u(x,t)dx \ . \tag{16}$$

The rate of absorbtion of particles, the rate of decrease of probabiltiy, may be calculated by using the heat equation and the boundary condition. Differentiating (16) with respect to t and using the heat equation for the right side then integrating gives

$$\begin{aligned} -\dot{p}(t) &= \int_{x>0} \partial_t u(x,t)dx \\ &= \int_{x>0} \frac{1}{2}\partial_x^2 u(x,t)dx \\ \dot{p}(t) &= \frac{1}{2}\partial_x u(0,t) \; . \end{aligned} \tag{17}$$

Note that both sides of (17) are positive. The left side because $P(\tau \leq t)$ is an increasing function of t, the right side because $u(0,t) = 0$ and $u(x,t) > 0$ for $x > 0$. The identity (17) leads us to interpret the left side as the probability "flux" (or "density flux if we are thinking of diffusing particles). The rate at which probability flows (or particles flow) across a fixed point ($x = 0$) is proportional to the derivative (the gradient) at that point. In the heat flow interpretation this says that the rate of heat flow across a point is proportional to the temperature gradient. This natural idea is called Fick's law (or possibly "Fourier's law").

2.19. Images and Reflections: We want a function $u(x,t)$ that satisfies the heat equation when $x > 0$, the boundary condition $u(0,t) = 0$, and goes to δ_{x_0} as $t \downarrow 0$. The "method of images" is a trick for doing this. We think of δ_{x_0} as a unit "charge" (in the electrical, not financial sense) at x_0 and $g(x - x_0, t) = \frac{1}{\sqrt{2\pi}} e^{-(x-x_0)^2/2t}$ as the response to this charge, if there is no absorbing boundary. For example, think of puting a unit drop of ink at x_0 and watching it spread along the x axis in a "bell shaped" (i.e. gaussian) density distribution. Now think of adding a negative "image charge" at $-x_0$ so that $u_0(x) = \delta_{x_0} - \delta_{-x_0}$ and correspondingly

$$u(x,t) = \frac{1}{\sqrt{2\pi t}} \left(e^{-(x-x_0)^2/2t} - e^{-(x+x_0)^2/2t} \right) \; . \tag{18}$$

This function satisfies the heat equation everywhere, and in particular for $x > 0$. It also satisfies the boundary condition $u(0,t) = 0$. Also, it has the same initial data as g, as long as $x > 0$. Therefore, as long as $x > 0$, the u given by (18) represents the density of unabsorbed particles in a Brownian motion with absorption at $x = 0$. You might want to consider the image charge contribution in (18), $\frac{1}{\sqrt{2\pi}} e^{-(x-x_0)^2/2t}$, as "red ink" (the ink that represents negative quantities) that also diffuses along the x axis. To get the total density, we subtract the red ink density from the black ink density. For $x = 0$, the red and black densities are the same because the distance to the sources at $\pm x_0$ are the same. When $x > 0$ the black density is higher so we get a positive u. We can think of the image point, $-x_0$, as the reflection of the original source point through the barrier $x = 0$.

2.20. The reflection principle: The explicit formula (18) allows us to evaluate $p(t)$, the probability of touching $x = 0$ by time t starting at $X_0 = x_0$. This is

$$p(t) = 1 - \int_{x>0} u(x,t)dx = \int_{x>0} \frac{1}{\sqrt{2\pi t}} \left(e^{-(x-x_0)^2/2t} - e^{-(x+x_0)^2/2t} \right) dx \; .$$

Because $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-(x-x_0)/2t} dx = 1$, we may write

$$p(t) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi t}} e^{-(x-x_0)^2/2t} dx + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-(x+x_0)^2/2t} dx \; .$$

Of course, the two terms on the right are the same! Therefore

$$p(t) = 2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi t}} e^{-(x-x_0)^2/2t} dx \; .$$

This formula is a particular case the Kolmogorov reflection principle. It says that the probability that $X_s < 0$ for some $s \leq t$ is (the left side) is exactly twice the probability that $X_t < 0$ (the integral on the right). Clearly some of the particles that cross to the negative side at times $s < t$ will cross back, while others will not. This formula says that exactly half the particles that touch for some $s \leq t$ have $X_t > 0$. Kolmogorov gave a proof of this based on the Markov property and the symmetry of Brownian motion. Since $X_\tau = 0$ and the increments of X for $s > \tau$ are independent of the increments for $s < \tau$, and since the increments are symmetric Gaussian random variables, they have the same chance to be positive $X_t > 0$ as negative $X_t < 0$.