Stochastic Calculus Notes, Lecture 5

Last modified October 17, 2002

1 Brownian Motion

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, all the other diffusion processes 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.

1.1. Path space: I will call brownian motion paths $W(t)$ or W_t. In other places people might use B_t, b_t, $Z(t)$, Z_t, etc. The probability space ω will be the space of continuous functions of t for $t \geq 0$ so that $W_0 = 0$. Later, we might consider other starting positions, but that will be explicitly stated when we get there. We might consider finite time or infinite time. That is, we might consider functions W_t, for $0 \leq t \leq T$ or for all $t \geq 0$. The $sigma-$algebra will be the algebra generated by all the "coordinate" functions $X(W) = X_t$ for various t values. Since this is an infinite collection of functions, what we really mean is to consider first finite collections, $t_1 < \cdots < t_n$, and take the $\sigma-$ algebra generated by all these. This complex definition of $\mathcal{F}$ leads to lots of technicality in complete rigorous discussions of Brownian motion. Also important are the $\sigma-$algebras, $\mathcal{F}_t$ with information up to time t. These are generated by the coordinate functions for $t_1 < \cdots < t_n \leq t$.

1.2. Increment probabilities: The probability measure for Brownian motion, called Wiener measure, is specified by giving the probabilities of generating events. These generating events are events generated by finitely many coordinate functions. Let $t_0 < t_1 < \cdots < t_n$. The Brownian motion increments (sometimes called "shocks" by finance people) are $X_k = W_{t_{k+1}} - W_{t_k}$. W_t is a Brownian motion if the increments form a multivariate Gaussian, distinct increments are independent, $E[X_k] = 0$, and

$$\text{var}[X_k] = E[X_k^2] = E[(W_{t_{k+1}} - W_{t_k})^2] = t_{k+1} - t_k \; . \tag{1}$$

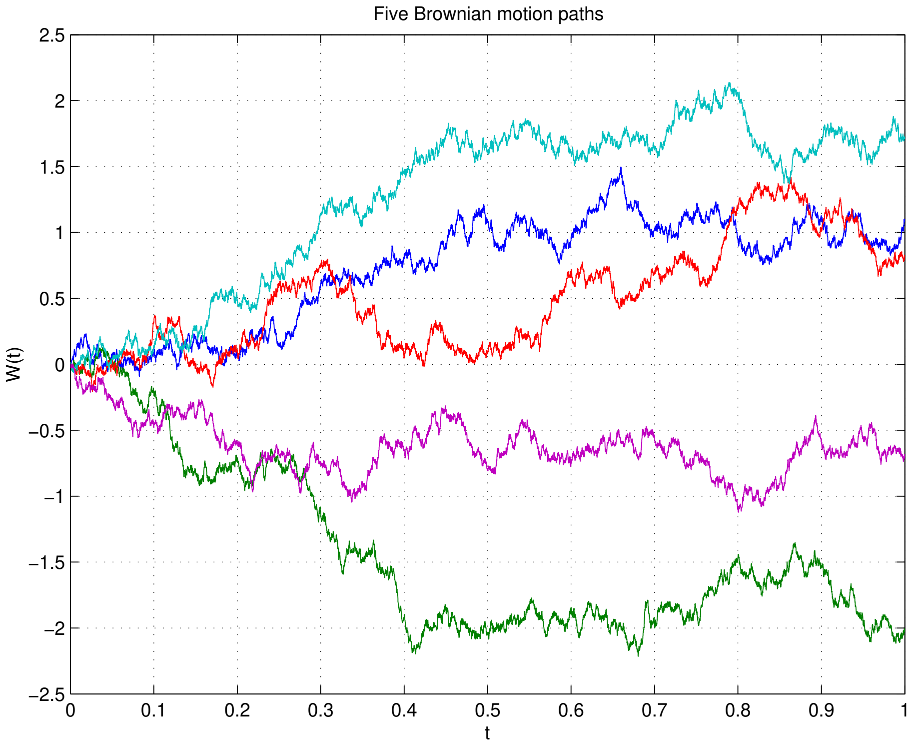
If (**??**) holds for every $n \geq 2$ and every set of times (increasing, of course), then the probability measure is Wiener measure.

1.3. Consistency: There is some technical mathematics between the claim of the above paragraph and it's proof. A first step might be to see that all the different probabilities for different n and t_k are consistent with each other.

There is something real here; if $\mathrm{var}(X_k) = (t_{k+1} - t_k)^2$, the probabilities are inconsistent (see below).

Suppose $m < n$ and we have two increasing sequences of times $t_1 < t_2 < \cdots < t_n$ and $\tilde{t}_n < \cdots \tilde{t}_m$. Suppose that the $\tilde{t}_k$ are a subset of the t_j. This means that the random variables $W_{\tilde{t}_k}$ are a subset of the random variables W_{t_j}. Call the joint probability density for the W_{t_j} $u(w_{t_1}, \ldots, w_{t_n})$, and let $\tilde{u}(w_{\tilde{t}_1}, \ldots, w_{\tilde{t}_m})$ be the density for the $W_{\tilde{t}_k}$. We should be able to get the probability density $\tilde{u}$ for the subset of the variables from the larger density u by integrating over the variables not present. That is, $\tilde{u}$ should be the marginal density for the $\tilde{t}_k$ derived from the the density u.

For example, suppose $n = 3$, $m = 2$, $t_1 < t_2 < t_3$ and $\tilde{t}_1 = t_1$ and $\tilde{t}_2 = t_3$. That is, the $\tilde{t}_k$ leave out the middle t. From (**??**), we find that the increments $X_1 = W_{t_2} - W_{t_1}$ and $X_2 = W_{t_3} - W_{t_2}$ are jointly gaussian with zero mean, correlation zero, and variance $\sigma_1^2 = t_2 - t_1$ and $\sigma_2^2 = t_3 - t_2$ respectively. For the $\tilde{t}_k$ we get that the increment $\tilde{X}_1 = W_{t_3} - W_{t_1}$ is gaussian with zero mean and variance $\tilde{\sigma}_1^2 = t_3 - t_1$. On the other hand, $\tilde{X}_1 = X_1 + X_2$ (check this from their definitions), so the distribution of the random variable $\tilde{X}_1$ is determined by those of X_1 and X_2. Are these two definitions of $\tilde{X}_1$ consistent? Yes. The sum of independent normals X_1 and X_2 is normal with variance $\sigma_1^2 + \sigma_2^2$. This shows that leaving out a single intermediate time gives consistent probability distributions. If we leave out times one at a time, we get the overall consistency statement. You should check that if the variance of X_k is not a linear function of $t_{k+1} - t_k$, the distributions are not consistent.



1.4. Rough paths, total variation: 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. For one thing, each of the paths, and any part of any path, has infinite "variation" (more technically, "total variation"). Consider times $T_1 < T_2$, choose a large number, n, and divide the time interval $[T_1, T_2]$ into $n-1$ equal side small subintervals t_k, t_{k+1}, where $t_k = T_1 + (k-1)\Delta t$, with $\Delta t = (T_2 - T_1)/(n-1)$. The quantity

$$V = \sum_{k=1}^{n-1} \left| W_{t_{k+1}} - W_{t_k} \right| \tag{2}$$

is the Δt variation of W between T_1 and T_2. By the independent increments property, the terms on the right side of (2) are independent. By (1), they have the same distribution. We estimate the sum of $n-1$ iid random variables using the Central Limit Theorem. The expected value is

$$E[V] = (n-1) \cdot E[|X_1|]$$

where $X_1 \sim \mathcal{N}(0, \Delta t)$. Therefore

$$E[|X_1|] \quad = \quad \frac{1}{2\pi\Delta t} \int_{x=-\infty}^{\infty} |x| \, e^{-x^2/(2\Delta t)} dx$$

$$
\begin{aligned}
&= 2 \cdot \frac{1}{2\pi\Delta t} \int_{x=0}^{\infty} x e^{-x^2/(2\Delta t)} dx \\
&= C\sqrt{\Delta t} \ ,
\end{aligned}
$$

where $C = \sqrt{2/\pi}$. Substituting the definition of Δt, this shows that $E[V] = const\sqrt{n-1}$, with $const = (2(T_2 - T_1)/\pi)^{1/2}$. As you take more and more intervals ($n \to \infty$), the total movement of W between T_1 and T_2 goes to infinity.

By contrast, suppose U_t is a differentiable function of time. Then

$$
|Ut_{k+1} - U_{t_k}| \approx \left| \frac{dU_t}{dt} \right| (t_{k+1} - t_k) \ ,
$$

so

$$
\sum_k |Ut_{k+1} - U_{t_k}| \to \int_{T_1}^{T_2} \left| \frac{dU}{dt} \right| dt < \infty \quad \text{as } n \to \infty.
$$

The variabion of a differentiable function has a limit, the "total variation" as the partition t_k gets finer. For Brownian motion, the finer you look, the more variation you see. Brownian motion paths are not differentiable in the ordinary sense of calculus. The Ito calculus is called for instead.

1.5. 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.

1.6. Quadratic variation: The quadratic variation for the partition t_k as above is

$$
Q(T_1, T_2, n) = \sum_{k=1}^{n-1} \left(W_{t_{k+1}} - W_{t_k} \right)^2 \ . \tag{3}
$$

This sum takes the squares of the increments, $X_k = W_{t_{k+1}} - W_{t_k}$ rather than the absolute values. For continuous paths, small $\Delta t = (T_2 - T_1)/(n-1)$ should imply small X_k. Therefore the quadratic variation should be smaller than the total variation. In fact, for a differentiable function, $Q \to 0$ as $n \to \infty$. For Brownian motion, the quadratic variation terms are just small enough for the sum not to go to zero or infinity as $n \to \infty$. In fact, the basic formula (1) implies that

$$
E[Q(T_2, T_1, n)] = \sum_k (t_{k+1} - t_k) = T_2 - T_1 \ , \tag{4}
$$

for any partition. Since the sum in (3) has a large number of iid terms for large n, the Central Limit Theorem suggests that the sum should be close to its expected value. Thus, we have the quadratic variation as the limit

$$Q(T_1, T_2) = \lim_{n \to \infty} Q(T_1, T_2, n) = T_2 - T_1 \ .$$

For other diffusion processes, the quadratic variation limit exists but it's value depends on the path. The quadratic variation is an important ingredient in the Ito calculus.

1.7. 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.

1.8. Almost sure convergence: An event, A, is called "almost sure" if $P(A) = 1$. For example, a probabilist would say that the quadratic variation formula (4) is true almost surely and might write

$$Q_n \to Q \quad \text{as } n \to \infty \quad a.s. \ .$$

It might seem that this should be called "sure" because we have no doubt that it will happen. The "almost" refers to the fact that (4) is might not be true for every $W \in \Omega$. There are paths, continuous functions W_t, so that the limit is infinite and others so that the limit is zero (e.g. differentiable paths). In continuous probability, there are many events that are impossible because they have probability zero, not because the do not exist.

1.9. Markov property: Brownian motion has the Markov property. This is a consequence of the independent increments property. For any t, we have the $\sigma-$ algebras $\mathcal{F}_t$ generated by the W_s for $0 < s \leq t$ (representing past and present), $\mathcal{G}_t$ generated by W_t (representing the present), and $\mathcal{H}_\sqcup$ (representing the future). The Markov property is that for any function $F \in \mathcal{H}_t$, $E[F \mid \mathcal{G}_\sqcup] = E[F \mid \mathcal{F}_\sqcup]$. A function measurable with respect to $\mathcal{H}_t$ depends on the values W_s for $s \geq t$. But W_s for $s \geq t$ is determined by W_t and increments X for intervals (t_k, t_{k+1}) that are measurable in $\mathcal{H}_t$ and independent of all increments that are $\mathcal{F}_t$ measurable. blabla.

1.10. Conditional probabilities for intermediate times:

1.11. Brownian bridge construction:

1.12. 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 be nested (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 W_t instead of $W(t)$. For each t, we think of W_t as a function of ω with t simply being a parameter. The Brownian motion has the property that, for every ω (not almost every), the map $t \to W_t(\omega)$ is a continuous function of t. Other stochastic processes, such as the Poisson jump process, do not have continuous sample paths.

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

2 Brownian motion and the heat equation

We saw for Markov chains that actual calculations of probabilities and expectation values often make use of forward and backward equations, which we call evolution equations, for probabilities (here, probability densities) and conditional probabilities. For Brownian motion, both the forward and backward equations are "the" heat equation, though the backward equation is often called the "backward heat equation". We will also find heat equations with boundary conditions that allow us to compute hitting time probability densities and expectations that involve hitting times.

2.1. Forward equation for the probability density: For now we will write X_t for Brownian motion. A Brownian motion starting at $X_0 = 0$ will have probability density at time t that is $\mathcal{N}(0,t)$. We denote this density by

$$g(x,t) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2t} \ . \tag{5}$$

Directly calculating partial derivatives, we can verify that

$$\partial_t g = \frac{1}{2}\partial_x^2 gl. \tag{6}$$

This g will play a role below as the "transition density" for Brownian motion, which is more general than just the density for X_t. For example, we could also consider a more general initial density $X_0 \sim u_0(x)$, and independent Gaussian increments as before. (We write $Y \sim v(y)$ to indicate that v is the probability density for the random variable Y, and sometimes also $Y_1 \sim Y_2$ to mean that Y_1 and Y_2 have the same density.) Then the increment $X_t - X_0$ will be $\mathcal{N}(0,t)$ and independent of X_0. That is, X_t is the sum of independent random variables X_0, with density u_0, and $X_t - X_0$, with density $g(\cdot, t)$. Therefore, the density for X_t is

$$u(x,t) = \int_{y=-\infty}^{\infty} g(x-y,t)u_0(y)dy \ , \tag{7}$$

Again, direct calculation using (5) shows taht u satisfies

$$\partial_t u = \frac{1}{2}\partial_x^2 u \ . \tag{8}$$

This is the "heat equation", also called "diffusion equation". The equation is used in two ways. First, we can compute probabilities by finding the solution to the partial differential equation. Also, we may be able to find solutions to the partial differential equation if there is an independent way to calculate the probability density.

2.2. Heat equation via Taylor series: There is another way to see that the X_t probability density u satisfies the heat equation (8) that proceeds directly from (5). This technique has the advantage that we do not have to know the equation in advance. We suppose only that u is a smooth function of x and t and derive the equation by Taylor series calculations. The idea applies in more general situations. It is one approach to the Ito calculus.

The Brownian motion $X_t \sim u(x,t)$ has the property that its increment in a small time interval Δt is $Y = X_{t+\Delta t} - X_t \sim \mathcal{N}(0, \Delta t)$, independent of X_t. As above, this means that $X_{t+\Delta t} = X_t + Y$ has probability density $u(x, t+\Delta t)$ that satisfies

$$u(x,t+\Delta t) = \int g(x-y,\Delta t)u(y,t)dy \ , \tag{9}$$

where g is still given by (5). Now, for small Δt, the integrand on the right side of (9) is significantly different from zero only when $x-y$ is small (not much larger than the order of $\sqrt{\Delta t}$). If u is a smooth function of x, most of the integral will be determined by values of u for y near x. This motivates us to approximate $u(y)$ as a Taylor series about x:

$$u(y) = u(x) + \partial_x u(x)\cdot(y-x) + \frac{1}{2}\partial_x^2 u(x)\cdot(y-x)^2 + O(|x-y|^3)l; .$$

We integrate the right side of (9) with this expansion, remembering that $\int g(x-y,\Delta t)(x-y)^2 dy = \Delta t$, that being the variance of the Δt increment in X. The result is (You can verify that $\int g(y,\Delta t)\left|y\right|^3 dy = O(\Delta t^{3/2}.)$:

$$\int g(x-y,\Delta t)u(y,t)dy = u(x,t) + 0 + \Delta t\frac{1}{2}\partial_x^2 u(x,t) + O(\Delta t^{3/2}) \ .$$

Of course, we also have

$$u(x,t_\Delta t) = u(x,t) + \Delta t\partial_t u(x,t) + O(\Delta t^2) \ .$$

Using these series for the left and right sides of (9) gives

$$u(x,t) + \Delta t\partial_t u(x,t) + O(\Delta t^2) = u(x,t) + \Delta t\frac{1}{2}\partial_x^2 u(x,t) + O(\Delta t^{3/2}) \ .$$

We cancel the $u(x,t)$ then divide by Δt and let $\Delta t \to 0$, and we are left with (8).

2.3. The initial value problem: The heat equation (8) is the Brownian motion anologue of the forward equation for Markov chains. It is often called the forward equation, often to distinguish it from the backward equation discussed below. If we know the time 0 density $u(x,0) = u_0(x)$ and the evolution equation (8), the values of $u(x,t)$ are completely and uniquely determined (ignoring mathematical technicalities that would be unlikely to trouble an applied person). The task of finding $u(x,t)$ for $t > 0$ from $u_0(x)$ and (8) 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 (7). Given u_0, the integral (7) 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).

2.4. 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 (7), is does not depend continuously on the values of $u(x,T)$ for $T > t$ (trust me).

The heat equation (8) 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.

2.5. 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

$$f(x,t) = E[V(X_T) \mid X_t = x] \,. \tag{10}$$

The "modern" formulation of this gives $f_t = E[V(X_t) \mid \mathcal{F}_t]$, which is, as has been repeated, a function of $X_t = x$ only. Of course, these definitions mean the same thing. The definition is also sometimes written as $f(x,t) = E_{x,t}[X_t]$. This is in the spirit of writing $E_\alpha[]$ for expectation with respect to the given probability measure P_α. Here, the probability measure $P_{x,t}$ is Brownian motion starting from x at time t, which is defined by the densities of increments for times larger than t as before.

2.6. Backward equation by direct verification: The expectation (10) depends on the increment $X_T - X_t$, which is $\mathcal{N}(0, T-t)$ and independent of X_t. Thus, the conditional density of X_T given that $X_t = x$ is (as a function of y) $g(y-x, T-t)$. Writing the expectation $f(x,t)$ as an integral, we get

$$f(x,t) = \int_{-\infty}^{\infty} g(x-y, T-t)V(y)dy \ . \tag{11}$$

Since this depends on x and t only through g, we can again verify through explicit calculation that

$$\partial_t f + \frac{1}{2}\partial_x^2 f = 0 \ . \tag{12}$$

Note that the sign of ∂_t here is not what it was in (8), which is because we are calculating $\partial_t g(T-t)$ rather than $\partial_t g(t)$. This (12) is the "backward equation".

2.7. Backward equation by Taylor series: As with the forward equation (8), we can find the backward equation by Taylor series expansions. Indeed, since $\mathcal{F}_t \subset \mathcal{F}_{t+\Delta t}$, we have, in "modern" notation,

$$f_t = E[V(X_T) \mid \mathcal{F}_t] = E[E[V(X_T) \mid \mathcal{F}_{t+\Delta t}] \mid \mathcal{F}_t] = E[f(X_{t_\Delta t} \mid \mathcal{F}_t] \ .$$

Using the probability density for the increment $X_{t+\Delta t} - X_t$, this gives the integral relation

$$f_t(x,t) = \int_{y=-\infty}^{\infty} g(x-y, \Delta t) f_{t+\Delta t}(y) dy \ . \tag{13}$$

Using Taylor series on the right and left (in different ways as above) again leads to (12).

2.8. The final value problem: We get a well posed problem by giving the partial differential equation (12) together with the "final values" $f(x,T) = V(x)$ (The definition (10) makes this obvious.). The "backwards heat equation enables us to find values of f at early times from given values at later times. The initial value problem, finding f_t with $t > 0$ from f_0 is not well posed. Although there may be occaisonal solutions, it is not a useful way to find the general solution, either because the general solution does not exist or because the solution that happens to exist does not depend in a continuous way on the values f_0.

2.9. Duality: You can check directly the duality property that $\int f(x,t)u(x,t)dx$ is independent of t. As for the Markov chain case, this is a consistency relation between the forward and backward evolution equations that makes one "dual" to the other. Also as for Markov chains, the integral is an expression of the law of total probability, integrating the expected payout starting at x at time t multiplied by the probability density for being at x at time t. This is $E[V(X_T)]$, and is thus independent of t.

2.10. 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 (7). 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 (13); 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 "smooooth" 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.

2.11. 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). For t close to T, $f(x,t)$ will be a differentiable function of x, but the derivative will be very large in some places. In fact,

$$\max_x |\partial_x f(x,t)| \sim \frac{1}{\sqrt{T-t}} \; .$$

Higher derivatives of f "explode" faster as t approaches T. If $V(x) = x_+$ (x_+ being the "positive part" of x, either x or 0 depending on which is larger), then the $\partial_x f$ is bounded as t approaches T, but the curvature "blows up". The fact that derivatives of f blow up at t approaches T makes numerical solution of the backward equation difficult and inaccurate.

2.12. Diffusion: It sometimes helps the intuition to think of particles diffusing through some medium, ink particles diffusing through still water, for example. Then $u(x,t)$ can represent the density of particles about x at time t. If ink has been diffusing through water for some time, there might be dark regions with a high density of particles (large u) and lighter regions with smaller u. This helps us interpret, for example, solutions of the heat equation (8) without the requirement that $\int u(x,t)dx = 1$. For ink in water, it is a reasonable approximation to think of each particle performing it's own Brownian motion independent of all the others. If the density of particles were too high (e.g. all particles and no water), we would have to adjust the model. A physical argument that tiny particles in water should undergo Brownian motion, and that their density should satisfy the heat equation, was given by the German phycisist Albert Einstein, and was the basis of his Nobel Prize (relativity and quantum mechanics seeming too uncertain at the time).

2.13. 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 (8) 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 to create a more uniform equilibrium temperature. 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.

2.14. 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 $t \geq \tau$.

2.15. 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.

2.16. 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 (8). 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.

2.17. 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

$$1 - p(t) = \int_{x>0} u(x,t)dx \; . \tag{14}$$

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 (14) with respect to t and using the heat equation for the right side then integrating gives

$$\begin{aligned} -\dot{p}(t) &= \int_{x>0} \partial_t u(x,t)dx \\ &= \int_{x>0} \frac{1}{2}\partial_x^2 u(x,t)dx \\ \dot{p}(t) &= \frac{1}{2}\partial_x u(x,0) \; . \end{aligned} \tag{15}$$

Note that both sides of (15) 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 (15) 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").

2.18. 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

$$u(x,t) = \frac{1}{\sqrt{2\pi t}} \left(e^{-(x-x_0)/2t} - e^{-(x+x_0)/2t} \right) \; . \tag{16}$$

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 (16) 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 (16), $\frac{1}{\sqrt{2\pi}} e^{-(x-x_0)^2/2t}$, 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 $x = 0$.

2.19. The reflection principle: The explicit formula (16) allows us to evaluate $p(t)$, the probability of touching $x = 0$ by time t starting at $X_0 = x_0$. This is

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

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

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

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

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

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 $s \leq t$ $x = 0$ 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 $X_t < 0$.