Stochastic Calculus, Courant Institute, Fall 2018
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2018/index.html
Jonathan Goodman, September, 2018

Lesson 1, Brownian motion

1 Introduction to the course

These "Lessons" class notes for the Stochastic Calculus class of Fall, 2018. They contain the material from the lecture, and probably a little more. You will probably need to read the "lessons" to do the assignments.

This class uses the term *stochastic calculus* in two senses. In one sense, *stochastic calculus* refers to a set of tricks for calculating things related to random processes. One such trick is using the recursive backward equation to calculate expected values. Most of the information we have about stochastic processes comes from calculations like these. A clever proof usually relies on a clever calculation.

In another sense, *stochastic calculus* refers to the *Ito calculus* and related topics. The basic operations of ordinary differential and integral calculus may not work when applied to *diffusion processes* because they are not differentiable. The chain rule for diffusion processes, which is called *Ito's lemma*, requires you to calculate to second order in Taylor series in order to get the first "Ito derivative".

Stochastic calculus is more than a collection of mathematical facts. It is also a framework for creating mathematical models of physical or economic random processes. Most of these models involve simplifications and approximations.[1] For example, the Black Scholes equation is derived from a model in which stock trading takes place in continuous time and the stock price is a continuous function of time. Actual stock trades happen each millisecond (not continuously) and stocks "tick" up or down by small but non-zero jumps. Brownian motion itself is a simple model of a complex physical process of a small particle in water interacting with a large number of even smaller water molecules. The Brownian motion model is accurate on "coarse" time scales (larger intervals of time) but not on the time scale of individual collisions between the particle and water molecules. This is not specific to stochastic calculus. Newton's laws of planetary motion neglect special relativity, quantum mechanics, magnetic fields, etc. Nevertheless, they are useful for modeling the motion of planets.

The class was originally created for the Mathematics in Finance program, but it was always meant to be generic and useful to others with different applications in mind. Many of the examples are not from finance. Still, the choice of topics was influenced by financial applications. In particular, there is less about steady states and correlation functions than a course aimed at physics or chemistry students would have.

[1] The statistician George Box was commenting on the modeling process when he said: "All models are wrong. Some models are useful."

The reasoning in this class isn't rigorous in the pure math sense, but it is serious. Someone with the right background in measure theory would be able to make many of the arguments rigorous if she or he were interested. My wish for the class is to add as much "value" to students as possible in 13 lectures. That means sacrificing proofs to make time for applications.

Computing is an essential part of present and future applied mathematics. Since this class is applied mathematics, it would be wrong to do it without computing. In fact, the computational methods – simulation and PDE solving, etc. – are core elements of modern stochastic calculus.

2 Introduction to Brownian motion

Brownian motion is the name of the phenomenon that small particles in water, when you look at them with a powerful enough microscope, seem to move in a random fashion. It is named after a Brit named Brown, but the Wikipedia page suggests that it was first observed elsewhere (France?). It also is the name of a mathematical model of this particle motion. In the Brownian motion model, X_t is the location of a randomly moving particle at time t. The *path* is written as just X. The value of the path, the location, is X_t. We use subscripts instead of function notation, $X(t)$, because, well, because probability people do.

The class starts with Brownian motion because it's the simplest example of a *diffusion process*, and diffusion processes are the main topic of the class. Many of the properties of diffusion processes can be seen in Brownian motion first and then generalized to more general processes. For example, the backward and forward equations for Brownian motion are special cases of the backward and forward equations for general diffusions.

The *central limit theorem* is behind the fact that Brownian motion is a model for the random motion of small particles, and for many other random processes. You can view the motion the the particle as the result of a large number of small and independent steps. The position X_t is thought of as the result of a large number (an infinite number in the *Brownian motion limit*) of small independent steps. The sum of a large number of independent identically distributed steps, according to the central limit theorem, is approximately Gaussian. The Brownian motion limit produces X_t that is exactly Gaussian.

But the Brownian motion limit is about more than the distribution of X_t. It's about other properties of the whole Brownian motion path. For example, is is about the *hitting probability*

$$\Pr(\,|X_t| \geq R \text{ for some } t < T) \ . \tag{1}$$

There is a path version of the central limit theorem, called the *Donsker invariance principle*.[2] The invariance principle says that you can estimate probabilities like (1) for complicated physical processes like the physical Brownian particle motion using the simple mathematical Brownian motion model.

[2]Monroe Donsker was a Courant Institute mathematician, a great mathematician, and an interesting person.

Finally, Brownian motion serves as a model of the random noise that "drives" other diffusion processes. This allows us to express general diffusions as functions of Brownian motion. The Ito calculus, developed several lessons from now, is the tool for doing this. Brownian motion is used in computer simulation of general diffusions through what is called the *Euler Mayurama* method.

Brownian motion is a random function of time. The position of a particle at time t is X_t. We suppose $X_0 = 0$ and model the motion for $t > 0$. The defining properties of Brownian motion are

1. X_t is a continuous function of t.

2. The *increment* $X_{t_2} - X_{t_1}$ is Gaussian with mean zero and variance $t_2 - t_1$ ($t_2 > t_1$ for this to make sense).

3. X is a *Markov process*, which means that conditional on X_t, the *future* (X_s, with $s > t$)is independent of the past (X_s with $s < t$).

The Markov property of the mathematical Brownian motion reflects the fact that the increments of Brownian motion after time t_1 are the result of small steps after time t_1 that are independent of whatever happened before t_1. The random forces moving the particle after t_1 are independent of the forces that moved it before t_1. The increment variance formula (2.) depends only on the time difference. The Brownian motion model is statistically homogeneous in time in the sense that the distribution of the random increment doesn't depend on X_{t_1} or t_1, but only on the amount of time in the increment. From a microscopic point of view, this reflects the idea that whatever in the environment that is causing X_t to move is homogeneous in time. The physical Brownian particle moving in a fluid is like this (to some degree of approximation).

A *stochastic process* (also called *random process*) is a function of t whose values are random. Brownian motion is a stochastic process. A random variable with a specific distribution may be called a *sample* of the distribution. A sample of a stochastic process may be called a sample *path*. A *diffusion process* is a random process that is a Markov process and has continuous sample paths. Brownian motion is the central and most basic example of a diffusion process. Other diffusion processes have non-Gaussian increments, or Gaussian increments with non-zero mean.

Brownian motion is important for many reasons, among them

1. It is a good model for many physical processes.

2. It illustrates the properties of general diffusion processes.

3. It can be used to construct other diffusion processes through the Ito calculus.

This first lesson focuses on Brownian motion itself, with some basic motivation and properties. One important point is some things about Brownian motion that can be calculated either directly or with the help of partial differential equations. Another point is the relation between Brownian motion and random walk, which may be seen as a fancy version of the central limit theorem. This relation motivates properties 1, 2 and 3. It also suggests computing methods that give approximate solutions to some partial differential equations related to Brownian motion.

3 Transition probabilities and value functions

The Brownian motion path is too complicated to be described by a single probability density. But there are useful probability densities for simpler quantities related to the Brownian motion path. These densities do not describe the whole path. They are densities of some simple functions of a Brownian motion path. For example, X_t is the position at a specific time t. We denote the PDF of X_t by $u(x,t)$. There is a simple Gaussian formula for $u(x,t)$, which comes from the fact that $X_t - X_0$ is the increment of Brownian motion for the time interval ending at $t_2 = t$ and starting at $t_1 = 0$. The increment is equal to X_t because $X_0 = 0$. It (by property 2) is Gaussian with mean zero and variance t. This is[3]

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

This probability density describes the probability density at time t but it says little about the path X_s for $0 < s < t$.

Properties 2 and 3 lead to formulas for the joint density of several *observations* of the Brownian motion path at several times. To start, suppose $0 < t_1 < t_2$. Write X_1 for X_{t_1}, etc. We want the joint density function $u_2(x_1, x_2, t_1, t_2)$, which is defined by

$$\begin{aligned} &u_2(x_1, x_2, t_1, t_2)\, dx_1 dx_2 \\ &\qquad = \Pr(\, x_1 \le X_1 \le x_1 + dx_1 \text{ and } x_2 \le X_2 \le x_2 + dx_2) \ . \end{aligned}$$

The expression for u_2 comes from the density of X_1 and the conditional density of X_2 given X_1. The PDF for X_1 is (2) with $x = x_1$. The conditional probability of X_2 given X_1 is given by (not writing t_1 and t_2 to shorten the formulas)

$$u(x_2|x_1)\, dx_2 = \Pr(\, x_2 \le X_2 \le x_2 + dx_2 \mid X_1 = x_1) \ .$$

Property 2 implies that this is Gaussian with mean x_1 and variance $t_2 - t_1$. Thus

$$u(x_2|x_1) = \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{1}{2(t_2-t_1)}(x_2-x_1)^2} \ . \tag{3}$$

This is called the *transition density* of Brownian motion because it describes the probability density of transitions from x_1 at time t_1 to x_2 at time t_2.

[3] The Gaussian density with mean μ and variance σ^2 is $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. Here, we have $\mu = 0$ and $\sigma^2 = t$.

Bayes' rule for this context is

$$u_2(x_1, x_2) = u(x_1)u(x_2|x_1) \ .$$

Specifically, we get (with some algebra)

$$u_2(x_1, x_2, t_1, t_2) = \frac{1}{\sqrt{(2\pi)^2 (t_2 - t_1)t_1}} e^{-\frac{1}{2}\left[\frac{(x_2-x_1)^2}{t_2-t_1} + \frac{x_1^2}{t_2}\right]} \ . \tag{4}$$

The formula for the joint density of three or more observations is analogous, but its derivation requires the Markov property 3.

Suppose we want to know the expected value of some function of the Brownian motion position at time T:

$$f_0 = \mathrm{E}[\, V(X_T)] \ . \tag{5}$$

The *value function* is the conditional expectation of $V(X_T)$, conditional on the location at an earlier time

$$f(x,t) = \mathrm{E}[\, V(X_T) \mid X_t = x] \ . \tag{6}$$

This is often written

$$f(x,t) = \mathrm{E}_{x,t}[\, V(X_T)] \ .$$

The subscript on the expectation indicates which "probability measure" is used for the expectation. Future lessons will say more about probability measure, but without giving all the mathematical details. Either the conditional probability formula (3), or the property 2 that it comes from, gives (writing x for x_2, y for x_1, T for t_2 and t for t_1)

$$f(x,t) = \frac{1}{\sqrt{2\pi(T-t)}} \int_{-\infty}^{\infty} V(x) e^{-\frac{1}{2}\frac{(x-y)^2}{T-t}} \, dy \ . \tag{7}$$

In more complicated problems, the expectation (5) is calculated by solving the partial differential equation that the value function satisfies. The value function gives f_0 because $f_0 = f(0,T)$.

4 Partial differential equations

The probability density (2) satisfies a partial differential equation, or *PDE*, that is called (depending on the context and the generality) the *heat equation* or the *diffusion equation* or the *Kolmogorov forward equation.* We write partial derivatives as

$$\frac{\partial u}{\partial t} = \partial_t u \ , \ \ \frac{\partial^2 u}{\partial x^2} = \partial_x^2 u \ , \ \text{etc} \ .$$

This is simpler, and it emphasizes differentiation as an *operator* that can be applied to a function. Direct calculation with (2) shows that

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

Later on there will be derivations of this equation that are more satisfying than direct calculation.

The heat equation (8) is more general than the Gaussian probability density even though many derivations rely on something being Gaussian. For example, suppose the distribution of X_0 is random with PDF $u_0(x_0)$. Suppose (as will be justified more below) also that the transition density from time 0 to time t is the Gaussian so that a transition $y \to x$ has the PDF [4]

$$G(x, y, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} \; . \tag{9}$$

Then, as before (only the notation has changed, and that only a little) Bayes' rule gives the PDF for X_t as

$$u(x, t) = \int_{-\infty}^{\infty} u_0(y) G(x, y, t) \, dy \; . \tag{10}$$

This is the integral over places X may have started, called y with PDF $u_0(y)$, multiplied by the probability density to go from y to x in time t, which is (9). The more general probability density $u(x, t)$ in (10) is Gaussian only if u_0 is Gaussian.[5] But u still satisfies the heat equation (8). You can see this by differentiating the integral on the right with respect to t and (twice) with respect to x. The partial derivatives go under the integral and land on G. Direct calculation (the same one as before) shows that G satisfies the heat equation. To summarize: the probability density of the Brownian motion location X_t satisfies the heat equation (8).

The value function (6) satisfies a PDE that is similar but has an important difference in sign. The partial derivatives $\partial_x^2 f$ and $\partial_t f$ may be calculated from the integral representation (7). Notice that (7) may be written using (9) as

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

The sign difference, relative to (8), comes from

$$\partial_t G(x, y, T - t) = -\frac{1}{2} \partial_x^2 G(x, y, T - t) \; .$$

The result is commonly written in the form

$$\partial_t f(x, t) + \frac{1}{2} \partial_x^2 f(x, t) = 0 \; . \tag{11}$$

This is commonly called the *backward equation* (or *Kolmogorov* backward equation) to indicate that the unknown f is a value function and that the sign is different. There is another derivation of this PDE that is more general and

[4]The G is for both Green and Gauss. In the theory of partial differential equations, functions that play the role of G are called *Green's functions*.

[5]This intuitive fact has a mathematical proof using the Fourier transform.

does not rely on algebraic calculation like this. The meaning of "backward" and "forward" will be explained soon.

There is a *duality* relation between the forward and backward equations. For any t in the range $0 < t < T$, the expected value (5) may be expressed in terms of the conditional expected values (6) as

$$f_0 = \int_{-\infty}^{\infty} u(x,t) f(x,t)\, dx \; . \tag{12}$$

Formulas like this are sometimes called the *law of total probability.* The expected value on the left is expressed in terms of conditional expectations on the right and all possible values of X_t, weighted by their probability density. The duality relation is that the forward equation implies the backward equation and conversely. Here, we give the calculation to show that the forward and backward equations are consistent. The proof that one implies the other is more technical.

The calculation relies on the fact that f_0 does not depend on t. If we differentiate with respect to t, the left side is zero. On the right side we may take ∂_t inside the integral and apply the product rule. The result is

$$\int_{-\infty}^{\infty} \left[\partial_t u(x,t)\right] f(x,t)\, dx + \int_{-\infty}^{\infty} u(x,t) \left[\partial_t f(x,t)\right]\, dx = 0 \; .$$

Suppose u satisfies the forward equation (8). Then this becomes

$$\frac{1}{2}\int_{-\infty}^{\infty} \left[\partial_x^2 u(x,t)\right] f(x,t)\, dx + \int_{-\infty}^{\infty} u(x,t) \left[\partial_t f(x,t)\right]\, dx = 0 \; .$$

We may integrate by parts twice in the first integral then combine terms. The result is

$$\int_{-\infty}^{\infty} u(x,t) \left[\partial_t f(x,t) + \frac{1}{2}\partial_x^2 f(x,t)\right]\, dx = 0 \; .$$

The term $[\cdots]$ is zero if f satisfies the backward equation (11). A mathematical proof would have to ask several questions: What about the boundary terms in the integration by parts? How do we know that $\partial_x^2 f(x,t)$ exists? These questions will be answered to some extent later in the course. It is "hard to imagine" that this integral is always zero even though $[\cdots] \neq 0$. Unfortunately, "hard to imagine" is not a mathematical proof. But it is related to one. In this course, we will derive a backward equation and then use the duality relation (12) to derive the corresponding forward equation.

The u equation (8) is called *forward* because it describes how u changes (evolves) as time moves forward. The *initial data* $u_0(x)$ are specified in a somewhat arbitrary way. Then the equation (8) describes how u changes as t increases. The solution of (8) with initial data u_0 is given by the integral representation (10). The expression (9) makes sense only if t has moved in the forward direction from 0 to $t > 0$. If you want to specify $u(x,t)$ for $t > 0$ completely, it suffices to say that u satisfies the PDE (8) and has initial data $u(x,0) = u_0(x)$. The problem of finding u for $t > 0$ is called the *initial value* *problem*, because the initial values $u_0(x)$ must also be given. It is a theorem that the solution of the initial value problem is unique – there is only one. We showed that the integral (10) satisfies the PDE. It is not hard to see (exercise) that this formula has $u(x,t) \to u_0(x)$ as $t \to 0$, so the initial condition is satisfied. Therefore, the integral represents the unique solution.

The equation (11) is called "backward" for a similar reason. The value function f satisfies *final conditions* $f(x,T) = V(x)$. This is "obvious" from the abstract definition (6). If we know $X_T = x$, then the expected value is irrelevant, you just get $V(x)$. The solution to the *final value problem* is also unique, so the formula (7) represents this unique solution. It is necessary that $t < T$. The value function evolves backwards in time from its given final condition. In the integral, $G(x,y,T-t)$ makes sense only if $t < T$.

5 Hitting probabilities

Suppose $a > 0$ is some "marker". The *hitting time* for a is the first time X "hits" a. This is written

$$\tau_a = \min\{\, t \mid X_t = a \} \ . \tag{13}$$

Hitting times may be defined for other processes. It is possible, for other processes, that there is no hitting time because X never hits a. We define $\tau_a = \infty$ in those cases. We will see that for Brownian motion,

$$\Pr(\, \tau_a = \infty) = 0 \ .$$

This means that the hitting probability is one. Philosophers argue whether this means that the hitting time is "always" finite. Probabilists say *almost surely* for events that have probability one.

Hitting probabilities may be calculated using either the forward or the backward equation. The unknown in the equation is defined to capture the hitting event. The forward equation represents probability density, which is defined in terms of $\Pr(\, X_t \in (x, x+dx))$. For hitting problems, we refine this to require that X_t has not hit the boundary before time t

$$u(x,t)dx = \Pr(\, X_t \in (x, x+dx) \text{ and } X_s \neq s \text{ for } 0 \leq a \leq t) \ . \tag{14}$$

This definition is applied only for $x < a$. The hitting probability is one minus the probability of not hitting, so

$$1 - \Pr(\, \tau_a < t) = \Pr(\, \tau_a > t) = \int_{-\infty}^{a} u(x,t)dx < 1 \ . \tag{15}$$

The unknown u may be calculated by solving a partial differential equation with a *boundary condition* at $x = a$.

I hope the following facts are easy to believe. We do not prove them rigorously in this class but we give more reasons to believe them in future lessons.

1. For $x < a$, the density (14) satisfies the forward equation (8). This is because X_t is continuous so if $X_t < a$, then for times close enough to t, the process does not "know about" the boundary a so it "looks like" a regular Brownian motion.

2. As $t \to 0$, and for $x < a$, $u(x,t)$ "looks like" the unconditional density (2). The reason is the same – X_t does not feel the boundary for small enough time.

3. $u(x,t) \to 0$ as $x \to a$. Since u is continuous for $t > 0$ at $x = a$ (this is as hard to prove as the other properties), this is the same as $u(a,t) = 0$.

It is possible to find a formula for $u(x,t)$ from these facts.

The *method of images* is a trick for finding u. The trick is to find something that satisfies the forward equation for $x \leq a$ by subtracting from the "free" solution (2) another function that satisfies the forward equation. The free solution may be thought of as the result of starting with a unit of probability at the point $x = 0$. This is called a unit *charge* of probability starting at $x = 0$. We subtract a negative unit charge starting at $x = 2a$. This is the *image charge* that gives the method its name. The image charge solution is

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

Subtraction gives

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

Let us check that this formula has the three properties above. For $t > 0$ this satisfies the forward equation, because the forward equation is linear and both parts satisfy it. As $t \to 0$, and for $x < a$, the solution looks like the free solution. This is because the image part goes to zero very rapidly for $x \leq a$ as $t \to 0$. If you fix $x \leq a$, then the exponential part of the image term is $e^{-\frac{-(x-2a)^2}{2t}} \leq e^{-\frac{-a^2}{2t}}$ (because $(x-2a)^2 \geq a^2$ for $x \leq a$). Therefore the exponential part goes to zero "exponentially" as $t \to 0$. The *prefactor*, which is $(2\pi t)^{-\frac{1}{2}}$, "blows up" (goes to infinity) as $t \to 0$, but it blows up more slowly than the exponential goes to zero. As a result, the whole image term goes to zero as $t \to 0$ for $x \leq a$. This is not true for $x = 2a$, where the image part blows up as $t \to 0$. Fortunately, $x = 2a$ is not in the domain where we care what $u(x,t)$ is.

The *boundary condition* $u(a,t) = 0$ is satisfied by the combination (16) by symmetry. It is easy to check that if you take $x = a$ then the two exponential terms cancel. The symmetry is that $u(x,t)$ takes "equal and opposite" values at equal distances from $x = a$. The two symmetric points are x and $2a - x$ (draw a picture). The point $2a - x$ is the *image point* for x, reflected about the point a The symmetry property under reflection is

$$u(x,t) = -u(2a - x, t) \ .$$

The negative image charge is placed at $2a$ to insure this symmetry. If you put in $x = a$, you get $u(x,t) = -u(a,t)$, so $u(a,t) = 0$ (if u is continuous). This is the property 3 above. This shows that the formula (16) is the probability density we're looking for.

The formulas (16) and (15) give a way to calculate the probability density of the hitting time τ_a. Call this density $v_a(t)$. It is defined by

$$\Pr(t \leq \tau_a \leq t + dt) = v_a(t)dt \ .$$

It is found from

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

(Check the sign: v_a is positive and the survival probability $\Pr(\,\tau_a > t)$ is a decreasing function of t.) We could substitute the explicit formula (16) into this and calculate, but there is less calculation if we instead use the forward equation that u satisfies:

$$\begin{aligned}
\frac{d}{dt}\Pr(\,\tau_a > t) &= \frac{d}{dt}\int_{-\infty}^{a} u(x,t)dx \\
&= \int_{-\infty}^{a} \partial_t u(x,t)dx \\
&= \frac{1}{2}\int_{-\infty}^{a} \partial_x^2 u(x,t)dx \\
&= \frac{1}{2}\partial_x u(a,t) \ .
\end{aligned}$$

In the last line, the boundary term at $x = -\infty$ is zero because $\partial_x u(x,t) \to 0$ as $x \to -\infty$, which you can believe based on the fact that it's very unlikely for x to have a large negative value, or you can check in the formula (16). The derivative at the end is calculated by

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

Altogether, we find

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

This formula says a lot about hitting.

When $t \to 0$, the exponential factor dominates the power law prefactor. This shows that it's "exponentially" unlikely to hit the boundary quickly. You can find the "most likely" hitting time by finding the maximum of v_a over t. The result is $t_* = \frac{1}{3}a^2$. Since a is the distance from the starting point, this says that the time to reach a point is roughly proportional to the square of the distance. This is the basic *scaling* of Brownian motion – "time goes like space squared". When $t \to \infty$, the exponential factor goes to 1, so only the prefactor matters. That is $v_a(t) \approx \frac{1}{\sqrt{2\pi}} t^{-\frac{3}{2}}$. The *tail* of a probability distribution is the part far from the most likely values. A probability distribution has *heavy tails* if the density goes to zero slowly and *light tails* if it goes to zero rapidly. The Gaussian density has light tails, since the density goes to zero exponentially. The hitting time density (17) has heavy tails because the density has a "power law" decay with power (the power of t) $-\frac{3}{2}$. This means that it isn't so unlikely for τ_a to be very big. It's unlikely, but not very unlikely.