Brownian Motion

Jianing Yao
Department of MSIS-RUTCOR
Rutgers University, the State University of New Jersey
Piscataway, NJ 08854 USA

February 27, 2015

Note: Remarks and section with $*$ are beyond the scope of the class. Interested readers can read and discuss with me if you want to know more about stochastic analysis.

1 Continuous-time stochastic process; generalities and terminology

1.1 Basics

Recall that a *stochastic process* is a collection of random variables,

$$\mathbf{X} = \{X(t); t \in \mathcal{J}\}$$

defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here, $\mathcal{J}$ is the index set used to label the elements of $\mathbf{X}$, it can be any, non-empty set. We restrict our attention to real-valued stochastic process, although it can be more general than that. Thus, for each $t \in \mathcal{J}$, $X(t)$ is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$[1]. In other words, $X(t)$ is a random variable in the usual sense: $X(t)^{-1}(U) \in \mathcal{F}$ for any $U \in \mathcal{B}(\mathbb{R}^n)$.

The index t will generally serve as a time parameter, when $\mathcal{J}$ is a subset of consecutive integers, $\mathbf{X}$ is a *discrete stochastic process*, such as random walk. We will be mainly studying rather the continuous-time processes for the rest of the semester, that is, $\mathcal{J}$ is a sub-interval of $\mathbb{R}$. Usually, $\mathcal{J}$ will stand for one of the interval, $[0, T)$, $[0, T]$ or $[0, \infty)$[2]. One thing to always keep in mind is that $X(t)$ is a function on Ω; we shall write $X(t)(\omega)$ when it is necessary to indicate its dependence on ω explicitly.

A stochastic process $\mathbf{X}$ maybe also be viewed as a joint function,

$$(t, \omega) \mapsto X(t)(\omega), \ \forall (t, \omega) \in \mathcal{J} \times \Omega$$

[1] $\mathcal{B}(\mathbb{R}^n)$ is the *Borel-σ*-algebra of $\mathbb{R}^n$.

[2] There is no significance to choosing $t = 0$ as left-endpoint of $\mathcal{J}$ beyond its convenience for representing the initial time

The definition of stochastic process requires only that $X(t)$ be a random variable for each separate t, which is not adequate for analysing its joint dependence on t and ω. For this, a stronger property is needed. We say a process $\mathbf{X}$ is *measurable* if, as a joint function of (t, ω), it is measurable with respect to the product σ-algebra, $\mathcal{B}(\mathbb{R}) \times \mathcal{F}$ (If you don't feel comfortable with this notion, you can temporarily skip it. Because, in practice, it will not be necessary to assume measurability ad hoc, because it follows either from an explicit construction or from the imposition of other properties).

For a process $textbf X$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and for a **given** $\omega \in \Omega$, consider the function, $t \mapsto X(t)(\omega)$, defined on $\mathcal{J}$. It is called a *sample path* of $\mathbf{X}$ and is denoted by $X(\cdot)(\omega)$(see in *figure* 1). We defined $\mathbf{X}$ as a collection of random variables, but, in many ways, it is more

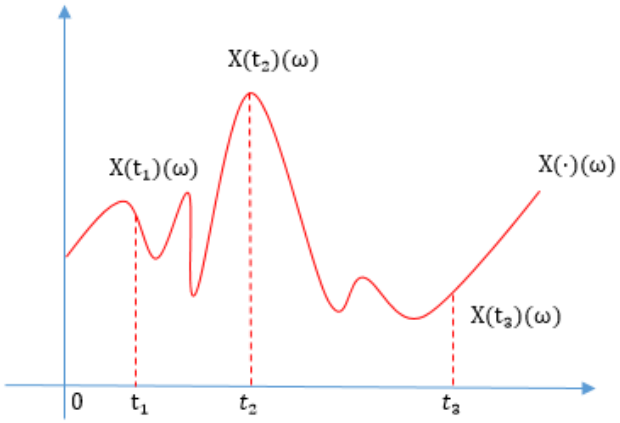

Figure 1: Sample Path

natural to think of it as a random path, that is, as a map, $\omega \mapsto X(\cdot)(\omega)$, from a probability space into a space of sample paths.

Remark 1.1 Pursuing this viewpoint leads to the important notion of *canonical process*. Suppose we want to model a random process we observe in the world. We could do this directly with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If we view an outcome of the process as a path it follows, Ω, the outcome space, should be defined as a set of possible process paths, that is, as a collection of functions on an appropriate index set $\mathcal{J}$. Then the measure $\mathbb{P}$ in this approach assigns probabilities directly to subsets of paths. The result is a direct, *'canonical'* model of the process, without explicit recourse to random variables. However, let us now define

$$X(t)(\omega) = \omega(t), \ t \in \mathcal{J}, \omega \in \Omega$$

Then $\mathbf{X}$ is a process on Ω with a special property: the sample path $X(\cdot)(\omega)$ is ω itself. For this reason, $\mathbf{X}$ is called the *canonical process* on Ω. Clearly, when studied as a process on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbf{X}$ re-expresses the canonical probability space model in the language of random variables.

None of the definitions so far have depended on the probability measure $\mathbb{P}$; they are really just about families of measurable functions on a measurable space. Now suppose some measure $\mathbb{P}$ is given, and let $\mathbf{X}$ be a stochastic process indexed by $\mathcal{J}$. For any finite subset $\mathcal{T} = \{t_1, ..., t_n\}$ of $\mathcal{J}$, the random vector $(X(t_1), ..., X(t_n))$ induces a probability measure

$$\mathbf{P}_{\mathcal{T}}^{\mathbf{X}} = \mathbb{P}\left((X(t_1), ..., X(t_n)) \in A\right)$$

on the *Borel* subsets A of $\mathbb{R}^n$. This is called a *finite-dimensional distribution* of process $\mathbf{X}$. Really, what is important and interesting about a stochastic process is not the particular, underlying probability space on which it is defined, nor the explicit definition of $X(t)(\omega)$ as function of t and ω, but its finite-dimensional distribution. Intuitively, a full set of finite-dimensional distribution uniquely determines a process in a statistical sense.

Remark 1.2 Consider an experiment with a random, real-valued outcome. In practice, how do we model it using the language of random variables? Not by defining an actual random variable! Rather, we propose a cumulative distribution function for the outcome. For example, think about the displacement after 1 second of a particle undergoing Brownian motion. Since a Brownian displacement is the sum of many small, independent perturbations, due to random collisions between the particle and molecules in the surrounding fluid, the *Central Limit Theorem* suggests it should have a normal distribution. Only after we have determined this distribution do we build an actual random variable model for the Brownian displacement. It's a similar story with a random process. We do not set out by defining an actual stochastic process to model it, but rather, by prescribing what its finite-dimensional distributions ought to be. Again, consider Brownian motion. Its finite dimensional distribution are derived from physical assumptions. If the molecular collisions in disjoint time intervals are independent, undirected, and unvarying in their statistical behaviour, so also are the Brownian displacements they cause, and , as we will see, this assumption, together with normality of each displacement, implies finite-dimensional distributions are all jointly normal, of a special type. Hence, there is a basic question to resolve. Given a full family of finite dimensional distributions, does there actually exist a matching stochastic process? If so, how may it be constructed? The *Kolmogorov Extension Theorem* answers both these questions at once, imposing only minimal consistency assumptions on the given family of finite-dimensional distributions. We will not discuss this but just to let you know such probability measure do exist.

1.2 Filtration and Adaptedness and measurability with respect to a filtration

Assume $\mathcal{J} = [0, T)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the underlying probability space on which all processes and random variables will be defined.

Definition 1.1 A *filtration* on this space is an increasing family $\mathbb{F} = \{\mathcal{F}_t; t \in \mathcal{J}\}$ of sub-σ-algebras of $\mathcal{F}$; that is, $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s < t$.

This is a straightforward generalization of the notion of a filtration $\{\mathcal{F}_n; n = 0, 1, 2, ...\}$ in discrete time. A 4-tuple, $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$, where $\mathbb{F}$ is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *filtered probability space*.

Let $\mathbf{X}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. For each $t \in \mathcal{J}$, define

$$\mathcal{F}_t^X = \sigma(X(s), s \leq t),$$

the σ-algebra generated by $X(s)$ for $s \leq t$. This defines a filtration, which is called the *filtration generated by* $\boldsymbol{X}$ and is denoted by $\mathbb{F}^X$. Most filtrations that arise in practice are built on filtrations generated by a processes. To be more specific, let $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$, define the event

$$A = \{\omega : X_i \in U_i, i = 1, ..., n, \text{ where } U_i \in \mathcal{B}(\mathbb{R}^n)\}$$

Then, $\mathcal{F}_t$ contains all such events and all events that can be obtained from them by taking countable unions and intersections. In the context of conditioning, sub-σ-algebras represent partial information. To explain, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and imagine an outcome $\omega \in \Omega$ is chosen at a random . The partial information $\mathcal{G}$ contains about ω is the knowledge of whether or not A contains ω for every $A \in \mathcal{G}$; equivalently, it is knowledge of the value of $\mathbf{1}_A(\omega)$, for every $A \in \mathcal{G}$. A conditional expectation $\mathbb{E}[Y|\mathcal{G}]$, as a $\mathcal{G}$-measurable function, is the expectation of Y given this partial information. Not a lot is gained mathematically by this interpretation, but it is a big aid to intuition. For an example, consider $\mathcal{F}_t^X$; it is easy to check that knowing $\mathbf{1}_A(\omega)$ for all $A \in \mathcal{F}_t^X$ is the same as knowing the value of $X_s(\omega)$ for all $s \leq t$.

By extension, filtration model the accumulation of information or knowledge as time progresses. commonly, $\mathcal{F}_t$ represents the "past" – all events that have occurred up to time t. Because conditional expectations and probabilities given the past, studied as t progresses, have turned out to be such important tools, the filtered probability space is the 'proper' abstract framework for a general theory of stochastic process.

Definition 1.2 A stochastic process $\mathbf{X}$ on $(\omega, \mathcal{F})$ is said to be *adapted* to $\mathbb{F}$ if $X(t)$ is $\mathcal{F}_t$-measurable for all $t \in \mathcal{J}$. When the filtration is clear from the context, we will just say that X is adapted.

Adaptedness is equivalent to demanding that $\mathcal{F}_t^X \subset \mathcal{F}_t$ for all $t \in \mathcal{J}$. In the loose language we introduce for interpreting filtrations, the partial information contained in $\mathcal{F}_t$ includes observation of $X(s)$ for all $s \leq t$.

2 Brownian Motion

2.1 From Random Walk to Brownian Motion

Recall the simple random walk, we have ξ_i's that are independent, identical distributed random variable that can take 1 and -1 with half probability. And

$$S_n = \sum_{i=1}^{n} \xi_i$$

with $S_0 = 0$. The recursive formula is: $S_n = S_{n-1} + \xi_n$, $n = 1, 2, ...$ We know that

$$\mathbb{E}[S_n] = 0, \ \mathrm{Var}[S_n] = \mathrm{Var}[\sum_{i=1}^{n} \xi_i] = n$$

Also,

$$\mathrm{Cov}(S_n, S_m) = \min(n, m)$$

Let's change the setting, fix an interval $[0, 1]$, split it into N equal size sub-intervals, i.e., $\delta_t := \frac{1}{N}$. We define $\{X_t^{(N)}\}_{t \in \{0, \delta_t, 2\delta_t, \cdots\}}$,

$$X^{(N)}_{(i+1)\delta_t} = X^{(N)}_{i\delta_t} + \begin{cases} +\delta_{X^{(N)}}, & \text{with prob. } \frac{1}{2} \\ -\delta_{X^{(N)}}, & \text{with prob. } \frac{1}{2} \end{cases}$$

with $X_0^{(N)} = 0$ and $t = i\delta_t = \frac{i}{N}$, $i = 1, ..., N$. Again,

$$\mathbb{E}[X_{i\delta_t}] = 0, \ \mathrm{Var}[X^{(N)}_{i\delta_t}] = i(\delta_{X^{(N)}})^2$$

In particular, $\mathrm{Var}[X_1^{(N)}] = N(\delta_{X^{(N)}})^2$. Since $i = tN$,

$$\mathrm{Var}(X_t^{(N)}) = tN(\delta_{X^{(N)}})^2$$

Let's define the step size: $\delta_{X^{(N)}} = \frac{1}{\sqrt{N}}$, then

$$\mathrm{Var}[X_t^{(N)}] = t$$

To compute covariance, assume $0 \le s \le t$ so that $s = \frac{j}{N}$ and $t = \frac{i}{N}$,

$$\mathrm{Cov}(X_s^{(N)}, X_t^{(N)}) = s$$

Let's write

$$X_t^{(N)} - X_s^{(N)} = \sum_{k=j+1}^{i} \frac{\xi_k}{\sqrt{N}}$$

where ξ_k's are i.i.d. and take 1 and -1 with equal probability.

Theorem 2.1 (*Central Limit Theorem*) Suppose $\{X_1, X_2, ...\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < +\infty$. Then as n approaches ∞, the random variable $\sqrt{n}(\sum_{i=1}^{n} X_i - \mu)$ converge in distribution to a normal $N(0, \sigma^2)$, i.e.,

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \to N(0, \sigma^2) \text{ in distribution.}$$

With the help of C.L.T, we can claim that

$$\frac{\sum_{k=j+1}^{i} \xi_k}{\sqrt{N(t-s)}} \sqrt{t-s} \sim N(0, t-s)$$

as $n \to \infty$.

This motivates the definition of *Brownian motion*. If we let the partition goes finer and finer, the simple random walk will 'converge to' the Brownian motion. Here comes the formal definition:

Definition 2.1 (*Standard Brownian motion*) The continuous-time stochastic process $\{W_t\}_{t\geq 0}$ is called *Brownian motion*:

1. if $0 \leq t \leq s$, then $W(s) - W(t) \sim N(0, s-t)$;
2. if $0 \leq t_1 < t_2 < \cdots < t_n$, then the increments $W(t_{i+1}) - W(t_i)$, $i = 1, ..., n-1$ are independent;
3. $W_0 = 0$;
4. With probability 1 all points of $W(t)(\omega)$ are continuous.

Note if 3 is not satisfied, it is called just *Brownian motion*.
A very important observation of Brownian motion is that it has *Markov Property*:

Theorem 2.2 (*Markov Property*) Let $W(t)$ be a *Brownian motion* and fix $s \geq 0$. Then $W(t+s) - W(s)$ is a *Standard Brownian motion* independent of $\{W_t | 0 \leq t \leq s\}$.

Proof. It is clear that $W(t+s)$ is a Brownian motion. Subtracting a constant only changes the starting point, and in particular, subtracting $W(s)$ makes the process a standard Brownian motion. Independence of $W(t)$ before time s follows from the independence of increments of Brownian motion. □

2.2 Stopping time & Brownian Motion

Let's recall the definition of *stopping time*,

Definition 2.2 (*stopping-time*) A random variable τ satisfying $\tau(\omega) \geq 0$ for all ω is called stopping time with respect to $\{\mathcal{F}_t\}_{t\geq 0}$ (or an $\{\mathcal{F}_t\}_{t\geq 0}$-stopping time) if

$$\{\omega : T(\omega) \leq t\} \in \mathcal{F}_t \text{ for every } t \geq 0$$

Instead of *Brownian motion*, let's first consider general continuous-time stochastic process that has path that is continuous:

Example 2.3 Assume that sample paths, $t \mapsto X(t)(\omega)$ are continuous for all ω. For simplicity, assume that $X(0)(\omega) = 0$ for all ω. For any constant $b > 0$, define

$$T_b(\omega) = \inf\{t \geq 0; X(t)(\omega) = b\}.$$

This is the first time that $X(t)$ 'hits' b. If $X(t)(\omega)$ never hits b, set $T_b(\omega) = \infty$. We claim: T_b is a stopping time w.r.t $\{\mathcal{F}_t^X\}_{t\geq 0}$.

The intuition behind this claim is that for any t, one needs only to know the values of $X(s)(\omega)$ for $s \leq t$ in order to know if $T_b(\omega) \leq t$ or not. Indeed, if $X(s^*)(\omega) \geq b$ for some $s^* \leq t$, then clearly, $T_b(\omega) \leq s^* \leq t$. On the other hand, if $X(s)(\omega) < b$ for all $s \leq t$, then certainly, $T_b(\omega) \geq t$, and since $X(t)(\omega) < b$ and X is continuous, in fact $T_b(\omega) > t$. More mathematically,

$$\{\omega : T_b(\omega) \leq t\} = \bigcup_{s\leq t} \{\omega; X(s)(\omega) \geq b\}$$

So it's obvious T_b is an $\{\mathcal{F}_t^X\}_{t\geq 0}$–stopping time, right? **Not quite!** Each set in the union on the right-hand side is indeed in $\mathcal{F}_t^X$, but the union is taken over an uncountable number of sets, one set for each $s \leq t$, and we only know $\mathcal{F}_t^X$ is closed under countable unions. So it is no possible to conclude $\{T_b \leq t\} \in \mathcal{F}_t$. Is it a contradiction? **No !** The result is still true but just need delicate analysis.

Remark 2.4 The day is saved by the following construction: fix any t and let $(q_1, q_2, ...)$ enumerate all the rational numbers in $[0, t]$, then

$$\{\omega; T_b(\omega) > t\} = \{\omega; X(\omega) < b \text{ for all } s \leq t\} = \cup_{m=1}^{\infty} A_m, \tag{1}$$

where

$$A_m = \bigcap_{n=1}^{\infty} \{\omega; X(q_n)(\omega) \leq b - \frac{1}{m}\}$$

Accepting this identity for the moment, note that A_m belongs to $\mathcal{F}_t^X$ for each m, because it is a countable intersection of sets in $\mathcal{F}^X(t)$. It then follows that $\{T_b > t\}$ is in $\mathcal{F}^X(t)$, since it is a countable union of events in $\mathcal{F}^X(t)$. Its complement $\{T_b \leq t\}$ is thus also in $\mathcal{F}_t^X$, which is what we need to show to prove it is a stopping time.

We still need to prove the correctness of (1). To see this, note first that A_m is the set of ω such that $X(s)(\omega) \leq b - \frac{1}{m}$ **for all** s, not just rational s; this is because if $s \leq t$ is not rational, there is a sequence $\{u_n\}$ of rational number less than t converging to s and so $X(s)(\omega) = \lim_{n\to\infty} X(u_n)(\omega)$, and it follows that if $X(u_n)(\omega) \leq b - \frac{1}{m}$ for each n, $X(s)(\omega) \leq b - \frac{1}{m}$ also. Next note that $\bigcup_1^\infty A_m$ is just the set of ω such that $X(s)(\omega) < b$ for all $s \leq t$, since if $X(s)(\omega) < b$ for all $s \leq t$, there will exists some $m > 0$ such that, in fact, $X(s)(\omega) \leq b - \frac{1}{m}$ for all $s \leq t$. Finally, $T_b(\omega) > t$ if and only if $X(s)(\omega) < b$ for all $s \leq t$. Observe that the continuity assumption was important in establishing this result.

Example 2.5 Let X be as in above example, let $b > 0$, but this time define

$$\tau_b(\omega) = \inf\{t; X(t)(\omega) > b\}.$$

This is called the first time that X enters the set (b, ∞). Since X has continuous paths and starts at 0 at time 0, $X(\tau_b(\omega)) = b/$ Therefore, $\tau_b(\omega) \geq T_b(\omega)$ for all ω.

This random time may not be $\{\mathcal{F}_t^X\}_{t\geq 0}$–stopping time. The reason is a little bit subtle. Indeed, suppose there are outcomes ω and ω' such that $X(s)(\omega) = X(s)(\omega') < b$ for $s \leq t$,

$X(t)(\omega) = X(t)(\omega') = b$, $X(s)(\omega) < b$ for $s > t$, and $X(s)(\omega') < b$ for $s < t$. Thus $X(s)(\omega)$ crosses b at time t, while $X(s)(\omega')$ only hits b at time t but goes no higher. Then $\tau_b(\omega) = t$, but $\tau_b(\omega) = \infty$. However, since the two processes are indistinguishable up to time t, there is no way of knowing from observing $X(s)$, $s \leq t$, which case we might be in and we will not be able to determine if $\tau_b(\omega) = t$ or $\tau_b(\omega) > t$. We can only do this by looking into the future beyond time t. Thus τ_b is not an $\{\mathcal{F}_t^X\}_{t\geq 0}$–stopping time.

Remark 2.6 More generally speaking, if we set first contact time in to A as

$$\tau_{c,A} = \inf\{t; X(t)(\omega) \in A\}$$

and suppose that X is a right continuous $\mathcal{F}_t$–adapted process, then

1. if A is closed and X has left limits at each $t > 0$ or if A is compact, then $\tau_{c,A}$ is an $\{\mathcal{F}_t^X\}_{t\geq 0}$–stopping time;
2. if A is open then $\tau_{c,A}$ is **not** an $\{\mathcal{F}_t^X\}_{t\geq 0}$–stopping time.

Example 2.7 Let X be a continuous process and let τ be the time it achieves its maximum value over all times s, $0 \leq s \leq 1$. In general this is not a stopping time. For example, to decide if $\tau(\omega) < 0.5$, knowing the values of $X(s)(\omega)$ for $s \leq 0.5$ is not enough. It is necessary to know $X(s)(\omega)$ for all $0 \leq s \leq 1$ to check whether it achieves its maximum before time $t = 0.5$.

If we replace X by Brownian motion W (with augmented filtration, which will be explained in the next recitation notes),

Theorem 2.8 When W is a Brownian motion, $\mathbb{P}(T_b = \tau_b)$.

In words, this theorem says that for any fixed b, W crosses into the region (b, ∞) immediately after hitting b, with probability one. Intuitively speaking, Brownian motion is such a process that it is wiggling so dramatically that after it touches certain fixed level, it will go across w.p. 1.

In *Theorem* 2.2, we already proved the strong markov property of Brownian motion, but actually it is still valid if we replace deterministic time t by a stopping time:

Theorem 2.9 Let τ be an $\{\mathcal{F}_t^W\}_{t\geq 0}$ –stopping time, let $W^\tau(u) = W(\tau + u) - W(\tau)$ be the process of the increments of W after τ. Then the process $\{W^\tau(u)\}_{u\geq 0}$ is a Brownian motion independent of all information in the filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$.

The proof is a little bit technical, but, intuitively, it should be clear.

3 The reflection principle of Brownian motion and its applications

3.1 Notations

Let W be a Brownian motion with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Let τ be a stopping time. We will define a new process by changing the sign of the increments $W(t) - W(\tau)$ when $t > \tau$,

$$B^\tau(t) = \begin{cases} W(t), & \text{if } t \leq \tau; \\ W(\tau) - [W(t) - W(\tau)], & \text{if } t > \tau, \end{cases} \tag{2}$$

we say that B^τ obtained by reflecting the increments of W after τ. This construction will be especially useful when $\tau = T_m$, where T_m is the first time that W hits m. Then $W(T_m) = m$ and the reflection procedure gives the process:

$$B(t) = \begin{cases} W(t), & \text{if } t \le T_m; \\ m - [W(t) - m], & \text{if } t > T_m, \end{cases} \tag{3}$$

This is obtained by reflecting the graph of W in the line $y = m$ after time T_m.

Theorem 3.1 (*Reflection Principle*) B^τ is a *Brownian motion*.

We know that $W(\tau + t) - W(\tau)$ is a Brownian motion independent of $W(t)$ for $t \le \tau$. If $Z(t)$ is any Brownian motion, so is $-Z(t)$, since multiplying by -1 does not change the independent increment property and since $Z(t) - Z(s)$ is normal with mean 0 and variance $t - s$, the same is true for $-(Z(t) - Z(s))$. Thus the reflected increment process $\{-[W(\tau + t) - W(t)]; t \ge 0\}$ is also a Brownian motion independent of $W(t)$ for $t \le \tau$. So by attaching it to $\{W(t); 0 \le t \le \tau\}$ instead of $\{[W(\tau + t) - W(t)]; t \ge 0\}$ one still obtains a Brownian motion. With a little work, this can be turned into a rigorous proof.

3.2 Brownian motion and its running maximum

Again, $\{W(t)\}_{t\ge 0}$ is our *Brownian motion* and $\{\mathcal{F}_t^W\}_{t\ge 0}$is the filtration for it. Let's define the running maximum process $\{M(t)\}_{t\ge 0}$, i.e.,

$$M(t) = \max_{0\le u\le t} W(u)$$

Denote $f_{M,W,t}(m, w)$ as the joint density of $(M(t), W(t))$, we want to compute $f_{M,W,t}(m, w)$. Note: $f_{M,W,t}(m, w) = 0$ if $m \le 0$ or $w > m$, because $\mathbb{P}(M(t) > 0) = 1$, $\mathbb{P}(W)(t) > M(t)) = 0$ (see *figure* 3). Let's first compute the probability of the following event (see *figure* **??**):

$$\mathbb{P}\left(M(t) \ge m, W(t) \le w\right) = \int_n^\infty \int_{-\infty}^w f_{M,W,t}(m', w')dw'dm'$$

It follows that

$$\frac{\partial}{\partial m}\mathbb{P}\left(M(t) \ge m, W(t) \le w\right) = -\int_{-\infty}^w f_{M,W,t}(m, w')dw'$$

and

$$\frac{\partial^2}{\partial w \partial m}\mathbb{P}\left(M(t) \ge m, W(t) \le w\right) = -f_{M,W,t}(m, w)$$

Hence,

$$f_{M,W,t}(m, w) = -\frac{\partial^2}{\partial w \partial m}\mathbb{P}\left(M(t) \ge m, W(t) \le w\right)$$

Observe that $\{M(t) \ge m, W(t) \le w\}$ if and only if $\{W^m(t) \ge 2m - \omega\}$ (see *figure* **??**) Thus,

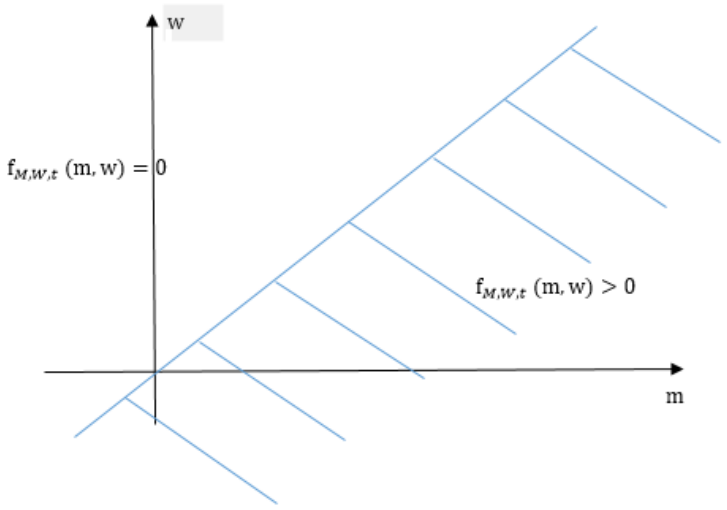

Figure 2: Visualization in 2-D

$$\begin{aligned}\mathbb{P}\left(M(t) \geq m, W(t) \leq w\right) &= \mathbb{P}\left(W^m(t) \geq 2m - w\right) \\ &= 1 - \mathbb{P}\left(W^m(t) < 2m - w\right) \\ &= 1 - F_{W^m}(2m - w)\end{aligned}$$

where F_{W^m} is the cumulative distribution function of W^m. Since W^m is a Brownian motion

$$F'_{W^m}(x) = \frac{e^{-x^2/2t}}{\sqrt{2\pi t}}$$

Thus,

$$\begin{aligned}f_{M,W,t}(m,w) &= -\frac{\partial^2}{\partial w \partial m}[1 - F_{W^m}(2m - w)] \\ &= \frac{\partial}{\partial w} 2F'_{W^m}(2m - w) = -2F''_{W^m}(2m - w) \\ &= -2\frac{\partial}{\partial x}\frac{e^{-x^2/2t}}{\sqrt{2\pi t}}|_{x=2m-w} \\ &= \frac{2x}{t\sqrt{2\pi t}} e^{-x^2/2t}|_{x=2m-w} \\ &= \frac{2(2m-w)}{t\sqrt{2\pi t}} e^{-(2m-w)^2/2t}, \ m > 0, w \leq m\end{aligned}$$

To summarize:

$$f_{M,W,t}(m,w) = \begin{cases} \frac{2(2m-w)}{t\sqrt{2\pi t}} e^{-(2m-w)^2/2t}, & \text{when } m > 0,\ w \leq m, \\ 0, & \text{otherwise.} \end{cases}$$

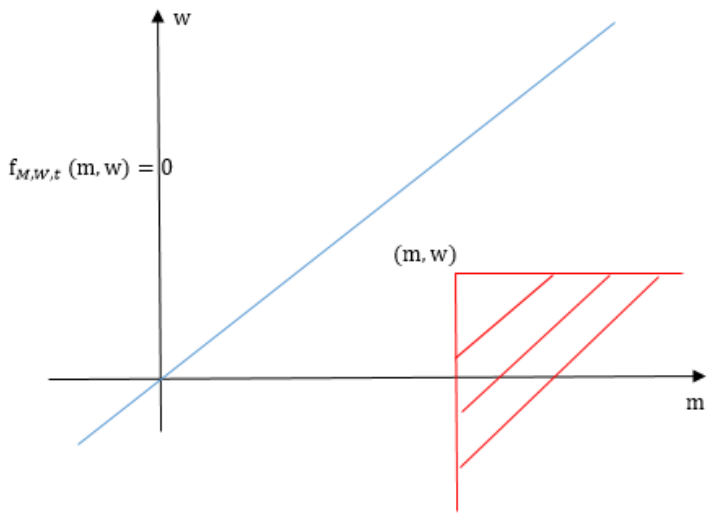

Figure 3: Region to be computed

3.3 Scaling Property of Brownian Motion

Sometimes, we may want to study transformations of functions which leave certain property invariant, and it is natural to ask what transformations of $W(t)$ preserve the same distribution.

Certainly, $-W(t)$ is a *Brownian motion*, because independence and continuity are maintained easily, in addition, since the normal distribution is symmetric about 0, all the increments have the same mean and variance as $W(t)$. Let's explore more transformations:

Proposition 3.2 If $W(t)$ is a standard Brownian motion, then so is the process $X(t) = aW(\frac{t}{a^2})$, for all $a > 0$.

Proof. Continuity and independence of increments still hold. For all $t > s \geq 0$, the normal random variable $X(t) - X(s) = a\left(W(\frac{t}{a^2}) - W(\frac{s}{a^2})\right)$ is distributed $aN(0, \frac{t-s}{a^2})$ that is equivalent to $N(0, t-s)$ in terms of distribution, so $X(t) - X(s) \sim N(0, t-s)$ as desired. □

This proposition tells us that $W(t)$ is a Brownian motions on all time scales as long as we compensate for the change in variance of the increments by taking a scalar multiple of the process. More surprisingly, we can invert the domain of $W(t)$ and still have a Brownian motion.

Proposition 3.3 (*Time-inversion*) Let $W(t)$ be a standard Brownian motion. Then the process

$$X(t) = \begin{cases} 0, & t = 0, \\ tW(\frac{1}{t}), & t \neq 0. \end{cases}$$

is also a standard Brownian motion.

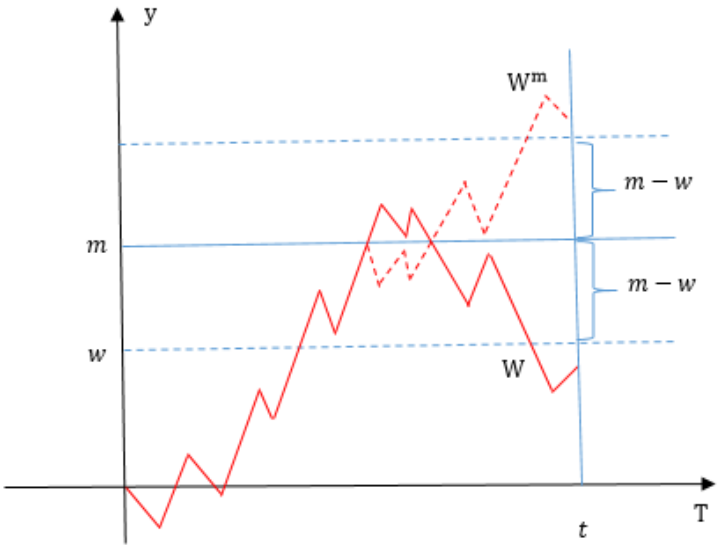

Figure 4: Reflection Principle

Proof. For Brownian motions,

$$\text{Cov}\left(W(t), W(t+s)\right) = \min(t, t+s) = t$$

for all $t, s \geq 0$. For our process $X(t)$,

$$\begin{aligned}\text{Cov}(X(t), X(t+s)) &= (tW(\frac{1}{t}), (t+s)W(\frac{1}{t+s})) \\ &= t(t+s)\frac{1}{t+s} = t\end{aligned}$$

So $\text{Cov}(X(t), X(t+s) - X(t)) = \text{Cov}(X(t), X(t+s)) - \text{Var}(X(t)) = t - t = 0$. Because the random variables $W(t+s)$ and $W(t)$ are normal, uncorrelated implies $X(t+s) - X(t)$ and $X(t)$ are independent. And

$$\begin{aligned}&\text{Var}\left(X(t+s) - X(t)\right) \\ =&\text{Var}(X(t+s)) + \text{Var}(X(t)) - 2\text{Cov}\left(X(t+s), X(t)\right) \\ =&t + s + t - 2t = s\end{aligned}$$

Thus, we get the right variance.

Continuity is clear for $t > 0$, thus it suffices to show the continuity at 0, i.e.,

$$\lim_{t\to 0} X(t) = \lim_{t\to 0} tW(\frac{1}{t}) = \lim_{t\to\infty} \frac{W(t)}{t} = 0 \text{ a.s.},$$

or equivalently, for any $\epsilon > 0$,

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{|W(t)|}{t} > \epsilon\right) = 0$$

Let $a > 0$, for $t \geq s \geq 0$,

$$\begin{aligned}
&\mathbb{E}[e^{aW(t)}|\mathcal{F}_s] \\
=&\mathbb{E}[e^{a(W(t)-W(s)+W(s))}|\mathcal{F}_s] \\
=&\mathbb{E}[e^{a(W(t)-W(s))}e^{aW(s)}|\mathcal{F}_s] \\
=&e^{aW(s)}\mathbb{E}[e^{a(W(t)-W(s))}]
\end{aligned}$$

Since $Y = W(t) - W(s) \sim N(0, t-s)$, by moment generating function of a normal random variable,

$$\mathbb{E}[e^{tY}] = e^{a\mu+\frac{1}{2}\sigma^2a^2} = e^{\frac{1}{2}a^2(t-s)} \geq 1$$

Thus, $e^{aW(t)}$ is a sub-martingale. By *Doob's martingale inequality*:

$$\mathbb{P}\left(\sup_{s\leq t} W(s) \geq \lambda\right) = \mathbb{P}\left(\sup_{s\leq t} e^{aW(s)} \geq e^{a\lambda}\right) \leq \frac{\mathbb{E}[e^{aW(t)}]}{e^{a\lambda}} = e^{a^2t/2-a\lambda}$$

Setting $a = \lambda/t$, we have

$$\mathbb{P}\left(\sup_{s\leq t} W(s) \geq \lambda\right) \leq e^{-\lambda^2/2t}$$

On the other hand, notice that

$$\sup_{n\leq s\leq n+1} \frac{|W(s)|}{s} \geq \epsilon$$

implies there exists $s \in [n, n+1]$ such that $|W(s)| > \epsilon s$, that is, $\exists s \in [n, n+1]$ such that $|W(s)| > \epsilon n \geq \frac{\epsilon(n+1)}{2}$. This leads to

$$\sup_{s\leq n+1} |W(s)| \geq \frac{\epsilon(n+1)}{2}$$

Thus, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\geq N} \frac{|W(t)|}{t} > \epsilon\right) \\
\leq &\sum_{n=N}^{\infty} \mathbb{P}\left(\sup_{n\leq t\leq n+1} \frac{|W(t)|}{t} > \epsilon\right) \\
\leq &\sum_{n=N+1}^{\infty} \mathbb{P}\left(\sup_{s\leq n} |W(s)| \geq \frac{\epsilon n}{2}\right) \\
\leq &2\sum_{n=N+1}^{\infty} e^{-\epsilon^2 n/8} = O_\epsilon(e^{-\epsilon^2 n/8})
\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{|W(t)|}{t} > \epsilon\right) = \lim_{N\to\infty} \mathbb{P}\left(\sup_{t\geq N} \frac{|W(t)|}{t} > \epsilon\right) = 0$$

□

4 Reference

1. Jean Jacod, Philip Protter, "Probability Esentials", Springer, 2004;
2. E.Cinlar, "Probability and Stochastics", Springer, 2011
3. Daniel, Ocone, "Notes on Ergodic Theory", 2012
4. Alison Etheridge, "A course in financial calculus", Cambridge, 2002