CHAPTER 2

Brownian motion

1. Scaling limit of random walks.

Our first goal is to understand *Brownian motion*, which is used to model "noisy fluctuations" of stocks, and various other objects. This is named after the botanist Robert Brown, who observed that the microscopic movement of pollen grains appears random. Intuitively, Brownian motion can be thought of as a process that performs a random walk in continuous time.

We begin by describing Brownian motion as the scaling limit of discrete random walks. Let X_1, X_2, ..., be a sequence of i.i.d. random variables which take on the values ± 1 with probability $1/2$. Define the time interpolated random walk $S(t)$ by setting $S(0) = 0$, and

$$(1.1) \qquad S(t) = S(n) + (t - n)X_{n+1} \quad \text{when } t \in (n, n+1] \,.$$

Note $S(n) = \sum_1^n X_i$, and so at integer times S is simply a symmetric random walk with step size 1.

Our aim now is to rescale S so that it takes a random step at shorter and shorter time intervals, and then take the limit. In order to get a meaningful limit, we will have to compensate by also scaling the step size. Let $\varepsilon > 0$ and define

$$(1.2) \qquad S_\varepsilon(t) = \alpha_\varepsilon S\left(\frac{t}{\varepsilon}\right),$$

where α_ε will be chosen below in a manner that ensures convergence of $S_\varepsilon(t)$ as $\varepsilon \to 0$. Note that S_ε now takes a random step of size α_ε after every ε time units.

To choose α_ε, we compute the variance of S_ε. Note first

$$\text{Var}\, S(t) = \lfloor t \rfloor + (t - \lfloor t \rfloor)^2,$$

and[1] consequently

$$\text{Var}\, S_\varepsilon(t) = \alpha_\varepsilon^2\left(\left\lfloor \frac{t}{\varepsilon} \right\rfloor + \left(\frac{t}{\varepsilon} - \left\lfloor \frac{t}{\varepsilon} \right\rfloor\right)^2\right).$$

In order to get a "nice limit" of S_ε as $\varepsilon \to 0$, one would at least expect that $\text{Var}\, S_\varepsilon(t)$ converges as $\varepsilon \to 0$. From the above, we see that choosing

$$\alpha_\varepsilon = \sqrt{\varepsilon}$$

immediately implies

$$\lim_{\varepsilon \to 0} \text{Var}\, S_\varepsilon(t) = t \,.$$

---

[1]Here $\lfloor x \rfloor$ denotes the greatest integer smaller than x. That is, $\lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leqslant x\}$.

THEOREM 1.1. *The processes $S_\varepsilon(t) \stackrel{\text{def}}{=} \sqrt{\varepsilon}S(t/\varepsilon)$ "converge" as $\varepsilon \to 0$. The limiting process, usually denoted by W, is called a (standard, one dimensional) Brownian motion.*

The proof of this theorem uses many tools from the modern theory of probability, and is beyond the scope of this course. The important thing to take away from this is that Brownian motion can be well approximated by a random walk that takes steps of variance ε on a time interval of size ε.

2. A crash course in measure theoretic probability.

Each of the random variables X_i can be adequately described by finite probability spaces. The collection of all X_i's can not be, but is still "intuitive enough" to to be understood using the tools from discrete probability. The limiting process W, however, can not be adequately described using tools from "discrete" probability: For each t, $W(t)$ is a continuous random variable, and the collection of all $W(t)$ for $t \geqslant 0$ is an *uncountable* collection of *correlated* random variables. This process is best described and studied through measure theoretic probability, which is very briefly described in this section.

DEFINITION 2.1. The *sample space* Ω is simply a non-empty set.

DEFINITION 2.2. A *σ-algebra* $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ is a *non-empty* collection of *subsets* of Ω which is:
  (1) closed under compliments (i.e. if $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$),
  (2) and closed under *countable* unions (i.e. A_1, A_2, ... are all elements of $\mathcal{G}$, then the union $\cup_1^\infty A_i$ is also an element of $\mathcal{G}$).
Elements of the σ-algebra are called *events*, or $\mathcal{G}$-measurable events.

REMARK 2.3. The notion of σ-algebra is *central* to probability, and represents *information*. Elements of the σ-algebra are events whose probability are known.

REMARK 2.4. You should check that the above definition implies that $\emptyset, \Omega \in \mathcal{G}$, and that $\mathcal{G}$ is also closed under countable intersections and set differences.

DEFINITION 2.5. A *probability measure* on $(\Omega, \mathcal{G})$ is a *countably additive* function $\boldsymbol{P}\colon \mathcal{G} \to [0,1]$ such that $\boldsymbol{P}(\Omega) = 1$. That is, for each $A \in \mathcal{G}$, $\boldsymbol{P}(A) \in [0,1]$ and $\boldsymbol{P}(\Omega) = 1$. Moreover, if A_1, $A_2, \dots \in \mathcal{G}$ are *pairwise disjoint*, then

$$\boldsymbol{P}\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \boldsymbol{P}(A_i) \,.$$

The triple $(\Omega, \mathcal{G}, \boldsymbol{P})$ is called a *probability space*.

REMARK 2.6. For a $\mathcal{G}$-measurable event A, $\boldsymbol{P}(A)$ represents the probability of the event A occurring.

REMARK 2.7. You should check that the above definition implies:
  (1) $\boldsymbol{P}(\emptyset) = 0$,
  (2) If $A, B \in \mathcal{G}$ are disjoint, then $\boldsymbol{P}(A \cup B) = \boldsymbol{P}(A) + \boldsymbol{P}(B)$.
  (3) $\boldsymbol{P}(A^c) = 1 - \boldsymbol{P}(A)$. More generally, if $A, B \in \mathcal{G}$ with $A \subseteq B$, then $\boldsymbol{P}(B - A) = \boldsymbol{P}(B) - \boldsymbol{P}(A)$.

(4) If $A_1 \subseteq A_2 \subseteq A_3 \cdots$ and each $A_i \in \mathcal{G}$ then $\boldsymbol{P}(\cup A_i) = \lim_{n \to \infty} \boldsymbol{P}(A_n)$.

(5) If $A_1 \supseteq A_2 \supseteq A_3 \cdots$ and each $A_i \in \mathcal{G}$ then $\boldsymbol{P}(\cap A_i) = \lim_{n \to \infty} \boldsymbol{P}(A_n)$.

DEFINITION 2.8. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that for every $\alpha \in \mathbb{R}$, the set $\{\omega \in \Omega \mid X(\omega) \leqslant \alpha\}$ is an element of $\mathcal{G}$. (Such functions are also called $\mathcal{G}$-*measurable*, *measurable with respect to* $\mathcal{G}$, or simply *measurable* if the σ-algebra in question is clear from the context.)

REMARK 2.9. The argument ω is *always* suppressed when writing random variables. That is, the event $\{\omega \in \Omega \mid X(\omega) \leqslant \alpha\}$ is simply written as $\{X \leqslant \alpha\}$.

REMARK 2.10. Note for any random variable, $\{X > \alpha\} = \{X \leqslant \alpha\}^c$ which must also belong to $\mathcal{G}$ since $\mathcal{G}$ is closed under complements. You should check that for every $\alpha < \beta \in \mathbb{R}$ the events $\{X < \alpha\}$, $\{X \geqslant \alpha\}$, $\{X > \alpha\}$, $\{X \in (\alpha, \beta)\}$, $\{X \in [\alpha, \beta)\}$, $\{X \in (\alpha, \beta]\}$ and $\{X \in (\alpha, \beta)\}$ are all also elements of $\mathcal{G}$.

Since $\boldsymbol{P}$ is defined on all of $\mathcal{G}$, the quantity $\boldsymbol{P}(\{X \in (\alpha, \beta)\})$ is mathematically well defined, and represents the chance that the random variable X takes values in the interval (α, β). For brevity, I will often omit the outermost curly braces and write $\boldsymbol{P}(X \in (\alpha, \beta))$ for $\boldsymbol{P}(\{X \in (\alpha, \beta)\})$.

REMARK 2.11. You should check that if X, Y are random variables then so are $X \pm Y$, XY, X/Y (when defined), $|X|$, $X \wedge Y$ and $X \vee Y$. In fact if $f : \mathbb{R} \to \mathbb{R}$ is any reasonably nice (more precisely, a Borel measurable) function, $f(X)$ is also a random variable.

EXAMPLE 2.12. If $A \subseteq \Omega$, define $\boldsymbol{1}_A : \Omega \to \mathbb{R}$ by $\boldsymbol{1}_A(\omega) = 1$ if $\omega \in A$ and 0 otherwise. Then $\boldsymbol{1}_A$ is a ($\mathcal{G}$-measurable) random variable if and only if $A \in \mathcal{G}$.

EXAMPLE 2.13. For $M \in \mathbb{N}$, $i \in \{1, \dots, M\}$, $a_i \in \mathbb{R}$ and $A_i \in \mathcal{G}$ be such that $A_i \cap A_j = \emptyset$ for $i \neq j$, and define

$$X \stackrel{\text{def}}{=} \sum_{i=1}^{M} a_i \boldsymbol{1}_{A_i}.$$

Then X is a ($\mathcal{G}$-measurable) random variable. (Such variables are called *simple* random variables.)

Note that if the a_i's above are all distinct, then $\{X = a_i\} = A_i$, and hence $\sum_i a_i \boldsymbol{P}(X = a_i) = \sum_i a_i \boldsymbol{P}(A_i)$, which agrees with our notion of expectation from discrete probability.

DEFINITION 2.14. For the simple random variable X defined above, we define *expectation of X* by

$$\boldsymbol{E}X = \sum_{i=1}^{M} a_i \boldsymbol{P}(A_i).$$

For general random variables, we define the expectation by approximation.

DEFINITION 2.15. If Y is a nonnegative random variable, define

$$\boldsymbol{E}Y \stackrel{\text{def}}{=} \lim_{n \to \infty} \boldsymbol{E}X_n \quad \text{where } X_n \stackrel{\text{def}}{=} \sum_{k=0}^{n^2-1} \frac{k}{n} \boldsymbol{1}_{\{\frac{k}{n} \leqslant Y < \frac{k+1}{n}\}}.$$

REMARK 2.16. Note each X_n above is simple, and we have previously defined the expectation of simple random variables.

DEFINITION 2.17. If Y is any (not necessarily nonnegative) random variable, set $Y^+ = Y \vee 0$ and $Y^- = Y \wedge 0$, and define the expectation by

$$\boldsymbol{E}Y = \boldsymbol{E}Y^+ - \boldsymbol{E}Y^-,$$

provided at least one of the terms on the right is finite.

REMARK 2.18. The expectation operator defined above is the *Lebesgue integral* of Y with respect to the probability measure $\boldsymbol{P}$, and is often written as

$$\boldsymbol{E}Y = \int_\Omega Y \, d\boldsymbol{P}.$$

More generally, if $A \in \mathcal{G}$ we define

$$\int_A Y \, d\boldsymbol{P} \stackrel{\text{def}}{=} \boldsymbol{E}(\boldsymbol{1}_A Y),$$

and when $A = \Omega$ we will often omit writing it.

PROPOSITION 2.19 (Linearity). *If $\alpha \in \mathbb{R}$ and X, Y are random variables, then* $\boldsymbol{E}(X + \alpha Y) = \boldsymbol{E}X + \alpha \boldsymbol{E}Y$.

PROPOSITION 2.20 (Positivity). *If $X \geqslant 0$ almost surely, then $\boldsymbol{E}X \geqslant 0$ Moreover, if $X > 0$ almost surely, $\boldsymbol{E}X > 0$. Consequently, (using linearity) if $X \leqslant Y$ almost surely then $\boldsymbol{E}X \leqslant \boldsymbol{E}Y$.*

REMARK 2.21. By $X \geqslant 0$ almost surely, we mean that $\boldsymbol{P}(X \geqslant 0) = 1$.

The proof of positivity is immediate, however the proof of linearity is surprisingly not as straightforward as you would expect. It's easy to verify linearity for simple random variables, of course. For general random variables, however, you need an approximation argument which requires either the *dominated* or *monotone* convergence theorem which guarantee $\lim \boldsymbol{E}X_n = \boldsymbol{E} \lim X_n$, under modest assumptions. Since discussing these results at this stage will will lead us too far astray, we invite the curious to look up their proofs in any standard measure theory book. The main point of this section was to introduce you to a framework which is capable of describing and studying the objects we will need for the remainder of the course.

3. A first characterization of Brownian motion.

We introduced Brownian motion by calling it a certain scaling limit of a simple random walk. While this provides good intuition as to what Brownian motion actually is, it is a somewhat unwieldy object to work with. Our aim here is to provide an intrinsic characterization of Brownian motion, that is both useful and mathematically convenient.

DEFINITION 3.1. A Brownian motion is a *continuous process* that has stationary independent increments.

We will describe what this means shortly. While this is one of the most intuitive definitions of Brownian motion, most authors choose to prove this as a theorem, and use the following instead.

DEFINITION 3.2. A Brownian motion is a *continuous process* W such that:

(1) W has independent increments, and
(2) For $s < t$, $W(t) - W(s) \sim N(0, \sigma^2(t - s))$.

REMARK 3.3. A *standard* (one dimensional) Brownian motion is one for which $W(0) = 0$ and $\sigma = 1$.

Both these definitions are equivalent, thought the proof is beyond the scope of this course. In order to make sense of these definitions we need to define the terms *continuous process*, *stationary increments*, and *independent increments*.

3.1. Continuous processes.

DEFINITION 3.4. A *stochastic process* (aka *process*) is a function $X \colon \Omega \times [0, \infty)$ such that for every time $t \in [0, \infty)$, the function $\omega \mapsto X(t, \omega)$ is a random variable. The ω variable is usually suppressed, and we will almost always use $X(t)$ to denote the random variable obtained by taking the slice of the function X at time t.

DEFINITION 3.5. A *continuous process* (aka *continuous stochastic process*) is a stochastic process X such that for (almost) every $\omega \in \Omega$ the function $t \mapsto X(t, \omega)$ is a continuous function of t. That is,

$$\boldsymbol{P}\Big(\lim_{s \to t} X(s) = X(t) \text{ for every } t \in [0, \infty)\Big) = 1 \,.$$

The processes $S(t)$ and $S_\varepsilon(t)$ defined in (1.1) and (1.2) are continuous, but the process

$$\tilde{S}(t) \stackrel{\text{def}}{=} \sum_{n=0}^{\lfloor t \rfloor} X_n \,,$$

is not.

In general it is not true that the limit of continuous processes is again continuous. However, one can show that the limit of S_ε (with $\alpha_\varepsilon = \sqrt{\varepsilon}$ as above yields a continuous process.

3.2. Stationary increments.

DEFINITION 3.6. A process X is said to have stationary increments if the distribution of $X_{t+h} - X_t$ does not depend on t.

For the process S in (1.1), note that for $n \in \mathbb{N}$, $S(n + 1) - S(n) = X_{n+1}$ whose distribution *does not* depend on n as the variables $\{X_i\}$ were chosen to be independent and identically distributed. Similarly, $S(n + k) - S(n) = \sum_{n+1}^{n+k} X_i$ which has the same distribution as $\sum_{1}^{k} X_i$ and is independent of n.

However, if $t \in \mathbb{R}$ and is not necessarily an integer, $S(t+k) - S(t)$ will in general depend on t. So the process S (and also S_ε) do not have stationary increments.

We claim, that the limiting process W does have stationary (normally distributed) increments. Suppose for some fixed $\varepsilon > 0$, both s and t are multiples of ε. In this case

$$S_\varepsilon(t) - S_\varepsilon(s) \sim \sqrt{\varepsilon} \sum_{i=1}^{\lfloor t-s \rfloor / \varepsilon} X_i \xrightarrow{\varepsilon \to 0} N(0, t - s) \,,$$

by the central limit theorem. If s, t aren't multiples of ε as we will have in general, the first equality above is true up to a remainder which can easily be shown to vanish.

The above heuristic argument suggests that the limiting process W (from Theorem 1.1) satisfies $W(t) - W(s) \sim N(0, t - s)$. This certainly has independent increments since $W(t + h) - W(t) \sim N(0, h)$ which is independent of t. This is also the reason why the normal distribution is often pre-supposed when defining Brownian motion.

3.3. Independent increments.

DEFINITION 3.7. A process X is said to have *independent increments* if for every finite sequence of times $0 \leqslant t_0 < t_1 \cdots < t_N$, the random variables $X(t_0)$, $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, ..., $X(t_N) - X(t_{N-1})$ are all jointly independent.

Note again for the process S in (1.1), the increments at *integer times* are independent. Increments at non-integer times are correlated, however, one can show that in the limit as $\varepsilon \to 0$ the increments of the process S_ε become independent.

Since we assume the reader is familiar with independence from discrete probability, the above is sufficient to motivate and explain the given definitions of Brownian motion. However, notion of independence is important enough that we revisit it from a measure theoretic perspective next. This also allows us to introduce a few notions on σ-algebras that will be crucial later.

3.4. Independence in measure theoretic probability.

DEFINITION 3.8. Let X be a random variable on $(\Omega, \mathcal{G}, \boldsymbol{P})$. Define $\sigma(X)$ to be the σ-algebra generated by the events $\{X \leqslant \alpha\}$ for every $\alpha \in \mathbb{R}$. That is, $\sigma(X)$ is the *smallest* σ-algebra which contains each of the events $\{X \leqslant \alpha\}$ for every $\alpha \in \mathbb{R}$.

REMARK 3.9. The σ-algebra $\sigma(X)$ represents all the information one can learn by observing X. For instance, consider the following game: A card is drawn from a shuffled deck, and you win a dollar if it is red, and lose one if it is black. Now the likely hood of drawing any particular card is $1/52$. However, if you are blindfolded and only told the outcome of the game, you have no way to determine that each gard is picked with probability $1/52$. The only thing you will be able to determine is that red cards are drawn as often as black ones.

This is captured by σ-algebra as follows. Let $\Omega = \{1, \ldots, 52\}$ represent a deck of cards, $\mathcal{G} = \mathcal{P}(\Omega)$, and define $\boldsymbol{P}(A) = \text{card}(A)/52$. Let $R = \{1, \ldots 26\}$ represent the red cards, and $B = R^c$ represent the black cards. The outcome of the above game is now the random variable $X = \mathbf{1}_R - \mathbf{1}_B$, and you should check that $\sigma(X)$ is exactly $\{\emptyset, R, B, \Omega\}$.

We will use σ-algebras extensively but, as you might have noticed, we haven't developed any examples. Infinite σ-algebras are "hard" to write down explicitly, and what one usually does in practice is specify a *generating* family, as we did when defining $\sigma(X)$.

DEFINITION 3.10. Given a collection of sets A_α, where α belongs to some (possibly infinite) index set $\mathcal{A}$, we define $\sigma(\{A_\alpha\})$ to be the *smallest* σ-algebra that contains each of the sets A_α.

That is, if $\mathcal{G} = \sigma(\{A_\alpha\})$, then we must have each $A_\alpha \in \mathcal{G}$. Since $\mathcal{G}$ is a σ-algebra, all sets you can obtain from these by taking complements, countable unions and countable intersections intersections must also belong to $\mathcal{G}$.[2] The fact that $\mathcal{G}$ is the smallest σ-algebra containing each A_α also means that if $\mathcal{G}'$ is any other σ-algebra that contains each A_α, then $\mathcal{G} \subseteq \mathcal{G}'$.

REMARK 3.11. The smallest σ-algebra under which X is a random variable (under which X is measurable) is exactly $\sigma(X)$. It turns out that $\sigma(X) = X^{-1}(\mathcal{B}) = \{X \in B \mid B \in \mathcal{B}\}$, where $\mathcal{B}$ is the *Borel* σ-algebra on $\mathbb{R}$. Here $\mathcal{B}$ is the *Borel* σ-algebra, defined to be the σ-algebra on $\mathbb{R}$ generated by all open intervals.

DEFINITION 3.12. We say the random variables $X_1, \ldots, X_N$ are independent if for every $i \in \{1 \ldots N\}$ and every $A_i \in \sigma(X_i)$ we have

$$\boldsymbol{P}\big(A_1 \cap A_2 \cap \cdots \cap A_N\big) = \boldsymbol{P}(A_1)\,\boldsymbol{P}(A_2)\cdots\boldsymbol{P}(A_N)\,.$$

REMARK 3.13. Recall two events A, B are independent if $\boldsymbol{P}(A \mid B) = \boldsymbol{P}(A)$, or equivalently A, B satisfy the multiplication law: $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)$. A collection of events $A_1, \ldots, A_N$ is said to be independent if *any* sub collection $\{A_{i_1}, \ldots, A_{i_k}\}$ satisfies the multiplication law. *This is a stronger condition than simply requiring* $\boldsymbol{P}(A_1 \cap \cdots \cap A_N) = \boldsymbol{P}(A_1)\cdots\boldsymbol{P}(A_N)$. You should check, however, that if the random variables $X_1, \ldots, X_N$, are all independent, then any collection of events of the form $\{A_1, \ldots A_N\}$ with $A_i \in \sigma(X_i)$ is also independent.

PROPOSITION 3.14. *Let $X_1, \ldots, X_N$ be N random variables. The following are equivalent:*

(1) The random variables $X_1, \ldots, X_N$ are independent.

(2) For every $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$, we have

$$\boldsymbol{P}\Big(\bigcap_{j=1}^{N}\{X_j \leqslant \alpha_j\}\Big) = \prod_{j=1}^{N}\boldsymbol{P}(X_j \leqslant \alpha_j)$$

(3) For every collection of bounded continuous functions $f_1, \ldots, f_N$ we have

$$\boldsymbol{E}\Big[\prod_{j=1}^{N} f_j(X_j)\Big] = \prod_{j=1}^{N}\boldsymbol{E}f_j(X_j)\,.$$

(4) For every $\xi_1, \ldots, \xi_N \in \mathbb{R}$ we have

$$\boldsymbol{E}\exp\Big(i\sum_{j=1}^{N}\xi_j X_j\Big) = \prod_{j=1}^{N}\boldsymbol{E}\exp(i\xi_j X_j)\,, \quad \text{where } i = \sqrt{-1}\,.$$

REMARK 3.15. It is instructive to explicitly check each of these implications when $N = 2$ and X_1, X_2 are simple random variables.

---

[2] Usually $\mathcal{G}$ contains *much more* than all countable unions, intersections and complements of the A_α's. You might think you could keep including all sets you generate using countable unions and complements and arrive at all of $\mathcal{G}$. It turns out that to make this work, you will usually have to do this *uncountably* many times!

This won't be too important within the scope of these notes. However, if you read a rigorous treatment and find the authors using some fancy trick (using Dynkin systems or monotone classes) instead of a naive countable unions argument, then the above is the reason why.

REMARK 3.16. The intuition behind the above result is as follows: Since the events $\{X_j \leqslant \alpha_j\}$ generate $\sigma(X_j)$, we expect the first two properties to be equivalent. Since $\mathbf{1}_{(-\infty, \alpha_j]}$ can be well approximated by continuous functions, we expect equivalence of the second and third properties. The last property is a bit more subtle: Since $\exp(a + b) = \exp(a)\exp(b)$, the third clearly implies the last property. The converse holds because of "completeness of the complex exponentials" or Fourier inversion, and again a through discussion of this will lead us too far astray.

REMARK 3.17. The third implication above implies that independent random variables are uncorrelated. The converse, is of course false. However, the normal correlation theorem guarantees that jointly normal uncorrelated variables are independent.

REMARK 3.18. If moment generating functions of the random variables are defined in an interval around 0, then one can test independence using real exponentials instead of the complex exponentials used in the last condition in Proposition 3.14. Explicitly, in this case $X_1, \ldots, X_N$ are independent if and only if for every $t_1, \ldots, t_N$ in some small interval containing 0 we have

$$\boldsymbol{E}\exp\Big(\sum_{j=1}^{N} t_j X_j\Big) = \prod_{j=1}^{N}\boldsymbol{E}\exp(t_j X_j)\,.$$

3.5. The covariance of Brownian motion. The independence of increments allows us to compute covariances of Brownian motion easily. Suppose W is a standard Brownian motion, and $s < t$. Then we know $W_s \sim N(0, s)$, $W_t - W_s \sim N(0, t - s)$ and is independent of W_s. Consequently $(W_s, W_t - W_s)$ is jointly normal with mean 0 and covariance matrix $\left(\begin{smallmatrix} s & 0 \\ 0 & t-s \end{smallmatrix}\right)$. This implies that (W_s, W_t) is a jointly normal random variable. Moreover we can compute the covariance by

$$\boldsymbol{E}W_s W_t = \boldsymbol{E}W_s(W_t - W_s) + \boldsymbol{E}W_s^2 = s\,.$$

In general if you don't assume $s < t$, the above immediately implies $\boldsymbol{E}W_s W_t = s \wedge t$.

4. The Martingale Property

A martingale is "fair game". Suppose you are playing a game and $M(t)$ is your cash stockpile at time t. As time progresses, you learn more and more information about the game. For instance, in blackjack getting a high card benefits the player more than the dealer, and a common card counting strategy is to have a "spotter" betting the minimum while counting the high cards. When the odds of getting a high card are favorable enough, the player will signal a "big player" who joins the table and makes large bets, as long as the high card count is favorable. Variants of this strategy have been shown to give the player up to a 2% edge over the house.

If a game is a martingale, then this extra information you have acquired *can not* help you going forward. That is, if you signal your "big player" at any point, you will not affect your expected return.

Mathematically this translates to saying that the *conditional expectation* of your stockpile at a later time given your present accumulated knowledge, is exactly the present value of your stockpile. Our aim in this section is to make this precise.

4.1. Conditional probability. Suppose you have an incomplete deck of cards which has 10 red cards, and 20 black cards. Suppose 5 of the red cards are *high cards* (i.e. ace, king, queen, jack or 10), and only 4 of the black cards are high. If a card is chosen at random, the *conditional probability* of it being high given that it is red is $1/2$, and the *conditional probability* of it being high given that it is black is $1/5$. Our aim is to encode both these facts into a single entity.

We do this as follows. Let R, B denote the set of all red and black cards respectively, and H denote the set of all high cards. A σ-algebra encompassing all the above information is exactly

$$\mathcal{G} \stackrel{\text{def}}{=} \{\emptyset, R, B, H, H^c, R \cap H, B \cap H, R \cap H^c, B \cap H^c,$$
$$(R \cap H) \cup (B \cap H^c), (R \cap H^c) \cup (B \cap H), \Omega\}$$

and you can explicitly compute the probabilities of each of the above events. A σ-algebra encompassing only the color of cards is exactly

$$\mathcal{G} \stackrel{\text{def}}{=} \{\emptyset, R, B, \Omega\}.$$

Now we define the *conditional probability* of a card being high given the color to be the **random variable**

$$\boldsymbol{P}(H \mid \mathcal{C}) \stackrel{\text{def}}{=} \boldsymbol{P}(H \mid R)\mathbf{1}_R + \boldsymbol{P}(H \mid B)\mathbf{1}_B = \frac{1}{2}\mathbf{1}_R + \frac{1}{5}\mathbf{1}_B.$$

To emphasize:

(1) What is given is the σ-algebra $\mathcal{C}$, and not just an event.
(2) The conditional probability is now a $\mathcal{C}$-*measurable random variable* and not a number.

To see how this relates to $\boldsymbol{P}(H \mid R)$ and $\boldsymbol{P}(H \mid B)$ we observe

$$\int_R \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P} \stackrel{\text{def}}{=} \boldsymbol{E}\big(\mathbf{1}_R \boldsymbol{P}(H \mid \mathcal{C})\big) = \boldsymbol{P}(H \mid R)\,\boldsymbol{P}(R).$$

The same calculation also works for B, and so we have

$$\boldsymbol{P}(H \mid R) = \frac{1}{\boldsymbol{P}(R)} \int_R \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P} \quad \text{and} \quad \boldsymbol{P}(H \mid B) = \frac{1}{\boldsymbol{P}(B)} \int_B \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P}.$$

Our aim is now to generalize this to a non-discrete scenario. The problem with the above identities is that if either R or B had probability 0, then the above would become meaningless. However, clearing out denominators yields

$$\int_R \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P} = \boldsymbol{P}(H \cap R) \quad \text{and} \quad \int_B \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P} = \boldsymbol{P}(H \cap B).$$

This suggests that the defining property of $\boldsymbol{P}(H \mid \mathcal{C})$ should be the identity

(4.1) $$\int_C \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P} = \boldsymbol{P}(H \cap C)$$

for every event $C \in \mathcal{C}$. Note $\mathcal{C} = \{\emptyset, R, B, \Omega\}$ and we have only checked (4.1) for $C = R$ and $C = B$. However, for $C = \emptyset$ and $C = \Omega$, (4.1) is immediate.

DEFINITION 4.1. Let $(\Omega, \mathcal{G}, \boldsymbol{P})$ be a probability space, and $\mathcal{F} \subseteq \mathcal{G}$ be a σ-algebra. Given $A \in \mathcal{G}$, we define the conditional probability of A given $\mathcal{F}$, denoted by $\boldsymbol{P}(A | \mathcal{F})$ to be an $\mathcal{F}$-measurable random variable that satisfies

(4.2) $$\int_F \boldsymbol{P}(H \mid \mathcal{F}) \, d\boldsymbol{P} = \boldsymbol{P}(H \cap F) \quad \text{for every } F \in \mathcal{F}.$$

REMARK 4.2. Showing existence (and uniqueness) of the conditional probability isn't easy, and relies on the *Radon-Nikodym theorem*, which is beyond the scope of this course.

REMARK 4.3. It is crucial to require that $\boldsymbol{P}(H|\mathcal{F})$ is measurable with respect to $\mathcal{F}$. Without this requirement we could simply choose $\boldsymbol{P}(H|\mathcal{F}) = \mathbf{1}_H$ and (4.2) would be satisfied. However, note that if $H \in \mathcal{F}$, then the function $\mathbf{1}_F$ is $\mathcal{F}$-measurable, and in this case $\boldsymbol{P}(H \mid \mathcal{F}) = \mathbf{1}_F$.

REMARK 4.4. In general we can only expect (4.2) to hold for all events in $\mathcal{F}$, and it need not hold for events in $\mathcal{G}$! Indeed, in the example above we see that

$$\int_H \boldsymbol{P}(H \mid \mathcal{C}) \, d\boldsymbol{P} = \frac{1}{2}\boldsymbol{P}(R \cap H) + \frac{1}{5}\boldsymbol{P}(B \cap H) = \frac{1}{2} \cdot \frac{5}{30} + \frac{1}{5} \cdot \frac{4}{30} = \frac{11}{100}$$

but

$$\boldsymbol{P}(H \cap H) = \boldsymbol{P}(H) = \frac{3}{10} \neq \frac{11}{100}.$$

REMARK 4.5. One situation where you can compute $\boldsymbol{P}(A|\mathcal{F})$ explicitly is when $\mathcal{F} = \sigma(\{F_i\})$ where $\{F_i\}$ is a pairwise disjoint collection of events whose union is all of Ω and $\boldsymbol{P}(F_i) > 0$ for all i. In this case

$$\boldsymbol{P}(A \mid \mathcal{F}) = \sum_i \frac{\boldsymbol{P}(A \cap F_i)}{\boldsymbol{P}(F_i)} \mathbf{1}_{F_i}.$$

4.2. Conditional expectation. Consider now the situation where X is a $\mathcal{G}$-measurable random variable and $\mathcal{F} \subseteq \mathcal{G}$ is some σ-sub-algebra. The conditional expectation of X given $\mathcal{F}$ (denoted by $\boldsymbol{E}(X \mid \mathcal{F})$ is the "best approximation" of X by a $\mathcal{F}$ measurable random variable.

Consider the incomplete deck example from the previous section, where you have an incomplete deck of cards which has 10 red cards (of which 5 are high), and 20 black cards (of which 4 are high). Let X be the outcome of a game played through a dealer who pays you \$1 when a high card is drawn, and charges you \$1 otherwise. *However, the dealer only tells you the color of the card drawn and your winnings, and not the rules of the game or whether the card was high.*

After playing this game often the only information you can deduce is that your expected return is 0 when a red card is drawn and $-3/5$ when a black card is drawn. That is, you approximate the game by the random variable

$$Y \stackrel{\text{def}}{=} 0\mathbf{1}_R - \frac{3}{5}\mathbf{1}_B,$$

where, as before R, B denote the set of all red and black cards respectively.

Note that the events you can deduce information about by playing this game (through the dealer) are exactly elements of the σ-algebra $\mathcal{C} = \{\emptyset, R, B, \Omega\}$. By

construction, that your approximation Y is $\mathcal{C}$-measurable, and has the same averages as X on all elements of $\mathcal{C}$. That is, for every $C \in \mathcal{C}$, we have

$$\int_C Y \, d\boldsymbol{P} = \int_C X \, d\boldsymbol{P}\,.$$

This is how we define conditional expectation.

DEFINITION 4.6. Let X be a $\mathcal{G}$-measurable random variable, and $\mathcal{F} \subseteq \mathcal{G}$ be a σ-sub-algebra. We define $\boldsymbol{E}(X \mid \mathcal{F})$, the *conditional expectation of X given $\mathcal{F}$* to be a *random variable* such that:

(1) $\boldsymbol{E}(X \mid \mathcal{F})$ is $\mathcal{F}$-measurable.
(2) For every $F \in \mathcal{F}$, we have the *partial averaging* identity:

(4.3) $$\int_F \boldsymbol{E}(X \mid \mathcal{F}) \, d\boldsymbol{P} = \int_F X \, d\boldsymbol{P}.$$

REMARK 4.7. Choosing $F = \Omega$ we see $\boldsymbol{E}\boldsymbol{E}(X \mid \mathcal{F}) = \boldsymbol{E}X$.

REMARK 4.8. Note we can only expect (4.3) to hold for all events $F \in \mathcal{F}$. In general (4.3) *will not* hold for events $G \in \mathcal{G} - \mathcal{F}$.

REMARK 4.9. Under mild integrability assumptions one can show that conditional expectations exist. This requires the *Radon-Nikodym* theorem and goes beyond the scope of this course. If, however, $\mathcal{F} = \sigma(\{F_i\})$ where $\{F_i\}$ is a pairwise disjoint collection of events whose union is all of Ω and $\boldsymbol{P}(F_i) > 0$ for all i, then

$$\boldsymbol{E}(X \mid \mathcal{F}) = \sum_{i=1}^{\infty} \frac{\mathbf{1}_{F_i}}{\boldsymbol{P}(F_i)} \int_{F_i} X \, d\boldsymbol{P}\,.$$

REMARK 4.10. Once existence is established it is easy to see that conditional expectations are unique. Namely, if Y is any $\mathcal{F}$-measurable random variable that satisfies

$$\int_F Y \, d\boldsymbol{P} = \int_F X \, d\boldsymbol{P} \quad \text{for every } F \in \mathcal{F}\,,$$

then $Y = \boldsymbol{E}(X \mid \mathcal{F})$. Often, when computing the conditional expectation, we will "guess" what it is, and verify our guess by checking measurablity and the above partial averaging identity.

PROPOSITION 4.11. *If X is $\mathcal{F}$-measurable, then $\boldsymbol{E}(X \mid \mathcal{F}) = X$. On the other hand, if X is independent[3] of $\mathcal{F}$ then $\boldsymbol{E}(X \mid \mathcal{F}) = \boldsymbol{E}X$.*

PROOF. If X is $\mathcal{F}$-measurable, then clearly the random variable X is both $\mathcal{F}$-measurable and satisfies the partial averaging identity. Thus by uniqueness, we must have $\boldsymbol{E}(X \mid \mathcal{F}) = X$.

Now consider the case when X is independent of $\mathcal{F}$. Suppose first $X = \sum a_i \mathbf{1}_{A_i}$ for finitely many sets $A_i \in \mathcal{G}$. Then for any $F \in \mathcal{F}$,

$$\int_F X \, d\boldsymbol{P} = \sum a_i \boldsymbol{P}(A_i \cap F) = \boldsymbol{P}(F) \sum a_i \boldsymbol{P}(A_i) = \boldsymbol{P}(F)\boldsymbol{E}X = \int_F \boldsymbol{E}X \, d\boldsymbol{P}\,.$$

Thus the constant random variable $\boldsymbol{E}X$ is clearly $\mathcal{F}$-measurable and satisfies the partial averaging identity. This forces $\boldsymbol{E}(X \mid \mathcal{F}) = \boldsymbol{E}X$. The general case when X is not simple follows by approximation. □

The above fact has a generalization that is tremendously useful when computing conditional expectations. Intuitively, the general principle is to *average* quantities that are independent of $\mathcal{F}$, and *leave unchanged* quantities that are $\mathcal{F}$ measurable. This is known as the independence lemma.

LEMMA 4.12 (Independence Lemma). *Suppose X, Y are two random variables such that X is independent of $\mathcal{F}$ and Y is $\mathcal{F}$-measurable. Then if $f = f(x, y)$ is any function of two variables we have*

$$\boldsymbol{E}\big(f(X, Y) \,\big|\, \mathcal{F}\big) = g(Y)\,,$$

where $g = g(y)$ is the function[4] defined by

$$g(y) \overset{\text{def}}{=} \boldsymbol{E}f(X, y)\,.$$

REMARK. If p_X is the probability density function of X, then the above says

$$\boldsymbol{E}\big(f(X, Y) \,\big|\, \mathcal{F}\big) = \int_{\mathbb{R}} f(x, Y)\, p_X(x) \, dx\,.$$

Indicating the ω dependence explicitly for clarity, the above says

$$\boldsymbol{E}\big(f(X, Y) \,\big|\, \mathcal{F}\big)(\omega) = \int_{\mathbb{R}} f(x, Y(\omega))\, p_X(x) \, dx\,.$$

REMARK 4.13. Note we defined and motivated conditional expectations and conditional probabilities independently. They are however intrinsically related: Indeed, $\boldsymbol{E}(\mathbf{1}_A \mid \mathcal{F}) = \boldsymbol{P}(A \mid \mathcal{F})$, and this can be checked directly from the definition.

Conditional expectations are tremendously important in probability, and we will encounter it often as we progress. This is probably the first visible advantage of the measure theoretic approach, over the previous "intuitive" or discrete approach to probability.

PROPOSITION 4.14. *Conditional expectations satisfy the following properties.*

(1) (Linearity) If X, Y are random variables, and $\alpha \in \mathbb{R}$ then

$$\boldsymbol{E}(X + \alpha Y \mid \mathcal{F}) = \boldsymbol{E}(X \mid \mathcal{F}) + \alpha \boldsymbol{E}(Y \mid \mathcal{F})\,.$$

(2) (Positivity) If $X \leqslant Y$, then $\boldsymbol{E}(X \mid \mathcal{F}) \leqslant \boldsymbol{E}(Y \mid \mathcal{F})$ (almost surely).
(3) If X is $\mathcal{F}$ measurable and Y is an arbitrary (not necessarily $\mathcal{F}$-measurable) random variable then (almost surely)

$$\boldsymbol{E}(XY \mid \mathcal{F}) = X\boldsymbol{E}(Y \mid \mathcal{F})\,.$$

(4) (Tower property) If $\mathcal{E} \subseteq \mathcal{F} \subseteq \mathcal{G}$ are σ-algebras, then (almost surely)

$$\boldsymbol{E}(X \mid \mathcal{E}) = \boldsymbol{E}\Big(\boldsymbol{E}(X \mid \mathcal{F}) \,\Big|\, \mathcal{E}\Big)\,.$$

---

[3]We say a random variable X is independent of σ-algebra $\mathcal{F}$ if for every $A \in \sigma(X)$ and $B \in \mathcal{F}$ the events A and B are independent.

[4]To clarify, we are defining a function $g = g(y)$ here when $y \in \mathbb{R}$ is any real number. Then, once we compute g, we substitute in $y = Y(= Y(\omega))$, where Y is the given random variable.

PROOF. The first property follows immediately from linearity. For the second property, set $Z = Y - X$ and observe

$$\int_{\boldsymbol{E}(Z \mid \mathcal{F})} \boldsymbol{E}(Z \mid \mathcal{F}) \, d\boldsymbol{P} = \int_{\boldsymbol{E}(Z \mid \mathcal{F})} Z \, d\boldsymbol{P} \geqslant 0 \,,$$

which can only happen if $\boldsymbol{P}(\boldsymbol{E}(Z \mid \mathcal{F}) < 0) = 0$. The third property is easily checked for simple random variables, and follows in general by approximating. The tower property follows immediately from the definition. $\square$

4.3. Adapted processes and filtrations. Let X be any stochastic process (for example Brownian motion). For any $t > 0$, we've seen before that $\sigma(X(t))$ represents the information you obtain by observing $X(t)$. Accumulating this over time gives us the *filtration*.

DEFINITION 4.15. The *filtration generated X* is the family of σ-algebras $\{\mathcal{F}_t^X \mid t \geqslant 0\}$ where

$$\mathcal{F}_t^X \stackrel{\text{def}}{=} \sigma\Big(\bigcup_{s \leqslant t} \sigma(X_s)\Big) \,.$$

Clearly each $\mathcal{F}_t^X$ is a σ-algebra, and if $s \leqslant t$, $\mathcal{F}_s^X \subseteq \mathcal{F}_t^X$. A family of σ-algebras with this property is called a *filtration*.

DEFINITION 4.16. A *filtration* is a family of σ-algebras $\{\mathcal{F}_t \mid t \geqslant 0\}$ such that whenever $s \leqslant t$, we have $\mathcal{F}_s \subseteq \mathcal{F}_t$.

The σ-algebra $\mathcal{F}_t$ represents the information accumulated up to time t. When given a filtration, it is important that all stochastic processes we construct respect the flow of information because trading / pricing strategies can not rely on the price at a later time, and gambling strategies do not know the outcome of the next hand. This is called *adapted*.

DEFINITION 4.17. A stochastic process X is said to be *adapted* to a filtration $\{\mathcal{F}_t \mid t \geqslant 0\}$ if for every t the random variable $X(t)$ is $\mathcal{F}_t$ measurable (i.e. $\{X(t) \leqslant \alpha\} \in \mathcal{F}_t$ for every $\alpha \in \mathbb{R}$, $t \geqslant 0$).

Clearly a process X is adapted with respect to the filtration it generates $\{\mathcal{F}_t^X\}$.

4.4. Martingales. Recall, a martingale is a "fair game". Using conditional expectations, we can now define this precisely.

DEFINITION 4.18. A stochastic process M is a martingale with respect to a filtration $\{\mathcal{F}_t\}$ if:
   (1) M is adapted to the filtration $\{\mathcal{F}_t\}$.
   (2) For any $s < t$ we have $\boldsymbol{E}(M(t) \mid \mathcal{F}_s) = M(s)$, almost surely.

REMARK 4.19. A *sub-martingale* is an adapted process M for which we have $\boldsymbol{E}(M(t) \mid \mathcal{F}_s) \geqslant M(s)$, and a *super-martingale* if $\boldsymbol{E}(M(t) \mid \mathcal{F}_s) \leqslant M(s)$. Thus $\boldsymbol{E}M(t)$ is an increasing function of time if M is a sub-martingale, constant in time if M is a martingale, and a decreasing function of time if M is a super-martingale.

REMARK 4.20. It is crucial to specify the filtration when talking about martingales, as it is certainly possible that a process is a martingale with respect to one filtration but not with respected to another. For our purposes the filtration will almost always be the *Brownian filtration* (i.e. the filtration generated by Brownian motion).

EXAMPLE 4.21. Let $\{\mathcal{F}_t\}$ be a filtration, $\mathcal{F}_\infty = \sigma(\cup_{t \geqslant 0} \mathcal{F}_t)$, and X be any $\mathcal{F}_\infty$-measurable random variable. The process $M(t) \stackrel{\text{def}}{=} \boldsymbol{E}(X_\infty \mid \mathcal{F}_t)$ is a martingale with respect to the filtration $\{\mathcal{F}_t\}$.

In discrete time a random walk is a martingale, so it is natural to expect that in continuous time Brownian motion is a martingale as well.

THEOREM 4.22. *Let W be a Brownian motion, $\mathcal{F}_t = \mathcal{F}_t^W$ be the Brownian filtration. Brownian motion is a martingale with respect to this filtration.*

PROOF. By independence of increments, $W(t) - W(s)$ is certainly independent of $W(r)$ for any $r \leqslant s$. Since $\mathcal{F}_s = \sigma(\cup_{r \leqslant s} \sigma(W(r)))$ we expect that $W(t) - W(s)$ is independent of $\mathcal{F}_s$. Consequently

$$\boldsymbol{E}(W(t) \mid \mathcal{F}_s) = \boldsymbol{E}(W(t) - W(s) \mid \mathcal{F}_s) + \boldsymbol{E}(W(s) \mid \mathcal{F}_s) = 0 + W(s) = W(s) \,. \quad \square$$

THEOREM 4.23. *Let W be a standard Brownian motion (i.e. a Brownian motion normalized so that $W(0) = 0$ and $\mathrm{Var}(W(t)) = t$). For any $C_b^{1,2}$ function[5] $f = f(t, x)$ the process*

$$M(t) \stackrel{\text{def}}{=} f(t, W(t)) - \int_0^t \Big(\partial_t f(s, W(s)) + \frac{1}{2} \partial_x^2 f(s, W(s))\Big) \, ds$$

is a martingale (with respect to the Brownian filtration).

PROOF. This is an extremely useful fact about Brownian motion follows quickly from the Itô formula, which we will discuss later. However, at this stage, we can provide a simple, elegant and instructive proof as follows.

Adaptedness of M is easily checked. To compute $\boldsymbol{E}(M(t) \mid \mathcal{F}_r)$ we first observe

$$\boldsymbol{E}\big(f(t, W(t)) \mid \mathcal{F}_r\big) = \boldsymbol{E}\big(f(t, [W(t) - W(r)] + W(r)) \mid \mathcal{F}_r\big).$$

Since $W(t) - W(r) \sim N(0, t - r)$ and is independent of $\mathcal{F}_r$, the above conditional expectation can be computed by

$$\boldsymbol{E}\big(f(t, [W(t) - W(r)] + W(r)) \mid \mathcal{F}_r\big) = \int_{\mathbb{R}} f(t, y + W(r)) G(t - r, y) \, dy \,,$$

where

$$G(\tau, y) = \frac{1}{\sqrt{2\pi\tau}} \exp\Big(\frac{-y^2}{2\tau}\Big)$$

is the density of $W(t) - W(r)$.

---

[5]Recall a function $f = f(t, x)$ is said to be $C^{1,2}$ if it is C^1 in t (i.e. differentiable with respect to t and $\partial_t f$ is continuous), and C^2 in x (i.e. twice differentiable with respect to x and $\partial_x f$, $\partial_x^2 f$ are both continuous). The space $C_b^{1,2}$ refers to all $C^{1,2}$ functions f for which and f, $\partial_t f$, $\partial_x f$, $\partial_x^2 f$ are all bounded functions.

Similarly

$$\boldsymbol{E}\Big(\int_0^t \big(\partial_t f(s, W(s)) + \tfrac{1}{2}\partial_x^2 f(s, W(s))\big)\, ds \,\Big|\, \mathcal{F}_r\Big)$$

$$= \int_0^r \big(\partial_t f(s, W(s)) + \tfrac{1}{2}\partial_x^2 f(s, W(s))\big)\, ds$$

$$+ \int_r^t \int_{\mathbb{R}} \big(\partial_t f(s, y + W(r)) + \tfrac{1}{2}\partial_x^2 f(s, y + W(r))\big) G(s - r, y)\, dy\, ds$$

Hence

$$\boldsymbol{E}(M(t) \mid \mathcal{F}_r) - M(r) = \int_{\mathbb{R}} f(t, y + W(r)) G(t - r, y)\, dy$$

$$- \int_r^t \int_{\mathbb{R}} \big(\partial_t f(s, y + W(r)) + \tfrac{1}{2}\partial_x^2 f(s, y + W(r))\big) G(s - r, y)\, ds$$

$$- f(r, W(r))\,.$$

We claim that the right hand side of the above vanishes. In fact, we claim the (deterministic) identity

$$f(r, x) = \int_{\mathbb{R}} f(t, y + x) G(t - r, y)\, dy$$

$$- \int_r^t \int_{\mathbb{R}} \big(\partial_t f(s, y + x) + \tfrac{1}{2}\partial_x^2 f(s, y + x)\big) G(s - r, y)\, dy\, ds$$

holds for any function f and $x \in \mathbb{R}$. For those readers who are familiar with PDEs, this is simply the Duhamel's principle for the heat equation. If you're unfamiliar with this, the above identity can be easily checked using the fact that $\partial_\tau G = \frac{1}{2}\partial_y^2 G$ and integrating the first integral by parts. We leave this calculation to the reader. $\square$

4.5. Stopping Times. For this section we assume that a filtration $\{\mathcal{F}_t\}$ is given to us, and fixed. When we refer to process being adapted (or martingales), we implicitly mean they are adapted (or martingales) with respect to this filtration.

Consider a game (played in continuous time) where you have the option to walk away at any time. Let τ be the *random* time you decide to stop playing and walk away. In order to respect the flow of information, you need to be able to decide weather you have stopped using only information up to the present. At time t, event $\{\tau \leqslant t\}$ is exactly when you have stopped and walked away. Thus, to respect the flow of information, we need to ensure $\{\tau \leqslant t\} \in \mathcal{F}_t$.

DEFINITION 4.24. A stopping time is a function $\tau \colon \Omega \to [0, \infty)$ such that for every $t \geqslant 0$ the event $\{\tau \leqslant t\} \in \mathcal{F}_t$.

A standard example of a stopping time is *hitting times*. Say you decide to liquidate your position once the value of your portfolio reaches a certain threshold. The time at which you liquidate is a hitting time, and under mild assumptions on the filtration, will always be a stopping time.

PROPOSITION 4.25. *Let X be an adapted continuous process, $\alpha \in \mathbb{R}$ and τ be the first time X hits α (i.e. $\tau = \inf\{t \geqslant 0 \mid X(t) = \alpha\}$). Then τ is a stopping time (if the filtration is right continuous).*

THEOREM 4.26 (Doob's optional sampling theorem). *If M is a martingale and τ is a bounded stopping time. Then the stopped process $M^\tau(t) \stackrel{\text{def}}{=} M(\tau \wedge t)$ is also a martingale. Consequently, $\boldsymbol{E}M(\tau) = \boldsymbol{E}M(\tau \wedge t) = \boldsymbol{E}M(0) = \boldsymbol{E}M(t)$ for all $t \geqslant 0$.*

REMARK 4.27. If instead of assuming τ is bounded, we assume M^τ is bounded the above result is still true.

The proof goes beyond the scope of these notes, and can be found in any standard reference. What this means is that if you're playing a fair game, then you can not hope to improve your odds by "quitting when you're ahead". Any rule by which you decide to stop, must be a stopping time and the above result guarantees that stopping a martingale still yields a martingale.

REMARK 4.28. Let W be a standard Brownian motion, τ be the first hitting time of W to 1. Then $\boldsymbol{E}W(\tau) = 1 \neq 0 = \boldsymbol{E}W(t)$. This is one situation where the optional sampling theorem doesn't apply (in fact, $\boldsymbol{E}\tau = \infty$, and W^τ is unbounded).

This example corresponds to the gambling strategy of walking away when you make your "million". The reason it's not a sure bet is because the time taken to achieve your winnings is finite almost surely, but very long (since $\boldsymbol{E}\tau = \infty$). In the mean time you might have incurred financial ruin and expended your entire fortune.

Suppose the price of a security you're invested in fluctuates like a martingale (say for instance Brownian motion). This is of course unrealistic, since Brownian motion can also become negative; but lets use this as a first example. You decide you're going to liquidate your position and walk away when either you're bankrupt, or you make your first million. What are your expected winnings? This can be computed using the optional sampling theorem.

PROBLEM 4.1. Let $a \geqslant 0$ and M be any continuous martingale with $M(0) = x \in (0, a)$. Let τ be the first time M hits either 0 or a. Compute $\boldsymbol{P}(M(\tau) = a)$ and your expected return $\boldsymbol{E}M(\tau)$.