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

September 23: some typos fixed.

Class 2, Brownian motion and diffusion equations

1 Introduction

A *differential equation* is an equation satisfied by a function and some of its derivatives. A *partial* differential equation is a differential equation involving a function of more than one variable and partial derivatives of that function with respect to more than one variable. Partial differential equations, *PDEs*, are one of the main tools in stochastic calculus. Many quantities of interest in stochastic calculus satisfy PDEs. There are powerful techniques for solving PDEs and, if not solving them, at least learning about the solutions. Some of the PDEs have explicit solution formulas. We will see some examples today. Sometimes it is practical to compute the solution numerically. We will give an example of a numerical solution method, but there are more sophisticated methods that we cannot discuss in this course. Take the course Numerical Methods II for more on this.

There are analytical approximation methods for PDEs, particularly if a parameter in the stochastic process is small or large. For example, there are approximate methods to estimate what happens for small amounts of time.

2 Diffusion of probability density

Let X_t be a Brownian motion whose increments are Gaussian as explained last week

$$X_{t_2} - X_{t_1} \sim \mathcal{N}(0, t_2 - t_1) \; . \tag{1}$$

Here and throughout the course, the notation $\mathcal{N}(\mu, \sigma^2)$ refers to the Gaussian distribution with mean μ and variance σ^2. If Y is a random variable, then $Y \sim p$ means that Y has the probability distribution given by p. This p could be a PDF, as it is here, or it could be a different way to specify probability distribution, such as a probability measure (described in later classes). Recall that increments from disjoint time intervals are independent.

The probability density of X_t satisfies a PDE called "the" *heat equation* or "the" *diffusion equation*. I put "the" in quotes because there are more complicated heat equations and diffusion equations. Often "the" heat equation refers to the simplest heat equation, which happens to be the one for the PDF of X_t. Let $p(x,t)$ represent the PDF of X_t. That means that time t the random variable X_t has PDF $p(x,t)$, as a function of x. This (as we will see) satisfies the PDE

$$\partial_t p(x,t) = \frac{1}{2} \partial_x^2 p(x,t) \; . \tag{2}$$

I will use is notation for partial derivatives, so

$$\partial_t p(x,t) = \frac{\partial p(x,t)}{\partial t} \quad , \quad \partial_x^2 p(x,t) = \frac{\partial^2 p(x,t)}{\partial x^2} \quad , \quad \partial_x \partial_t p(x,t) = \frac{\partial^2 p(x,t)}{\partial x \partial t} \, , \text{etc.}$$

We use the term *operator* (*linear* operator) for an operation that takes a function and creates another function from it. Partial derivatives are operators in this sense. For example, the operator ∂_t takes a function $p(x,t)$ and produces the function $\partial_t p(x,t)$. An operator is like a linear transformation that might be represented by a square matrix, A. If you apply A twice, the resulting transformation is written A^2 and is given by the square of the matrix A. Similarly, if you apply the operator ∂_x twice, you get the operator ∂_x^2.

One way to see that p satisfies the heat equation (2) is by using a formula involving p that comes from the Gaussian increment formula (1) and the independent increments property. The independent increments property implies the *Markov* property for Brownian motion. An informal statement (which will make more sense soon) is: conditional on the present, the future is independent of the past. Suppose "the present" is a specific time t_1. Then "the past" refers to times $t < t_1$ and the future is times $t > t_1$. Suppose the Brownian motion is at y in "the present", which means $X_{t_1} = y$. This value is determined by all the increments between time 0 and time t_1. All the increments in the future are independent of these past increments, so if $t > t_1$, then $X_t - y = X_t - X_{t_1}$ is independent of all these past increments. We denote the future increment by Z, so $Z = X_t - X_{t_1}$. Of course $X_t = Z + y$, so X_t is not independent of $y = X_{t_1}$.

For this paragraph, let Y represent the Brownian motion location at t_1 and X the location at $t > t_1$. That is $Y = X_{t_1}$ and $X = X_t$. The previous paragraph explains that the random variables Y and $X - Y$ are independent. Therefore, the joint density of Y and $X - Y$ is the product, which we write (specifics explained below)

$$(Y, X - Y) \sim p(x, y, t_1, t) = p(y, t_1) G(x - y, t - t_1) \, . \tag{3}$$

It is common to use $p(\cdots)$ for more or less any probability density. The difference between different densities can be the number of arguments. In that spirit, $p(y, t_1)$ is the PDF of X_{t_1} and $p(x, y, t_1, t)$ is the joint density of (X_t, X_{t_1}). The "G" in (3) is for "Gaussian". The PDF of $X - Y$ is Gaussian with mean zero and variance $t - t_1$, which is the content of (1). The density is a product because of independence. The G density is given by the usual Gaussian formula. If $Z \sim \mathcal{N}(\mu, \sigma^2)$, its PDF is

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \, .$$

Here, Z is $X - Y$, $\mu = 0$ and $\sigma^2 = t - t_1$. Therefore, (written this way to make some future manipulations more clear)

$$G(x - y, t - t_1) = \frac{1}{\sqrt{2\pi}} \, (t - t_1)^{-\frac{1}{2}} \, e^{-\frac{(x-y)^2}{2(t-t_1)}} \, . \tag{4}$$

This is the background.

Now the desired p formula. Since $p(x,t)$ is the density of X_t, and since we have a formula for the joint density of X_{t_1} and X_t, we can view $p(x,t)$ as the marginal of the joint density. You find the marginal density by integrating out the variables you're not interested in, X_{t_1} in this case. In the abstract, this is

$$p(x,t) = \int_{-\infty}^{\infty} p(x,y,t_1,t)\,dy \ .$$

We make this concrete by substituting in the specific relations (3) and (4). The result is either of these equivalent forms

$$p(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} p(y,t_1)(t-t_1)^{-\frac{1}{2}}\, e^{-\frac{(x-y)^2}{2(t-t_1)}}\,dy \ , \tag{5}$$

or

$$p(x,t) = \int_{-\infty}^{\infty} p(y,t_1)\,G(x-y,t-t_1)\,dy \ . \tag{6}$$

Representation formulas like this are used a lot in stochastic calculus.

We use the representation formula (5) or (6) to verify that p satisfies the heat equation (2). We calculate $\partial_t p$ and $\frac{1}{2}\partial_x^2 p$, and they turn out to be the same. Futures class will contain derivations of (2) that come from "first principles", and are not just calculations that verify a formula that obviously was discovered another way. But here, we just see that the variables x and t appear only in G, so the t and x derivatives in (2) only "hit" G. Also, x and t are just parameters in the integrals (5) (6), so they pass into the integral. For example,

$$\partial_t p(x,t) = \int_{-\infty}^{\infty} p(y,t_1)\,[\,\partial_t G(x-y,t-t_1)]\,dy \ .$$

We will see that p satisfies the heat equation because G does.

It may be helpful (it definitely is helpful to me) to start with the simpler case with $t_1 = 0$ and $y = 0$. Then (and only then)

$$G(x,t) = Ct^{-\frac{1}{2}} e^{-\frac{x^2}{2t}} \ .$$

We know that $C = \frac{1}{\sqrt{2\pi}}$, but that is irrelevant for this calculation. The time derivative calculation may be explained as

$$\begin{aligned} t^{-\frac{1}{2}} e^{-\frac{x^2}{2t}} \ &\xrightarrow{\partial_t} \ \left(\partial_t t^{-\frac{1}{2}}\right) e^{-\frac{x^2}{2t}} + t^{-\frac{1}{2}} \left(\partial_t e^{-\frac{x^2}{2t}}\right) \\ &= \left(-\frac{1}{2} t^{-\frac{3}{2}}\right) e^{-\frac{x^2}{2t}} + t^{-\frac{1}{2}} \left(\frac{x^2}{2t^2} e^{-\frac{x^2}{2t}}\right) \\ &= \frac{1}{2} t^{-\frac{3}{2}} \left(-1 + \frac{x^2}{t}\right) e^{-\frac{x^2}{2t}} \ . \end{aligned}$$

We do the two x derivatives one at a time:

$$t^{-\frac{1}{2}} e^{-\frac{x^2}{2t}} \xrightarrow{\partial_x} t^{-\frac{1}{2}} \left(-\frac{x}{t} \right) e^{-\frac{x^2}{2t}} = -t^{-\frac{3}{2}} x e^{-\frac{x^2}{2t}}$$
$$\xrightarrow{\partial_x} -t^{-\frac{3}{2}} e^{-\frac{x^2}{2t}} + t^{-\frac{5}{2}} x^2 e^{-\frac{x^2}{2t}}$$
$$= t^{-\frac{3}{2}} \left(-1 + \frac{x^2}{t} \right) e^{-\frac{x^2}{2t}} \ .$$

Look at these results and you see that $\partial_t G$ is the same as $\frac{1}{2}\partial_x^2 G$. You can do the calculations with $t_1 \neq 0$ and $y \neq 0$. You can put them in the integral (5) and see that p satisfies (2).

2.1 Evolution of the PDF

[We often use a $\cdot$ to represent a variable that does not get a name. This emphasizes that the important thing is the function, not its values at particular places. For example, if f is a function, then $f(x)$ is the value of f at the point x. Writing $f(\cdot)$ emphasizes that f is a function. This is important when a function has more than one argument and we want to describe the function of one of these arguments with the other held fixed. For example, if f is a function of two arguments, then $f(x, y)$ is the value when the arguments are x and y. For any particular value of y, there is a function of x that you get by fixing y. Some ways to denote this function are $g_y(x) = f(x, y)$ or $g_y(\cdot) = f(\cdot, y)$.]

Evolution (in this context) means changing over time. The formulas (5) or (6) are formulas for $p(\cdot, t)$ in terms of $p(\cdot, t_1)$, if $t > t_1$. This implies a *forward* evolution. As time moves forward, it is possible to determine the forward (future) probability density $p(\cdot, t)$ from the present density $p(\cdot, t_1)$.

This forward evolution may be written as

$$p(\cdot, t) = G(t - t_1) p(\cdot, t_1) \ . \tag{7}$$

The G is this formula is a *linear operator* that is represented by the integral formula (6). The operator $G(t)$ may be applied to functions that are not probability densities. The definition is

$$g = G(t) f \ \text{ means } \quad g(x) = \int_{-\infty}^{\infty} G(x - y, t) f(y) \, dy \ . \tag{8}$$

You can think of a linear operator as being like a linear transformation in n dimensional vector space, where G would be represented by an $n \times n$ matrix. But here there is no matrix. Instead the *action* of the operator $G(t)$ is given by the integral (8). Remember that t in this formula is just a parameter. We do not integrate over t, at least not in the definition (8).

Is it possible to go backwards in time? Can you determine $p(\cdot, t_1)$ from $p(\cdot, t)$ for $t > t_1$? If $G(t)$ were a linear transformation on an n dimensional vector space, then the operation of $G(t)$ could be reversed by the inverse matrix $G(t)^{-1}$. The

problem of reversing the direction of time is the same as the problem of inverting the linear operator (8).

Linear operators are in some ways analogous to linear transformation of finite dimensional vector spaces. So, suppose A is an $n \times n$ matrix and ask whether that "action" of A can be reversed. For $x \in \mathbb{R}^n$, the action of A is $x \rightsquigarrow y = Ax$. Undoing that would mean recovering x from y. The linear algebra formula is $y = A^{-1}x$. In linear algebra, there is a simple criterion for A^{-1} to exist. Suppose there is a vector $v \neq 0$ with $Av = 0$. Then if you know $y = 0$, you do not know whether $x = 0$ or $x = v$, because $A0 = 0$ and $Av = 0$. You can say that the memory of v is erased by A. We know (these are facts of linear algebra) that there is such a $v \neq 0$ if and only if $\det(A) = 0$. If $\det(A) \neq 0$, then A^{-1} exists. This kind of A^{-1} is a *left inverse*, because it satisfies $A^{-1}Ax = x$ for all x. The inverse is on the left in the matrix product.

A *right inverse* is an A^{-1} with $AA^{-1}y = y$ for all y. If $x = A^{-1}y$, then $x = Ay$. This is the problem of solving, which means starting with an arbitrary $y \in \mathbb{R}^n$ and finding an x with $Ax = y$. A matrix that has an inverse in this sense is *onto* (also called *surjective*, from the French word "sur", which means "on"). A matrix with a left inverse is *into* (also called *injective*). That means that if $x_1 \neq x_2$ then $y_1 = Ax_1 \neq y_2 = Ax_2$, which is the same as $A(x_1 - x_2) \neq 0$ if $x_1 - x_2 \neq 0$, which is how we expressed this in the previous paragraph. It is a fact of linear algebra that an $n \times n$ matrix has a right inverse if and only if it has a left inverse (if and only if A $\neq 0$). But "into" (left inverse) and "onto"(right inverse) seem to be different things.

Right and left inverse for the evolution operator $G(t)$ also seem different. For a right inverse, we would be given a $p(\cdot)$ and then we would look for q with $Gq = p$. We will soon see that G has a *smoothing property* which makes this impossible. Even if q is discontinuous, $p = Gq$ is differentiable. Therefore (we will see) there is no q if p is discontinuous – no right inverse.

The left inverse is more subtle. There is no q with $Gq = 0$. No q gets erased completely. But it is possible to show (this may be too long a detour for an applied math class) that there are functions q that are "reduced" by an arbitrarily large factor. For any $\epsilon > 0$ there is a q so that if $p = Gq$, then $\|p\| \leq \epsilon \|q\|$. If p is not known exactly, but has an error ϵ, then it is impossible to know whether it had this q in it or not. The technical term for this phenomenon is *ill-posededness*. A coming computing assignment will demonstrate that running the heat equation "backwards" (which is the same as finding the inverse of G (running the heat equation "forwards")) is ill posed. If you do it on the computer, the computation "blows up". The computer isn't damaged but the results are garbage.

2.2 Short time evolution

It is interesting and important to understand what happens in short amounts of evolution. The Brownian motion path is a continuous function of t, so it is natural to think that $G(t)p$ is not much different from p for small t. If X_t is close to X_0, then the PDF of X_t should be close to the PDF of X_0. First we

will see that this true, then we will see that it is not true. Pay attention to the difference.

Look at the *integral kernel* ("kernel" probably comes from the German word "Kern", which means core) in (4), with $t_1 = 0$ for simplicity. We simplify the operation (5) to

$$p(x) = \int_{-\infty}^{\infty} G(x-y,t)q(y)\,dy\ . \tag{9}$$

The kernel is positive and integrates to 1:

$$\int_{-\infty}^{\infty} G(x-y,t)\,dy = 0\ .$$

Therefore the value $p(x)$ is a weighted average of the values of $q(y)$. When t is small, the kernel (4) is "exponentially small" when y is far from x. Therefore, $p(x)$ is an average of $q(y)$ for y values close to x. If q is continuous, then $q(y) \approx q(x)$ for $y \approx x$, so

$$p(x) \approx \int_{-\infty}^{\infty} G(x-y,t)q(y)\,dy \approx \int_{-\infty}^{\infty} G(x-y,t)q(x)\,dy = q(x)\ .$$

If you have taken mathematical analysis, you can make a proof of this theorem: if q is bounded and continuous, then $p(x,t) \to q(x)$ as $t \downarrow 0$.

The English physicist Dirac (Sir Paul Adrian Maurice Dirac, F.R.S., to be precise) described this using what we call the *Dirac delta function*, which is written $\delta(x)$. This is not an actual function, but an abstract "generalized function" that has a unit "mass" at zero and is equal to zero elsewhere. That is

- $\delta(x) \geq 0$ for all x,
- $\delta(x) = 0$ if $x \neq 0$.
- $\int_{-\infty}^{\infty} \delta(x)\,dx = 1$.

These properties imply that

$$f(x) = \int_{-\infty}^{\infty} \delta(x-y)f(y)\,dy\ .$$

In the integral, $\delta(x-y) = 0$ if $x \neq y$, so only $x = y$ matters, and there f is $f(x)$.

Although the delta function is not a real function (with real function values for each x), it is possible for a family of real functions to converge to the delta function. Our G does that:

$$G(x,t) \to \delta(x) \quad \text{as} \quad t \downarrow 0\ .$$

You can see what this means by drawing a sequence of graphs of G as $t \downarrow 0$. The mathematical theorem is what we said just before, that $p(x,t) \to q(x)$ as $t \downarrow 0$.

Now consider the solution of the heat equation, the PDF of X_t, if $p(\cdot, 0)$ is not continuous. Specifically, suppose X_0 is random and uniformly distributed in $[0, 1]$. This is the same as saying that $p(x, 0) = 1$ if $0 \leq x \leq 1$, and $p(x, 0) = 0$ otherwise. If x is not near 0 or 1, then $p(x, 0)$ is continuous and $p(x, t) \to 0$ or 1 as $t \downarrow 0$. However, the jump discontinuities at $x = 0$ and $x = 1$ are replaced by rapid transitions. $p(x, t)$ goes from almost 0 for $x < 0$ to almost 1 for $x > 0$, making the transition rapidly. The width of the transition (for example, the distance between $p(x, t) = .2$ and $p(x, t) = .8$) is on the order of $t^{\frac{1}{2}}$. For $t > 0$, $p(\cdot, t)$ may be expressed in terms of the *cumulative normal* function

$$N(x) = \Pr_{\mathcal{N}(0,1)}(Z < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} \, dz \ .$$

The Z random variable Gaussian with mean zero and variance 1 (a standard normal). This goes to 0 as $x \to -\infty$ and 1 as $x \to \infty$. Near $x = 0$ and for small $t > 0$, we have the approximation

$$p(x, t) \approx N(t^{-\frac{1}{2}} x) \ .$$

For small t, $t^{-\frac{1}{2}}$ is large. Therefore $t^{-\frac{1}{2}} x$ is either far to the left of 0 (negative) or far to the right of 0 if $x \neq 0$. The corresponding values of $N(t^{-\frac{1}{2}} x)$ are close to 0 or close to 1. These formulas are verified in Exercise 1 of Assignment 2.

2.3 Probability flux

You can think of probabilitconditionaly density of Brownian motion as "flowing" along the x axis as $p(\cdot, t)$ evolves. If a is some point on the x axis, then probability cannot go from $x < a$ to $x > a$ without crossing a. There is a *probability flux* (also called *probability current* and written as $\mathbb{F}(a, t)$) that says how fast probability is flowing at location a at time t. If $\mathbb{F} > 0$, then probability is flowing to the right. If $\mathbb{F} < 0$, then probability is flowing to the left. The amount of probability between a and b at time t is

$$\Pr(a \leq X_t \leq b) = \int_a^b p(x, t) \, dx \ .$$

The rate of change described by the flux at the endpoints a and b. If $\mathbb{F}(a, t) > 0$, then probability is flowing into the interval $[a, b]]$ at the left endpoint. If $\mathbb{F}(b, t) > 0$, then probability is flowing out of the interval $[a, b]]$ at the right endpoint. The mathematical statement is

$$\frac{d}{dt} \Pr(a \leq X_t \leq b) = \frac{d}{dt} \int_a^b p(x, t) \, dx = \mathbb{F}(a, t) - \mathbb{F}(b, t) \ . \tag{10}$$

This formula is related to the fact that Brownian motion paths X_t are continuous functions of t. The Brownian motion cannot go from inside $[a, b]$ to outside without crossing one of the boundary points. The probability inside cannot leave or enter except by being flux at one of the endpoints.

We find a formula for $\mathbb{F}$ and verify the "local conservation" formula (10) using the heat equation and the fundamental theorem of calculus (the integral of $\partial_x^2 p$ is $\partial_x p$ evaluated at the endpoints)

$$\begin{aligned}\frac{d}{dt}\int_a^b p(x,t)\,dx &= \int_a^b \partial_t p(x,t)\,dx \\ &= \int_a^b \frac{1}{2}\partial_x^2 p(x,t)\,dx \\ &= \frac{1}{2}\partial_x p(b,t) - \partial_x p(a,t) \\ &= -\left(-\frac{1}{2}\partial_x p(b,t)\right) + \left(-\frac{1}{2}\partial_x p(a,t)\right)\ .\end{aligned}$$

This verifies the local conservation formula (10), with probability flux formula

$$\mathbb{F}(a,t) = -\frac{1}{2}\partial_x p(a,t)\ . \tag{11}$$

The heat equation itself may be written using the flux as

$$\partial_t p(x,t) + \partial_x \mathbb{F}(x,t) = 0\ . \tag{12}$$

To understand this, imagine that $\mathbb{F}$ is constant (independent of x). Then $\partial_x \mathbb{F} = 0$, so $\partial_t p = 0$. If the flux is constant at some point then the probability density is not changing there. The probability density isn't changing because probability flows in at the same rate it flows out. If $\mathbb{F}$ is increasing (as a function of x), then probability flows out of the right endpoint a little faster than it flows in at the left. Therefore the probability inside is decreasing (as a function of t. In (12), if $\partial_x \mathbb{F} > 0$, then $\partial_t p < 0$.

The specific flux formula (11) is called *Fick's law* or the *Fourier law*. It says that probability flows from regions of high probability density to regions of low probability density. If $\partial_x p > 0$ at some point, then $\mathbb{F} < 0$ at that point. The density is larger on the right so it is more likely that the particle goes from right to left than that it goes from left to right. The term *Fourier law* is often used when the heat equation is used to model the flow of heat (as opposed to probability). Fourier himself "wrote the book on" the flow of heat (the first book). *Fick's law* is often used when the heat equation models the diffusion of a chemical through a gas or liquid. The heat equation is often called the *diffusion equation* for that reason.

3 Hitting probability

Let X_t be Brownian motion starting from $X_0 = x_0$ which is not random. Suppose $x_0 > 0$. The *hitting probability* for $x = 0$ up to time t is

$$\Pr(X_s = 0 \ \text{ for some } 0 \le s \le t)\ .$$

The *hitting time* (also called *first hitting* time or *first passage time*) is

$$T = \min \{t \mid X_t = 0\} \ .$$

If $X_t > 0$ for all $t > 0$, we say $T = \infty$. This is a random variable that depends on the random path X_t. We call the point $x = 0$ the *absorbing boundary*. The point $x = 0$ is the boundary (the end) of the region $x > 0$ that we are interested in. The goal of this section is the PDF: $T \sim u(t)$.

This calculation does not consist of integrating a probability density for X_t (at least, not in a simple way). Indeed, in some sense there is no probability density for the path. There is a probability density for the values X_t for fixed t, but not for the whole path (we will return to this statement, which physicists may disagree with). Instead, the calculation uses the heat equation, the probability flux, and a trick for solving the relevant heat equation called the *method of images*.

To calculate hitting probabilities and the hitting time PDF, let $p(x,t)$ be the probability density for a path that has not hit yet. That is, if $x > 0$, then

$$p(x,t)dx = \Pr\left(x \leq X_t \leq x + dx \ \text{ and } T > t\right) \ . \tag{13}$$

This is only defined for $x > 0$. The *survival probability* is the complement to the hitting probability. It is given in terms of this density function by

$$S(t) = \Pr(T > t) = \int_0^\infty p(x,t)\, dx < 1 \ .$$

The density p of a surviving Brownian motion (13) satisfies the heat equation if $x > 0$. I do not give the proof here, but it depends on the fact that Brownian motion paths are continuous. If $X_{t_1} > 0$, and if the time increment $t - t_1$ is small, then X_t is so unlikely to reach the boundary that it does not effect (for very short time) the evolution of the PDF. The density (13) satisfies an *absorbing boundary condition* at the absorbing boundary, which is $p(0,t) = 0$. The intuition for this is that it is unlikely to find a surviving Brownian motion close to the boundary because it probably hit the boundary before time t_1. We will see this in simulations but I think a real proof is too long and technical to be of interest to a majority of students here. If the path starts at $X_0 = x_0$ which is some distance from the boundary, then X_t is unlikely to hit the boundary for small enough t. Therefore, for small t, the density (13) satisfies $p(x,t) \approx \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-x_0)^2}{2t}}$. To summarize, we find p by finding the function that satisfies

- (initial condition) $p(x,0) = \delta(x)$, which is the same as $p(x,t) \approx \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-x_0)^2}{2t}}$ for small t, if $x > 0$.
- (boundary condition) $p(0,t) = 0$.
- (PDE) p satisfies the heat equation if $x > 0$.

The trick for finding p is to look for a function (also called p) that satisfies these three conditions but is defined for all x, not just for $x > 0$. This function will be anti-symmetric with respect to the absorbing boundary, which means $p(-x,t) = -p(x,t)$. Since $p(x,t) > 0$ for $x > 0$ (being a probability density), this implies that $p(x,t) < 0$ if $x < 0$. Therefore the extended p cannot be interpreted as a probability density for $x < 0$. The skew-symmetric extension automatically satisfies the absorbing boundary condition (check by algebra or draw a picture), which is the principle behind the trick.

The initial condition for the skew symmetric extension has a delta function at $x = x_0$ and (by skew-symmetry) a negative delta function at $-x_0$. The solution is the sum of these two solutions

$$p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-x_0)^2}{2t}} - \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x+x_0)^2}{2t}} \ . \tag{14}$$

The first term comes from the delta function at x_0. The second term comes from the delta function at $-x_0$. Note that $x - (-x_0) = x + x_0$, which explains the exponent in the second term. We call the delta function at x_0 the *point charge* at x_0. It is a unit "charge" (probability mass) at x_0. The delta function at $-x_0$ is the *image charge*. The point $-x_0$ is the image of x_0 if you reflect through the reflecting boundary. You can check that this function satisfies the three conditions above. The initial condition there is only for $x > 0$. The extended p also has a charge at $-x_0$, which is consistent with that. You can check that p given by (14) satisfies $p(x,t) > 0$ for $x > 0$ (the negative part is smaller than the positive part because x is closer to x_0 than it is to $-x_0$).

The hitting time density satisfies

$$\begin{aligned} u(t)dt &= \Pr(t \le T \le t + dt) \\ &= S(t) - S(t + dt) \\ &= -\frac{d}{dt} S(t) dt \\ u(t) &= -\frac{d}{dt} S(t) \ . \end{aligned}$$

We argued above that

$$\frac{dS}{dt} = \mathbb{F}(0,t) \ \ , \quad \mathbb{F}(0,t) = -\frac{1}{2} \partial_x p(0,t) \ .$$

Therefore,

$$u(t) = \frac{1}{2} \partial_x p(0,t) \ .$$

The calculation from the explicit formula (14) finally gives

$$u(t) = \frac{2x_0}{\sqrt{2\pi}} t^{-\frac{3}{2}} e^{-\frac{x_0^2}{2t}} \ . \tag{15}$$

This is the hitting time probability density – a famous formula we will use a lot.

What does $u(t)$ look like for small t and for large t? For small t, the prefactor $Ct^{-\frac{3}{2}}$ goes to infinity, but the exponential $-\frac{x_0^2}{2t}$ goes to minus infinity. Since exponentials beat powers, this shows that $u(t) \to 0$ as $t \to 0$ "exponentially". It is very unlikely to hit the boundary very quickly. When $t \to \infty$, the exponent goes to zero. Therefore $e^{-\frac{x_0^2}{2t}} \to 1$ as $t \to \infty$ and $u(t) \sim Ct^{-\frac{3}{2}}$. As $t \to \infty$, this probability density goes to zero, but slowly. It can't be too slow because the integral has to converge. But it is so slow that the expected hitting time is infinite (the integrand is "like" $t^{-\frac{1}{2}}$ for large t, so the integral diverges):

$$\mathrm{E}[T] = \int_0^\infty t u(t)\, dt = \infty \, .$$

3.1 Reflection principle

The method of images formula (14) has a consequence that is (in this context in probability) called the *reflection principle*. That is,

$$\Pr(X_s \le 0 \ \text{ for some } s \le t) = 2 \Pr(X_t \le 0) \, . \tag{16}$$

If $X_t < 0$, then $X_s = 0$ for some $s \le t$. Therefor the probability of being "out" (outside the domain $x \ge 0$ at time t is less than the probability of ever having been out. But (16) says that the two probabilities differ exactly by a factor of 2.

To prove this formula, use the fact that

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

Also, the probability of being out at time t is

$$\Pr(X_t < 0) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-x_0)^2}{2t}} \, dx \, .$$

The hitting probability is

$$\int_0^\infty p(x,t)\, dx = 1 - \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-x_0)^2}{2t}} \, dx - \int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x+x_0)^2}{2t}} \, dx \, .$$

The two integrals on the right are equal (check this), and either of them is $\Pr(X_t \le 0)$. This proves the reflection principle (16).

The Russian mathematician Kolmogorov must have seen (16) in this way and wondered whether there is a proof that doesn't rely on the solution formula for Brownian motion. He found such a simple argument, so the reflection principle formula (16) is true whenever X_t is a "symmetric martingale". This is common in math. You do some calculations and come to an unexpected formula. Then you wonder why this unexpected formula might be true, and you find a derivation that is simpler and more natural than the computation that led you there in the beginning.