Stochastic Calculus, Courant Institute, Fall 2014
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2014/index.html

Section 3, Simple Gaussian processes in continuous time

Jonathan Goodman, October 6, 2014

1 Introduction

This section is on Brownian motion and a small extension of it, the Ornstein Uhlenbeck process. Brownian motion is in some senses the simplest continuous time continuous path stochastic process. It is a reasonable yet tractable model for many stochastic processes. It is used as a mathematical way to model random noise in most continuous time continuous sample path processes.

The name comes from English biologist Robert Brown. In the 1820's he was looking at pollen grains in his new more powerful microscope. Amazingly, the particles seemed to be swimming. They were all in rapid and seemingly random motion. Brown then observed fine particles of clay, which move in a similar way. Brown, and science in general, were at a loss to understand what makes small particles move about randomly in water.

The first physical explanation was given by Albert Einstein in 1905. Although clay particles are seem small to us, they are much larger than water molecules. Einstein suggested that the particles move because they are being pushed about by random collisions with water molecules. In this picture, the fine scale random motion of a clay particle is the large scale result of many small independent random impacts from water molecules. The clay particle motions are large scale relative to water molecules. Einstein made a quantitative theory of this and used it to give one of the first physical estimates of the size of water molecules. Chemists had believed in discrete molecules for a long time, but they only knew that they were very small, without knowing how small.

The mathematical model of Brownian motion, in one dimension, is a random number X_t defined for each $t > 0$. The subscript notation X_t suggests that X_t is the t component of a random object, X. It may be more natural to write it as $X(t)$ to emphasize that X is a function of t. It turns out that X_t is a continuous function of t, but not a smooth function. In this class, Brownian motion plays three important roles:

1. It describes the large scale behavior of a large class of stochastic processes, like the pollen grain.

2. It has a model for understanding diffusion processes (the main topic of stochastic calculus) in general.

3. It is a mathematical way to describe the noise that drives general diffusion processes.

2 The Brownian motion central limit theorem

Undergraduate probability deals with discrete or finite dimensional random variables, probabilities and probability densities. PhD level probability deals with abstract probability spaces and measures. Brownian motion is one of the first random objects where the power and generality of measure theory is useful and helpful. This class is too short to discuss measure theory completely. This technical section is an attempt to raise awareness rather than to answer all questions.

The central limit theorem describes the distribution of the sum of a large collection of independent random variables. Suppose Y_k is an i.i.d. family of mean zero variance 1 random variables, and define the sums as

$$S_n = \sum_{k=1}^{n} Y_k \ .$$

The conclusion of the central limit theorem (CLT) is that as $n \to \infty$, the distribution of S_n depends less and less on the distribution of the Y_k and is more and more nearly Gaussian. *Scaling* is a mathematical device that makes it possible to state this in a simple way. The *scaled sum* is

$$X_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k \ . \tag{1}$$

The central limit theorem states that the scaled sums converge *in distribution* to a Gaussian:

$$X_n \stackrel{d}{\rightharpoonup} Z \ , \quad Z \sim \mathcal{N}(0,1) \ , \quad \text{as } n \to \infty. \tag{2}$$

The convergence symbol in (2), $\stackrel{d}{\rightharpoonup}$, means convergence *in distribution*. It does not mean that the numbers X_n converge to the number Z as $n \to \infty$. It means that the probability distribution of X_n converges to the probability distribution of Z. The numbers themselves do not converge. For example, X_{2n} is probably not close to X_n because half of the Y_k that define X_{2n} are not involved in X_n. But the distribution of X_n is approximately normal, with the approximation improving as n increases. The little d over the convergence arrow in $\stackrel{d}{\rightharpoonup}$ is for convergence in *distribution*.

The half arrow $\rightharpoonup$ rather than the full arrow $\to$ indicates *weak convergence*. Weak convergence means that even the probability densities may not converge. Let $f_n(x)$ be the PDF of X_n and let $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ be the PDF of Z. Then *strong convergence* of the densities would be $f_n(x) \to f(x)$ as $n \to \infty$ for each x (pointwise strong convergence), or

$$\int |f_n(x) - f(x)| \, dx \to 0 \ , \quad \text{as } n \to \infty.$$

(L^1 convergence). But the CLT applies in situations where X_n these statements do not make sense. For example, if $Y = \pm 1$ (with equal probability), then

$E[Y] = 0$ and $\mathrm{var}(Y) = 1$, but the "density function" of X_n has delta functions at each of the numbers $x = \sqrt{k/n}$ with $-n \leq k \leq n$. If x^2 is an irrational number, then $f_n(x) = 0$ for all n.

Weak convergence, which is written $f_n \rightharpoonup f$ as $n \to \infty$, is possible in situations like this. We say that $X_n \stackrel{d}{\rightharpoonup} Z$ as $n \to \infty$ if, whenever $V(x)$ is a bounded and continuous function of x,

$$\mathrm{E}[V(X_n)] \to \mathrm{E}[V(Z)] \quad , \quad \text{as } n \to \infty. \tag{3}$$

We require V to be bounded so that $\mathrm{E}[V(X)]$ exists and is finite for any random variable X. To see why V should be continuous, suppose $X_n = \pm\frac{1}{n}$ (equal probability or not) and $X = 0$. If V is continuous, then $\mathrm{E}\big[V(\pm\frac{1}{n})\big] \to V(0) = \mathrm{E}[V(X)]$ as $n \to \infty$. But if V is allowed to be discontinuous, we could take $V(0) = 1$ and $V(x) = 0$ for $x \neq 0$. This V has $\mathrm{E}\big[V(\pm\frac{1}{n})\big] = 0$ and $\mathrm{E}[V(X)] = 1$. The CLT in finite dimensions states that if Y_k is an i.i.d. family with $\mathrm{E}[Y_k] = 0$ and $\mathrm{cov}[Y_k] = C$, and if $Z \sim \mathcal{N}(0, C)$, and if $V(x)$ is a bounded and continuous function of $x \in \mathbb{R}^n$, then (3) holds.

Brownian motion is a large scale description of a path that is a sum of a large number of small independent "shocks". The CLT is an approximation description of the distribution of S_n for large but fixed n. Brownian motion is an approximation to the way S_n depends on n. For the CLT we did a scaling, the scaled variable $X_n = \frac{1}{\sqrt{n}} S_n$ has a distribution with a (Gaussian) limit as $n \to \infty$. Now we rescale the sums S_n to get a random process. The *scaling parameter* is a small δ that will go to zero. The scaled sum process is

$$X_t^{(\delta)} = \sqrt{\delta} \sum_{k\delta < t} Y_k \ . \tag{4}$$

Brownian motion is the process X_t so that

$$X_t^{(\delta)} \stackrel{d}{\rightharpoonup} X_t \quad , \quad \text{as } \delta \to 0. \tag{5}$$

This theorem is called the *Donsker invariance principle*, named after Monroe Donsker, who was at the Courant Institute most of his career.

The invariance principle contains the CLT. If you take $t = 1$ and $\delta = \frac{1}{n}$, then $X_t^{(\delta)} = X^{(\frac{1}{n})}$ is exactly what we called X_n before, scaling and all. For any fixed t, the number of terms in the sum (4) is approximately t/δ and

$$\mathrm{var}(X_t^{(\delta)}) \to t \quad , \quad \text{as } \delta \to 0.$$

Since $X_t^{(\delta)}$ is the sum of a large number of i.i.d. random variables, its distribution is approximately Gaussian. This shows that the limit process X_t has values at any particular t that are Gaussian:

$$X_t \sim \mathcal{N}(0, t) \ .$$

The *independent increments* property goes further. The *increment* of Brownian motion between t and $t+s$ is the difference $X_{t+s} - X_t$. The value at $t+s$ is

the value at t plus the increment: $X_{t+s} = X_t + (X_{t+s} - X_t)$. One independent increments fact is that the increment $X_{t+s} - X_t$ is independent of the value X_t. But we can go further. The *path* up to time T, which is $X_{[0,T]}$, is a random object determined by the small "shocks" Y_k for $1 \leq k < T/\delta$. The increment path beyond T, which is $X_{T+s} - X_T$, is a random object determined by the shocks for k in the range $k \geq T/\delta$. Thus, the shocks that determine $X_{[0,T]}$ are independent of the shocks that determine the increment path. The increment path $X_{T+s} - X_T$ is thus independent of the path up to time T.

The increments are Gaussian. Indeed,

$$X^{(\delta)}_{T+s} - X^{(\delta)}_T = \sqrt{\delta} \sum_{T \leq k\delta < T+s} Y_k$$

is the sum of many independent contributions, so its distribution converges to Gaussian as $\delta \to 0$. The argument we gave earlier shows that the variance (in the limit $\delta \to 0$) is $T - s$, so

$$X_{T+s} - X_T \sim \mathcal{N}(0, s) \; .$$

The independent increments property implies that this increment is independent of X_T, so the joint distribution is a two dimensional uncorrelated normal. In general, suppose $0 = T_0 < T_1 < T_2 < \cdots < T_n$ and the increments are defined by

$$X_{T_j} = X_{T_{j-1}} + \xi_j \; .$$

Then the increments ξ_j are independent Gaussians with variance $T_j - T_{j-1}$. The joint PDF of the ξ_j is

$$u(\xi_1, \xi_2, \cdots, \xi_n) = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{j=1}^{n} \frac{\xi_j^2}{T_j - T_{j-1}} \right) \; .$$

This may be written in terms of the values of X_{T_j}:

$$u(X_{T_1}, X_{T_2}, \ldots, X_{T_n}) = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{j=1}^{n} \frac{(X_{T_j} - X_{T_{j-1}})^2}{T_j - T_{j-1}} \right) \; . \tag{6}$$

You may be uncertain that we can just substitute the X variables for the ξ variables to get the PDF for the joint distribution of the values $X_{t_1}, \ldots, X_{T_n}$ from the joint distribution of $\xi_1, \ldots, \xi_n$. If you are, review the Section 1 notes on *linear* change of variables in multi-variate densities. The n component vector $X_{t_1}, \ldots, X_{T_n}$ is a linear function of the n component vector $\xi_1, \ldots, \xi_n$.

The formulas (6) do not give a PDF for the random path $X_{[0,T]}$ itself. In fact, it is in principle impossible to give a PDF for the path for the following reason. If $Y \in \mathbb{R}^n$ is a random variable with PDF $u(y)$, this means that if $V(y)$ is a bounded continuous function, then

$$\mathrm{E}[V(Y)] = \int_{\mathbb{R}^n} V(y) u(y) \, dy \; .$$

In n dimensions, the *integration element* is $dy = dy_1 dy_2 \cdots dy_n$. There is no corresponding integration element in *path space*. We don't know how to interpret the formula

$$\text{(wrong)} \qquad \mathrm{E}\big[V(X_{[0,T]})\big] = \int_\Omega V(x_{[0,T]}) u(x_{[0,T]})\, dx_{[0,T]} \; . \qquad \text{(wrong)}$$

The probability space, Ω, consists of all continuous paths X_t defined for $0 \leq t \leq T$ with $X_0 = 0$ (the definition (4) and (5) gives $X_0 = 0$). The first two factors on the right are OK, $V(x_{[0,T]})$ and $u(x_{[0,T]})$ could be functions of the random path $x_{[0,T]}$, but there is no real definition of the integration element $dx_{[0,T]}$.

Abstract probability is something that can be defined even when there are no probability densities. There is a class of *measurable events* $A \subseteq \Omega$ that form a $\sigma-$algebra, $\mathcal{F}$. For each such event, there is a probability $P(A)$. The collection of all these probabilities forms a *probability measure* if it satisfies the normal properties of probabilities and is *countably additive*. The "normal properties" are pretty standard:

- $0 \leq P(A) \leq 1$ for any $A \in \mathcal{F}$
- $P(\Omega) = 1$, and $P(\emptyset) = 0$.
- If $A \in \mathcal{F}$ and $B \in \mathcal{F}$ and A is disjoint from B, then $P(A \cup B) = P(A) + P(B)$.

Countable additivity is the requirement that the probability measure works with limits. One of the many equivalent ways to state the property involves an increasing family of events $A_1 \subseteq A_2 \subseteq \cdots$. The "limit" of this expanding family of events is the union $A = \cup_j A_j$. Countable additivity says that the probabilities should have the correct limit

$$P(A) = \lim_{j \to \infty} P(A_j) \; .$$

In discrete probability, at the finest level it was possible to suppose that every event is measurable. This is not possible in continuous probability. No matter how much "information" you have, there still must be events whose probability is not known. You can find a proof of this statement by googling "non-measurable set". However, any event that has an explicit description is measurable (precise statement and proof omitted). There is an abstract definition of integration with respect to a probability measure that we may say a little about in a later section.

3 Brownian motion and the heat equation

One of the central themes of stochastic calculus is the relationship between stochastic processes like Brownian motion and partial differential equations (or

PDE's). The probability density is one of the many quantities related to X_t that satisfies a PDE. Let $u(x,t)$ be the PDF of X_t. If $X_0 = 0$, then $X_t \sim \mathcal{N}(0,t)$, so

$$u(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \ .$$

This function satisfies the PDE

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

We check this by direct verification:

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

Also

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

This verifies (7). There are more conceptual derivations that do not rely on knowing the answer in advance, as we will see. This is very important if we want to apply differential equation methods to problems where the PDF is not known in advance.

It is possible to consider a slightly more general Brownian motion where X_0 is random. Just adjust (4) to

$$X_t^{(\delta)} = X_0 + \sqrt{\delta} \sum_{k\delta < t} Y_k \ ,$$

and take the limit $\delta \to 0$. For this, reasoning we have used before gives

$$X_t \sim X_0 + \mathcal{N}(0,t) \ ,$$

where the two random variables on the right are independent. We make a small shift in notation to denote the Gaussian density

$$G(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \ . \tag{8}$$

The G stands either for "Gaussian" or for "Green's function" (a concept we will discuss more later, but common in engineering and physics). This G is the

PDF of the Gaussian contribution to X_t. The PDF of the sum of independent random variables is the convolution of the probability densities. Therefore

$$u(x,t) = \int_{-\infty}^{\infty} G(x-y,t)u_0(y)\,dy\ . \tag{9}$$

This representation of $u(x,t)$ implies that u satisfies the heat equation (7) regardless of the initial distribution $u(x,0)$. If you apply t and x derivatives to u on the left, the derivatives on the right fall onto the function G, which satisfies the heat equation.

A more general version of (9) is helpful. Suppose $0 \leq s < t$ is some intermediate time. Then $X_t = X_s + (X_t - X_s)$. The term in parentheses is the increment of Brownian motion between time s and time t. This increment is $\mathcal{N}(0, t-s)$ and independent of $X_{[0,s]}$. Therefore

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

In this context, the function G is called the *fundamental solution*, or the *Green's function*. Any solution can be build from the fundamental solution by an integral like this. If t is just a little larger than s, this gives an expression for the small change in u that happens over a small time interval.

The formula (10) applies only when $t > s$. There are many deep reasons for this, but you can see it superficially by noting that $G(x, t-s)$ in (8) is not defined for $t < s$. You can determine u in the future from u in the present, but you cannot determine u in the present from u in the future. You can "run" the heat equation (7) forward in time (determine u at time $t > s$ from u at time s), but you cannot run the heat equation backwards in time.

The formula (10) defines a family of *linear operators*, $G(t)$ that move the solution u forward in time by an amount t. We write this as

$$u(s) \quad \stackrel{G(t-s)}{\longrightarrow} \quad u(t)\ .$$

In this expression, $u(s)$ represents the function of x which is $u(x,s)$. It says that the operator $G(t-s)$ transforms $u(s)$ into $u(t)$. We write $u = G(t)v$ if the operator $G(t)$ transforms v to u. The formal definition is

$$u = G(t)v \qquad \text{if} \qquad u(x) = \int_{-\infty}^{\infty} G(x-y,t)v(y)\,dy\ .$$

The operator family $G(t)$ form the *solution operator* for the heat equation (7). The solution of the heat equation is given by the solution operator expression $u(t) = G(t)u(0)$, which is the same as (9).

The family of operators $G(t)$ satisfies the *semigroup property*. We can write $u(s) = G(s)u(0)$ and then $u(t) = G(t-s)u(s)$. This gives $u(t) = G(t-s)G(s)u(0)$. Also $u(t) = G(t)u(0)$. Together, these imply that

$$G(t) = G(t-s)G(s)\ . \tag{11}$$

With the x variable written in, this is

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

The mathematical term *group* refers to a collection of objects that can by "multiplied" and inverted. If A and B are elements of a group, the product AB also is an element of the group, and so is A^{-1}. The definition of A^{-1} is that $A^{-1}A$ is the *identity*. The operators $G(t)$ can be multiplied using the formula above, but they cannot be inverted. A collection of objects that can be multiplied but not inverted is a *semi-group*. The formula (11) expresses the fact that the solution operators for the heat equation form a semigroup.

Smoothing is an important property of the heat equation (7). Suppose the initial PDF $u(x,0)$ is uniform in the interval $[0,1]$. This PDF is discontinuous at $x = 0$ and $x = 1$. But the solution $u(x,t)$ given by (9) has no discontinuities. If t is small, then $u(x,t)$ is close to 1 if x if $\varepsilon < x < 1-\varepsilon$, and $u(x,t)$ is close to zero if $x < -\varepsilon$ or $x > 1+\varepsilon$. We will see soon that this is true for $t = .01$ and $\varepsilon = .2$, say. You can see that $u(x,t)$ is differentiable. For example, we can calculate that

$$\begin{aligned}\partial_x u(x,t) &= \int_{-\infty}^{\infty} \left[\partial_x G(x-y,t)\right] u(y,0)\,dy \\ &= \int_{-\infty}^{\infty} \left[-\frac{x-y}{t} G(x-y,t)\right] u(y,0)\,dy\ .\end{aligned}$$

The quantity in square brackets is bounded, so $\partial_x u(x,t)$ is bounded. The same goes for $\partial_x^2 u(x,t)$ and higher derivatives. For small t, $u(x,t)$ makes a finite speed, but fast, transition from near 0 to near 1 as x goes from $-\varepsilon$ to ε. But the exact discontinuity of $u(x,0)$ has disappeared.

The smoothing property is responsible for the fact that we cannot "run the heat equation backwards". Running the heat equation backwards would mean specifying $u(x,t)$ for some $t > 0$ and asking what $u(x,0)$ it would take to achieve that. But if we ask for $u(x,t)$ to be discontinuous, it is impossible to get a $u(x,0)$. Not only is discontinuous $u(x,t)$ not allowed, but even discontinuities in derivatives. The solution operator of the heat equation gives a *smooth* function, a function that has derivatives of all orders.

4 Hitting times

The *hitting time* at a for X_t is the first time X "hits" a. More technically,

$$\tau_a = \min\ \{\, t \mid X_t = a\}\ . \tag{12}$$

These are important in many applications of stochastic processes. How long will it take for a random stock price to reach a given level? How long will it take a randomly moving particle to touch the boundary of a container? They are

important in mathematical analysis of stochastic processes. If $|X_{t+s} - X_t| \leq \varepsilon$ for all $0 \leq s \leq \delta$, then X_t is a continuous function of t. This is a question about a hitting time like (12).

The hitting time τ_a is a random variable. We will use the heat equation to find the probability density $\tau_a \sim f_a(t)$. There is an explicit formula for $f_a(t)$. The derivation gives some idea how PDF (partial differential equation) methods are used in the study of Brownian motion and other diffusion processes.

One way to find f_a is to first find the probability density $u_a(x,t)$ of for a *stopped* Brownian motion process. For the rest of this section we will assume that $X_0 = 0$ and $a < 0$. We will see that it is natural to call a the *absorbing boundary*. Imagine a particle that executes Brownian motion until the first time it touches the absorbing boundary. After that you can imagine that the particle is removed or absorbed at the boundary. It help to imagine a large cloud of particles doing independent Brownian motion. Any particle that touches the absorbing boundary is immediately eaten. Then $u_a(x,t)$ is the density of *surviving* particles, which means particles that have not touched the boundary yet. The formal definition is

$$u_a(x,t)dx = \Pr(x \leq X_t \leq x + dx \text{ and } \tau_a > t) \ . \tag{13}$$

We find u_a by solving a PDE, which relies on three facts about u_a.

1. u_a satisfies the heat equation in the domain where $u > 0$, which is $x > a$.
2. $u_a(x,t)$ satisfies the *initial condition* $u_a(x,0) = \delta(x)$.
3. u_a satisfies the *boundary condition* at $x = a$, which is $u_a(a,t) = 0$.

Two of these conditions are easy to understand. u_a satisfies the heat equation for $x > a$ because (10) is approximately true at any point $x > a$ if $t - s$ is small enough. For any point $x > a$ you can take $t - s$ so small that in that interval of time there is almost no chance that a Brownian motion particle at time s will touch the boundary before time t. We will give more mathematical justification later. The initial condition $u_a(x,0) = \delta(x)$ just means that $X_0 = 0$.

The boundary condition is more subtle. It says not only that particles are eaten (absorbed) at the boundary, but that there are very few surviving particles close to the boundary – the density of surviving particles goes to zero as you approach the absorbing boundary. This is based on the following fact about Brownian motion, which we will come to understand better as the class progresses. If X_t is close to a, then it is very likely that $\tau_a < t$. In other words, if a Brownian particle is close to the boundary, then it is likely to have touched the boundary in the recent past. We put this into formulas using the conditional probability

$$P(t,x) = \Pr(\ X_s > a \quad \text{for } 0 \leq s \leq t \ \mid X_t = x) \ .$$

Clearly $u_a(x,t) = P(x,t)u(x,t)$. It is a property of Brownian motion (motivated but not strictly proved) that $P(x,t) \to 0$ as $x \to a$. Therefore, $u_a(x,t) \to 0$ as $x \to a$.

It is possible to find an explicit formula for u_a. The derivation has two ingredients, a *uniqueness* theorem and the *method of images* trick. *Uniqueness* means that you have given so many conditions that there is only one function that satisfies all of them. The uniqueness theorem for the heat equation states that if, say, $v(x,t)$ is bounded (as a function of x) for every t, and if v is differentiable enough for the PDE to make strict sense, and if it satisfies the three conditions above, then $v(x,t) = u_a(x,t)$. We will not give a proof, but the Courant graduate class on PDE does. The uniqueness theorem has the following consequence: If you can write a formula that satisfies the three conditions, then that formula is the solution. It is the solution even if you can't explain how you found the formula. If you can check that the formula works, it's the solution.

The *method of images* produces a such a formula. The formula is for a function, also called $u_a(x,t)$, that is defined for all x and $t > 0$. This formula satisfies the heat equation for all x and $t > 0$. It satisfies the initial condition only for $x > a$, which is enough. It satisfies the boundary condition $u_a(a,t) = 0$ because it is anti-symmetric about the boundary $x = a$:

$$u_a(2a - x, t) = -u_a(x,t) \; . \tag{14}$$

If $x > a$ is any point in the domain where the original u_a is defined, then $x' = 2a - x$ is the *image* point of x reflected about the point a. That means that $|x' - a| = |x - a|$, and $x' < a$. A function that is skew symmetric in the sense of (14) must be equal to zero at the point of symmetry: $u_a(a,t) = 0$.

The method of images works because of the symmetry of the Brownian motion diffusion process. The initial condition for u_a will be the desired "point charge" (delta function) at $x = 0$, and an equal strength but opposite charge at the image point $2a$:

$$u_a(x,0) = \delta(x) - \delta(x - 2a) \; .$$

This satisfies the initial condition 2 above because that condition applies only when $x > a$. The image "charge" at $x = 2a$ is outside the domain. The solution with this initial data is the sum of the positive fundamental solution centered on $x = 0$ and the negative fundamental solution centered at $x = 2a$:

$$u_a(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} - \frac{1}{\sqrt{2\pi t}} e^{-\frac{(2a-x)^2}{2t}} \; . \tag{15}$$

This satisfies the heat equation (7) because both terms to. It satisfies the boundary condition, because .. well .. just put in $x = a$.

The solution formula (15) carries much information about Brownian motion hitting times. In particular it leads to a formula $f_a(t)$, the PDF of the hitting time τ_a. We already saw that

$$\Pr(\, \tau_a > t) = \int_a^\infty u_a(x,t)\, dx \; .$$

This follows from the definition (13) by summing over all $x > a$. The hitting time PDF is related to this by

$$f_a(t) = -\frac{d}{dt} \Pr(\, \tau_a > t) \; .$$

(Check signs: the probability on the right is a decreasing function of t, so the minus sign on the right makes the left side positive.) We carry out the differentiation using the heat equation (7)

$$\begin{aligned} f_a(t) &= - - \frac{d}{dt} \int_a^\infty u_a(x,t)\, dx \\ &= - \int_a^\infty \partial_t u_a(x,t)\, dx \\ &= -\frac{1}{2} \int_a^\infty \partial_x^2 u_a(x,t)\, dx \\ f_a(t) &= \frac{1}{2} \partial_x u_a(a,t) \; . \end{aligned} \tag{16}$$

(Check signs: u_a is going from 0 to positive values as x increases from a, so the right side is positive.) If we differentiate the explicit formula (15), the result is

$$f_a(t) = \frac{1}{\sqrt{2\pi}} \frac{-a}{t^{3/2}} \, e^{\frac{-a^2}{2t}} \; . \tag{17}$$

(Check sign: $a < 0$ and the rest is positive, so $f_a > 0$.)

The explicit formula says some interesting things about $f_a(t)$. One is that $f_a(t)$ is always positive, so particles can reach the boundary from far away arbitrarily quickly. But this is unlikely, as $f_a(t)$ is exponentially small for small t. The time when X_t is most likely to hit is the maximum of $f_a(t)$ over t, which is found by differentiation:

$$\partial_t f_a(t) = \left[\frac{3}{2t} - \frac{a^2}{2t^2} \right] f_a(t) = \frac{1}{2t} \left[3 - \frac{a^2}{t} \right] f_a(t)$$

The result is $t = \frac{1}{3} a^2$. This illustrates the natural scaling of Brownian motion, t scales as a^2, or length (a) scales like the square root of time. How far from the boundary do you have to start a Brownian motion so that t is the most likely hitting time? The answer is on the order of $\sqrt{t}$.

An important feature of the formula (17) concerns it's "tail" behavior, how it goes to zero as $t \to \infty$. In that limit, the exponential factor goes to 1, so

$$f_a(t) \approx \frac{1}{\sqrt{2\pi}} \frac{-a}{t^{3/2}} \; .$$

This *power law tail* indicates that $f_a(t)$ goes to zero slowly as $t \to \infty$. It implies that there is a rather large chance of having a rather large τ. Large τ is so likely that the expected value

$$\mathrm{E}[\, \tau_a] = \frac{-a}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{1}{2}} \, e^{\frac{-a^2}{2t}} \, dt = \infty \; .$$

The large t behavior of the integrand is $t^{-\frac{1}{2}}$, which is not integrable. The hitting time is positive random variable that is finite (almost surely) and has infinite expected value.

4.1 Martingales, stopping times, fat tails, the gambler's ruin paradox, the dominated convergence theorem

Brownian motion and stopping times illustrate many things that will concern us in the rest of the class. This subsection quickly describes some of them.

A process Y_t is a *martingale* if, for any $s > t$, $E[|Y_t|] < \infty$ and

$$E[\, Y_s \mid \mathcal{F}_t] = Y_t \; . \tag{18}$$

A Brownian motion should be a martingale, because X_s is X_t plus some independent mean zero noise. Many processes are martingales without having the independent increments property. In fact, you could call (18) the *uncorrelated increments* property. The increment, $Y_s - Y_t$, is uncorrelated with any function of $Y_{[0,t]}$. Being uncorrelated does not imply being independent. As a trivial example in discrete time, suppose $Z_n \sim \mathcal{N}(0,1)$ is an i.i.d. family of standard normals, $\mathcal{F}_n$ is generated by $Z_{[1:n]}$, and

$$Y_{n+1} = Y_n + Z_n Z_{n+1} \; .$$

Then $E[\, Y_{n+1} \mid \mathcal{F}_n] = Y_n$, but Z_n is known at time n and the difference $Y_{n+1} - Y_n = Z_{n+1} Z_n$ is not independent of Z_n.

A positive random variable τ is a *stopping time* if the event $\{\tau \leq t\}$ is known at time t. More precisely, if $A_t = \{\tau \leq t\}$, then $A_t \in \mathcal{F}_t$. The idea is that you watch the process Y_t, and based on the path $Y_{[0,t]}$, you decide whether to stop at time t. Hitting times are examples of stopping times. Another example might be

$$\tau_a = \min \left\{ t \mid \int_0^t X_s \, ds = a \right\} \; .$$

This one depends on the whole path $X_{[0,t]}$, not just its value at time t. A stopping time may not rely on knowing the future. It is rare that the last time something happens is a stopping time, but it can be:

$$\tau_a = \max \left\{ t \mid \int_0^t X_s^2 \, ds = a \right\} \; .$$

Here, $X_t^2 > 0$ (almost surely), so the last time is also the first time. More typical example (sort of contrived):

$$\tau_a = \max \left\{ t \mid \int_0^t \left(X_s^2 - \sqrt{(s)} \right) \, ds = a \right\} \; .$$

If the integral is equal to a at time t, we don't know whether the integral will increase and then decrease back to a in the future.

If Y_t is a stochastic process and τ is a stopping time, the *stopped* process is the one that stops moving at time τ. The wedge notation is an efficient way to say this: $a \wedge b = \min(a, b)$. The stopped process has the value Y_t if $t < \tau$ and the value Y_τ if $t > \tau$:

$$\widetilde{Y}_t = Y_{t \wedge \tau} \; .$$

The *Doob stopping time theorem* states that if Y_t is a martingale and if τ is a stopping time, then the stopped process $\widetilde{Y}_t$ also a martingale. This will become obvious, particularly in the discrete time case, if you think about it for some while. The proof in the discrete time case is simple. The continuous time proof is more technical but isn't really different.

The stoping time theorem has a finance interpretation. The martingale Y_t is thought of as the fluctuating price of some asset. The martingale property says that you cannot make an expected profit by buying the asset at time t and selling at time $s > t$. The expected price at time s is the purchase price Y_t, see (18). The stopping time is a possibly more sophisticated strategy. You buy at time t then sell at time τ. You have to know at time s whether to sell at time s, because if you sell at a different time you don't get the price Y_s. The stopping time theorem says that even this larger family of strategies cannot produce an expected profit. If $\tau > t$ and $s > t$, then

$$E[\, Y_{\tau \wedge s} | \mathcal{F}_t] = Y_t \ . \tag{19}$$

If you give a final "give up" time, s, then you cannot make an expected profit.

The gambler's ruin paradox concerns what happens if you leave out the give up time s in (19). Let X_t be a Brownian motion starting at $X_0 = 0$ and let τ be the the first hitting time when $X_t = 1$. We know that $\tau < 0$ almost surely, and that $E[\tau] = \infty$. Clearly $X_\tau = 1$. We make an expected profit $X_{|tau} - X_0$ if we buy at time 0 and sell at time τ. The