STOCHASTIC CALCULUS

Amit Rajaraman

Last updated March 23, 2021

Contents

§1. Conditional Expectations and Martingales

1.1. Conditional Expectations

To motivate the general definition of conditional expectation, it is helpful to first work in a discrete setting.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X, Z be random variables such that each of them takes some finite set of values $\{x_1, \ldots, x_n\}$ and $\{z_1, \ldots, z_m\}$ respectively. What would the conditional expectation be if, say, $Z = z_j$? It is given by

$$\mathbf{E}[X \mid Z = z_j] = \sum x_i \Pr[X = x_i \mid Z = z_j].$$

Denote the above random variable by Y – if $Z(\omega) = z_j$, then $Y(\omega)$ is equal to the above quantity. Note that Y is constant on each Z-atom, and as a result, is measurable with respect to the σ-algebra $\mathcal{G}$ consisting of the 2^m possible unions of Z-atoms. Suppose Y takes value y_j on $\{Z = z_j\}$. Then,

$$\int_{\{Z=z_j\}} Y \, dP = y_j \Pr[Z = z_j] = \sum_j x_i \Pr[X = x_i \mid Z = z_j] \Pr[Z = z_j] = \sum_j \Pr[X = x_j, Z = z_j] = \int_{\{} Z = z_j\} X \, dP.$$

That is, for any $G \in \mathcal{G}$,

$$\int_G X \, dP = \int_G Z \, dP.$$

This is exactly the motivation behind the general definition.

Definition 1.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable with $\mathbf{E}[\|X\|] < \infty$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then, there exists a random variable Y such that

- Y is $\mathcal{G}$-measurable,

- $\mathbf{E}[\|Y\|] < \infty$, and

- for every $G \in \mathcal{G}$, $\int_G Y \, dP = \int_G X \, dP$.

Moreover, if Y and $\tilde{Y}$ are two random variables satisfying the above, then $Y = \tilde{Y}$ almost surely. Such a random variable Y is called the *conditional expectation* $\mathbf{E}[X \mid \mathcal{G}]$ of X given G, and we write $Y = \mathbf{E}[X \mid \mathcal{G}]$ almost surely.

The key insight here is that instead of viewing a random variable as something that takes values, we can view it as a partitioning of the space into equivalence classes, and defining the new random variable based on this partition instead. We do not care about *what* values Z takes, we only care about the probability that Z takes a certain value! Since the conditional expectation is almost surely unique, we mean by "the conditional expectation" any such version of a conditional expectation.

For a random variable Z, $\mathbf{E}[X \mid Z]$ is used to denote $\mathbf{E}[X \mid \sigma(Z)]$.

Now, if $F \in \mathcal{F}$ and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then we define $\Pr[F \mid \mathcal{G}]$ as $\mathbf{E}[\mathbb{1}_F \mid \mathcal{G}]$.

Lemma 1.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable in $\mathcal{L}^1$. Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$.

(i) If $Y = \mathbf{E}[X \mid \mathcal{G}]$, then $\mathbf{E}[Y] = \mathbf{E}[X]$.

(ii) If X is $\mathcal{G}$-measurable, then $\mathbf{E}[X \mid \mathcal{G}] = X$ almost surely.

(iii) $\mathbf{E}[a_1 X_1 + a_2 X_2 \mid \mathcal{G}] = a_1 \mathbf{E}[X_1 \mid \mathcal{G}] + a_2 \mathbf{E}[X_2 \mid \mathcal{G}]$ almost surely.

(iv) If $X \geq 0$, then $\mathbf{E}[X \mid \mathcal{G}] \geq 0$ almost surely.

(v) If $0 \leq X_n \uparrow X$, then $\mathbf{E}[X_n \mid \mathcal{G}] \uparrow \mathbf{E}[X \mid \mathcal{G}]$ almost surely.

(vi) If $X_n \geq 0$, then $\mathbf{E}[\liminf X_n \mid G] \leq \liminf \mathbf{E}[X_n \mid G]$ almost surely.

(vii) If $|X_n(\omega)| \leq V(\omega)$ for all n, $\mathbf{E}[V] < \infty$, and $X_n \to X$ almost surely, then $\mathbf{E}[X_n \mid \mathcal{G}] \to \mathbf{E}[X \mid \mathcal{G}]$ almost surely.

(viii) If $c : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbf{E}[|c(X)|] < \infty$, then $\mathbf{E}[c(X) \mid \mathcal{G}] \geq c(\mathbf{E}[X \mid \mathcal{G}])$. In particular, $\|\mathbf{E}[X \mid \mathcal{G}]\|_p \leq \|X\|_p$ almost surely.

(ix) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbf{E}[X \mid \mathcal{H}]$ almost surely. We often shorten this as $\mathbf{E}[X \mid \mathcal{G} \mid \mathcal{H}]$.

(x) If Z is $\mathcal{G}$-measurable and bounded, then $\mathbf{E}[ZX \mid \mathcal{G}] = Z\mathbf{E}[X \mid \mathcal{G}]$ almost surely. This also holds if $p > 1$, $p^{-1} + q^{-1} = 1$, $X = \mathcal{L}^p(\Omega, \mathcal{F}, P)$, and $Z = \mathcal{L}^q(\Omega, \mathcal{G}, p)$.

(xi) If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then $\mathbf{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbf{E}[X \mid \mathcal{G}]$ almost surely. In particular, if X is independent of $\mathcal{H}$, then $\mathbf{E}[X \mid \mathcal{H}] = \mathbf{E}[X]$ almost surely.

For future purposes, we also state four types of convergence of random variables, which the reader is hopefully familiar with.

Definition 1.2. Let X be a random variable and (X_n) be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$.

(a) $X_n \to X$ *almost surely* if $\Pr[\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}] = 1$.

(b) $X_n \to X$ *in probability* if $\Pr[|X_n - X| > \varepsilon] \to 0$ for every $\varepsilon > 0$.

(c) $X_n \to X$ *in $\mathcal{L}^p$* if $\|X_n - X\|_p \to 0$.

(d) $X_n \to X$ *in law* or *in distribution* if $\mathbf{E}[f(X_n)] \to \mathbf{E}[f(X)]$ for any bounded continuous function f.

We also have that (a),(c)$\Rightarrow$(b)$\Rightarrow$(d). Also, if $q \leq p$ and $X_n \to X$ in $\mathcal{L}^p$, then $X_n \to X$ in $\mathcal{L}^q$.

1.2. Discrete Time Processes

Before we get to stochastic processes, we must build a notion of "time" in a probability space. First, let us do so in a discrete setting.

Definition 1.3. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A (discrete time) *filtration* is an increasing sequence $(\mathcal{F}_n)$ of σ-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$. The quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ is called a *filtered probability space*.

$\mathcal{F}_n$ is essentially the "set of questions" we can ask at time n. For example, if we are flipping a coin and are at time n, we can ask and get the answer to "Did the coin flips up to time n give more heads than tails?" but we cannot know the answer to "Did the coin flips up to time $N > n$ give more heads than tails?". We can of course determine the probability of the latter case, but we cannot answer it with certainty.

Definition 1.4. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ be a filtered probability space. A process (X_n) is called $\mathcal{F}_n$-*adapted* if X_n is $\mathcal{F}_n$-measurable for each n, and $\mathcal{F}_n$-*predictable* if X_n is $\mathcal{F}_{n-1}$-measurable for each n.

We often refer to them as just adapted or predictable if it is clear from context what the filtration is.
In the adapted case, X_n represents something in the past or present, whereas in the predictable case, X_n represents something in the past.
Similarly, we can generate a filtration given a process – this is what we shall use more often.

Definition 1.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space and (X_n) be a process. The *filtration generated* by (X_n) is defined as $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$ for each n.

Note that the process X_n is $\mathcal{F}_n^X$-adapted by construction.
Even if some sequences don't explicitly have any relation to the (physical) notion of time, it is often useful to think of them as an adapted stochastic process – for example, a sequence of approximations $\mathbf{E}[X \mid Y_n]$ to $\mathbf{E}[X \mid Y]$.

1.3. Martingales

Definition 1.6. A process (X_n) is said to be a $\mathcal{F}_n$-*martingale* if it is $\mathcal{F}_n$-adapted and it satisfies $\mathbf{E}[X_n \mid \mathcal{F}_m] = X_m$ almost surely for every $m \leq n$.

We often refer to it as just a martingale if the corresponding filtration is obvious.

Observe that the final condition is equivalent to $\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$.

The nice name that martingales have comes from a betting strategy. We can think of playing a sequence of games, wherein X_n represents our total winnings after the nth game. In general of course, there need not be any relation between the X_i. In a fair setting however, we should make (and lose) no money on average. If we make money, then it is unfair towards the casino, and if we lose money, then it is unfair towards us (most casinos are unfair towards us). That is, if we have totally won X_m dollars at time m, then our expected winnings at some time point $n > m$ should be $\mathbf{E}[X_n \mid \sigma\{X_0, \ldots, X_m\}] = X_m$ as well!

Martingales turn out to be surprisingly pervasive in probability theory and are seen in several places. To begin with, let us relate martingales a bit more to the above betting strategy, while at the same time giving some intuition as to why martingales are important.

Lemma 1.2 (Doob Decomposition). Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(\mathcal{F}_n)$ be a filtration, and (X_n) be $\mathcal{F}_n$-adapted with $X_n \in \mathcal{L}^1$ for every n. Then, we can write $X_n = X_0 + A_n + M_n$ almost surely, where (A_n) is $\mathcal{F}_n$-predictable and M_n is a $\mathcal{F}_n$-martingale with $M_0 = 0$. Further, this decomposition is unique.

This is not too difficult to prove on setting $A_n = \sum_{k=1}^{n} \mathbf{E}[X_k - X_{k-1} \mid \mathcal{F}_{k-1}]$.

Definition 1.7. Let (M_n) be a martingale and (A_n) be a predictable process. Then $(A \cdot M)_n = \sum_{k=1}^{n} A_k(M_k - M_{k-1})$, known as the *martingale transform* of M by A, is a martingale, provided that A_n and $(A \cdot M)_n$ are in $\mathcal{L}^1$ for all n.

Keeping in line with our gambling analogy, here, M_n would represent our total winnings at time n if we were to stake one dollar in the game and A_k represents the number of dollars we stake. The A_n should then be predictable because we place our bet *before* the game occurs.

Let us now consider a particular betting strategy, which seems to go against the intuition built thus far for a martingale. First, choose some $a < b$. We play against a single friend[1] of ours. Whenever our own capital sinks below a, we bet one dollar until it exceeds b dollars. At this point, we stop betting money and wait until our capital sinks below a again.

In each "round" from a to b, we make $(b-a)$ dollars. This seems contradictory however! This would mean that if we played the game for a long time, we would repeatedly make $(b-a)$ dollars and become rich. However, any betting strategy should be a martingale – what are we missing?

If we face k reversals of fortune, we will make $k(b-a)$ dollars. The only part we have not considered is when we hit a dollars for the $(k+1)$th time but have not hit b dollars for the $(k+1)$th time yet. Since it must balance out, we must make a significant loss before we reach the next b. That is, the expected loss incurred from the last time we started staking money (when we hit a for the $(k+1)$th time) should be equal to $k(b-a)$.

The only logical conclusion in the case where the M_n is bounded is that we can only cross a and b a finite number of times! Otherwise, we could (expect to) make money by playing with the above strategy. That is, the M_n must converge to some random variable!

Lemma 1.3 (Doob's Upcrossing Lemma). Let (M_n) be a martingale and denote by $U_n(a, b)$ the number of "up-crossings" of $a < b$ up to time n. Then, $\mathbf{E}[U_n(a, b)] \leq \mathbf{E}[(a - M_n)^+]/(b - a)$.

Proof. This is easily proved as in the previous paragraph. Set $C_0 = 0$ and for $k > 0$, $C_k = \mathbb{1}_{C_{k-1}=1}\mathbb{1}_{M_{k-1}<b} + \mathbb{1}_{C_{k-1}=0}\mathbb{1}_{M_{k-1}\leq a}$. Let $X_n = (C \cdot M)_n$. Then $X_n \geq (b - a)U_n(a, b) - (a - M_n)^+$ (Why?). However, X_n is a martingale, so $\mathbf{E}[X_n] = X_0 = 0$, proving the result. ∎

---

[1] They may not be your friend at the end of the game, however.

We now come to perhaps the most important result in martingale theory.

Theorem 1.4 (Martingale Convergence Theorem). Let (M_n) be a $\mathcal{F}_n$-martingale such that $\sup_n \mathbf{E}[|M_n|] < \infty$, $\sup_n \mathbf{E}[(M_n)^+] < \infty$, or $\sup_n \mathbf{E}[(M_n)^-] < \infty$. Then, there exists a $\mathcal{F}_\infty$-measurable random variable $M_\infty \in \mathcal{L}^1$, where $\mathcal{F}_\infty = \sigma(\mathcal{F}_\setminus : n \in \mathbb{N})$, such that $M_n \to M_\infty$ almost surely.

Proof. It is not too difficult to show that the three conditions (involving expectation) are equivalent.
Let $\omega \in \Omega$. First of all, we claim that that $M_n(\omega)$ cannot have distinct $\limsup$ and $\liminf$ (so it converges) – if we choose some rational a, b such that $\liminf M_n(\omega) < a < b < \limsup M_n(\omega)$, then there would be infinitely many upcrossings. However,

$$\Pr[\exists a, b \in \mathbb{Q} \text{ such that } M_n \text{ crosses } a, b \text{ i.o.}] \leq \sum_{a,b \in \mathbb{Q}} \Pr[M_n \text{ crosses } a, b \text{ i.o.}] = \sum_{a,b \in \mathbb{Q}} \Pr[U_\infty(a,b) = \infty] = 0,$$

where the last step follows from using the Monotone Convergence Theorem on $\mathbf{E}[U_n(a,b)] \leq (|a| + \sup_n \mathbf{E}[|M_n|])/(b - a) < \infty$, so $\mathbf{E}[U_\infty(a,b)] < \infty$.
Therefore, M_n converges to some M_∞ almost surely. Using Fatou's Lemma, $\mathbf{E}[M_\infty] \leq \liminf \mathbf{E}[|M_n|] < \infty$, so M_∞ is both almost surely finite and in $\mathcal{L}^1$. ∎

Similar to martingales, we define the following.

Definition 1.8. A $\mathcal{F}_n$-adapted process (X_n) is said to be a *supermartingale* if it satisfies $\mathbf{E}[X_n \mid \mathcal{F}_m] \leq X_m$ almost surely for every $m \leq n$ and a *submartingale* if it satisfies $\mathbf{E}[X_n \mid \mathcal{F}_m] \geq X_m$ almost surely for every $m \leq n$.

For example, the winnings in casinos usually form a supermartingale – they thrive on the fact that you are expected to lose money.
The analogue of Lemma 1.2 for supermartingales is the following.

Lemma 1.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(\mathcal{F}_n)$ be a filtration, and (X_n) be $\mathcal{F}_n$-adapted with $X_n \in \mathcal{L}^1$ for every n. Then, we can write $X_n = X_0 + A_n + M_n$ almost surely, where (A_n) is a non-increasing predictable process and M_n is a martingale with $M_0 = 0$. Further, this decomposition is unique.

Lemma 1.6. Let M_n be an almost surely non-negative supermartingale and $k > 0$. Then

$$\Pr\left[\sup_n M_n \geq k\right] \leq \frac{\mathbf{E}[M_0]}{k}.$$

1.4. Stopping Times

Definition 1.9. A *stopping time* is a random time $\tau : \Omega \to \{0, 1, \ldots, \infty\}$ such that $\{\omega \in \Omega : \tau(\omega) \leq n\} \in \mathcal{F}_n$ for each n.

Continuing with the gambling analogy, this essentially represents the time we leave the table. For example, quitting while we're ahead and letting $\tau = \inf\{k : X_k \geq 20\}$ is a stopping time. The definition of a stopping time just says that our decision to leave is based solely on whatever events have transpired thus far. On the other hand, one might decide to leave when we are done with our final upcrossing (as defined before Lemma 1.3). However, it is impossible to determine whether or not this is the case since we cannot look into the future to see if there are no more upcrossings.

Definition 1.10. Let (X_n) be a stochastic process and $\tau < \infty$ be a stopping time. Then X_τ denotes the random variable $X_{\tau(\omega)}(\omega)$, that is, the process X_n evaluated at τ. In general, the stochastic process $X'_n(\omega) = X_{\min\{n, \tau(\omega)\}}(\omega)$ is called the *stopped process*. That is, $X'_n = X_n$ for $n < \tau$ and $X'_n = X_\tau$ for $n \geq \tau$.

As mentioned, this represents our winnings if we stop playing the game at τ.

Lemma 1.7 (Optional Stopping). Let M_n be a martingale and $\tau < \infty$ be a stopping time. Then $\mathbf{E}[M_\tau] = \mathbf{E}[M_0]$ if any of the following holds

(a) $\tau < K$ almost surely for some $K \in \mathbb{N}$.

(b) $|M_n| \leq k$ for some $k \in [0, \infty)$ and all n.

(c) $|M_n - M_{n-1}| \leq k$ almost surely for some $k \in [0, \infty)$ and $\mathbf{E}[\tau] < \infty$.

If any of these conditions hold for a supermartingale, then $\mathbf{E}[M_\tau] \leq \mathbf{E}[M_0]$.

Let us now look at the gambling idea we gave earlier of "quitting while we are ahead". Suppose $\xi_1, \xi_2, \ldots$ are iid and each take ± 1 with probability $1/2$ each. Let $M_n = M_0 + \sum_{i=1}^k \xi_i$. It is easy to see that (M_n) forms a martingale. In the gambling context, this just means that at each time step, we gain or lose a dollar with probability $1/2$.

- First of all, note that M_n does *not* converge (Why?). To solidify the notion of quitting while ahead, consider the stopping time $\tau = \inf\{n : M_n \geq 2M_0\}$. We wait until this happens, and then stop playing.
  The first question we ask is: is $\tau < \infty$? Perhaps unintuitively, the answer is yes. Indeed, $M_{\min\{n,\tau\}} \leq 2M_0$ almost surely for all n. Therefore, we may apply the martingale convergence theorem to get that $M_{\min\{n,\tau\}}$ converges. Further, since M_n only takes integer values and changes by 1 at each step, it is not too difficult to show that convergence can only occur if it "gets stuck" at M_0, that is, $\tau < \infty$ almost surely. This seems to be a contradiction – a way to earn free money!

- Next, note that while $M_{\min\{n,\tau\}} \to 2M_0$ almost surely, it does not converge in $\mathcal{L}^1$. Indeed, the latter would imply that the expectations converge as well, but this is clearly not the case since $\mathbf{E}[M_{\min\{n,\tau\}}]$ does not converge to $\mathbf{E}[M_\tau]$. Intuitively, a process does not converge in $\mathcal{L}^1$ only when the "outliers" of the process grow very rapidly. In the current context, this would mean that while we eventually double our capital, we incur massive losses in the middle in order to keep the game fair.

- To quantify this, suppose that there is also some maximum debt amount $-R$ we can incur, so the stopping time is now $\kappa = \inf\{n : M_n \geq 2M_0 \text{ or } M_n \leq -R\}$. Since $|M_{\max\{n,\kappa\}}| \leq \max\{R, 2M_0\}$, we can use Lemma 1.7 to get that $\mathbf{E}[M_\kappa] = \mathbf{E}[M_0]$. However, M_κ takes either $-R$ or $2M_0$, so we can actually calculate the probability that each occurs.
  The conclusion is that in a fair game, no money can be made on average *unless* we allow ourselves to

  - play for an arbitrarily long amount of time and
  - go arbitrarily far into debt.

1.5. Towards Continuous Time

In the continuous setting, we usually work with stochastic processes in either a time period $[0, T]$ or $[0, \infty)$. In either case, a *stochastic process* on the probability space $(\Omega, \mathcal{F}, P)$ is a family of (measurable) random variables (X_t) indexed by time t.

Definition 1.11. Let X_t and Y_t be two stochastic processes. Then X_t and Y_t are said to be *indistinguishable* if $\Pr[X_t = Y_t \text{ for all } t] = 1$ and *modifications* of each other if $\Pr[X_t = Y_t] = 1$ for all t.

First of all, it should be mentioned that we are being slightly messy here. The event "$X_t = Y_t$ for all t" is equal to $\bigcap_t \{\omega : X_t(\omega) = Y_t(\omega)\}$, which need not even be an element of the σ-algebra! We shall assume, however, that the σ-algebra is sufficiently rich and this set is an element.
Some texts define indistinguishability slightly different and instead say that this set must contain a subset of measure 1. That is, there exists some set A such that $\Pr[A] = 1$ and for every $\omega \in A$ and t, $X_t(\omega) = Y_t(\omega)$.

Second, it should be noted that indistinguishability of two stochastic processes implies that they are modifications of each other (Why?). Further, the converse holds in the discrete setting (countably many time points) since

$$\Pr\left[\bigcap_n \{\omega : X_n(\omega) = Y_n(\omega)\}\right] \geq 1 - \sum_n \Pr\left[\{\omega : X_n(\omega) \neq Y_n(\omega)\}\right].$$

The fact that the converse does not hold in general by shown by considering the probability space as $\Omega = ([0, \infty)$ and $\mathcal{B}([0, \infty)))$, the time period as $[0, \infty)$ and the processes as

$$X_t(\omega) = \begin{cases} 1, & t = \omega, \\ 0, & \text{otherwise,} \end{cases} \quad Y_t(\omega) = 0$$

for all $t \geq 0$ and $\omega \in \Omega$.

In the continuous setting, a filtration is given by a family of σ-algebras $(\mathcal{F}_t)$ indexed by time such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$.

Definition 1.12. Let X_t be a stochastic process on some filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and time set $\mathbb{T} \subseteq [0, \infty)$ (of the form $[0, T]$ or $[0, \infty)$). Then X_t is said to be *adapted* if X_t is $\mathcal{F}_t$ measurable for all t, *measurable* if the random variable $X. : \mathbb{T} \times \Omega \to \mathbb{R}$ is $\mathcal{B}(\mathbb{T}) \times \mathcal{F}$-measurable, and *progressively measurable* if $X. : [0, t] \cap \mathbb{T} \times \Omega \to \mathbb{R}$ is $\mathcal{B}([0, t] \cap \mathbb{T}) \times \mathcal{F}$-measurable for all t.

The first definition is familiar to the reader, and the second and third essentially say that

$$Y_t = \int_0^t X_s \, \mathrm{d}s$$

is well-defined and $\mathcal{F}$-measurable (for measurable) or $\mathcal{F}_t$-measurable (for progressively measurable). In particular, progressive measurability ensures that Y_t is adapted.
There exists a result that says that every adapted measurable process has a modification that is progressively measurable, but we shall not require it.

Life is made much easier if we restrict ourselves to continuous paths, that is, the function $t \mapsto X_t(\omega)$ is continuous for every ω. Indeed, if they are continuous, then it suffices to compare them at a countable dense subset, say the rationals.

Lemma 1.8. Let X_t and Y_t be stochastic processes with continuous paths. If X_t and Y_t are modifications, then they are indistinguishable. If X_t is adapted and measurable, then it is progressively measurable.

Proof. The first result is direct on considering a countably dense subset (say the rational time points) and using the second remark after Definition 1.11.
Construct a sequence of approximate processes $X_k : [0, t] \times \Omega \to \mathbb{R}$ such that $X_t^k(\omega) = X_t(\omega)$ for all $\omega \in \Omega$ and $t = 0, 2^{-k}, \ldots, 2^{-k} \lfloor 2^k t \rfloor$ such that the sample paths are piecewise linear. Then for $s \in [0, t]$ and $\omega \in \Omega$, $X_s^k(\omega) \to X_s(\omega)$ as $k \to \infty$. Each X_k is $\mathcal{B}([0, t]) \times \mathcal{F}$-measurable. Since the limit of a sequence of measurable maps is measurable, the result follows. $\blacksquare$

Theorem 1.9 (Martingale Convergence Theorem)**.** Let M_t be a martingale and suppose that each M_t has continuous sample paths. If $\sup_t \mathbf{E}[|M_t|] < \infty$, $\sup_t \mathbf{E}[(M_t)^+] < \infty$, or $\sup_t \mathbf{E}[(M_t)^-] < \infty$, then there exists a $\mathcal{F}_\infty$-measurable random variable $M_\infty \in \mathcal{L}^1$ such that $M_t \to M_\infty$ almost surely.

1.6. The Wiener Process

Brownian motion, which is the main focus of this section, can be thought of as the limit of a random walk such that the time step and the mean square displacement both go to 0. We wish to show that this is in fact well-defined.

1.6.1. Introduction and Definitions

Let ξ_n be iid random variables each with 0 mean and unit variance and define

$$x_t(N) = \sum_{n=1}^{\lfloor Nt \rfloor} \frac{\xi_n}{\sqrt{N}}. \tag{1.1}$$

The Wiener process is just some suitable limit of this sequence. The mere existence of a stochastic process which is the limit of the above, as we have remarked already, is not obvious and we shall construct this process explicitly. The basic tool which comes to mind to do this is the central limit theorem. Unfortunately, that does not work here because we have an uncountable collection of random variables (since there is a random variable for each $t \in [0, T]$). We *can* fix the limiting distribution at some finite number of time steps $x_{t_1}(N), \ldots, x_{t_n}(N)$ however. That is,

Lemma 1.10. For any finite set of times $t_1 < \cdots < t_n$ ($n < \infty$), the n-dimensional random variable $(x_{t_1}(N), \ldots, x_{t_n}(N))$ converges in law as $N \to \infty$ to an n-dimensional random variable $(x_{t_1}, \ldots, x_{t_n})$ such that $x_{t_1}, x_{t_2} - x_{t_1}, \ldots, x_{t_n} - x_{t_{n-1}}$ are each independent Gaussian random variables with 0 mean and variance $t_1, t_2 - t_1, \ldots, t_n - t_{n-1}$.

The above is easily proved since the increments $x_{t_k}(N) - x_{t_{k-1}}(N)$ are independent for any N.
Before getting to the definition of a Wiener process, we state the following, which justifies why we can think of the Wiener process to be continuous.

Lemma 1.11. Suppose we have constructed some stochastic process x_t whose finite dimensional distributions as those of Lemma 1.10. Then, x_t has a modification $\tilde{x}_t$ such that $t \mapsto \tilde{x}_t$ is continuous.

The proof of the above is quite similar to the proof we give later of the existence of a Wiener process.

Definition 1.13. A stochastic process W_t is called a *Wiener process* if

- the finite dimensional distributions of W_t are those of Lemma 1.10 and

- the sample paths of W_t are continuous.

An $\mathbb{R}^n$-valued process $W_t = (W_t^1, \ldots, W_t^n)$ is called an *n-dimensional Wiener process* if $W_t^1, \ldots, W_t^n$ are independent Wiener processes.

1.6.2. Existence and Uniqueness

To show that a Wiener process is well-defined, we must establish existence and some sort of uniqueness. We first show uniqueness.

Lemma 1.12. If W_t and W_t' are Wiener processes, then the $C([0, \infty))$-valued random variables $W_., W_.' : \Omega \to C([0, \infty))$ have the same law.[2]

The reader might be wondering exactly what σ-algebra $\mathcal{C}$ we are taking on $C([0, \infty))$. We have two options:

- For each t, consider the evaluation map $\pi_t : C([0, \infty)) \to \mathbb{R}$, $\pi_t(x) = x_t$. Then set $\mathcal{C} = \sigma\{\pi_t : t \in [0, \infty)\}$.

- Take the natural topology on $C([0, \infty))$ as the topology of uniform convergence on compact intervals. Then, take $\mathcal{C}$ as the Borel σ-algebra with respect to this topology.

It turns out that these two σ-algebras are the same, so our intention is unambiguous.

Proof. First, we must show that $W_.$ and $W_.'$ are in fact measurable (and thus random variables). Since W_t is measurable for every t (it is a stochastic process), $W_t^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{B}(\mathbb{R})$. Note that $W_t = \pi_t \circ W_.$.
Now, by our earlier remark, $\mathcal{C} = \sigma\{\pi_t^{-1}(A) : A \in \mathcal{B}(\mathbb{R}), t \in [0, \infty)\}$, so

$$W_.^{-1}(\mathcal{C}) = \sigma\{W_t^{-1}(A) : A \in \mathcal{B}(\mathbb{R}), t \in [0, \infty)\} \subseteq \mathcal{F}.$$

Therefore, $W_.$ is a $[0, \infty)$-valued random variable (and so is $W_.'$).

Next, we must show that the two random variables have the same law, that is, they induce the same probability measure on $(C([0, \infty)), \mathcal{C})$. To do this, we use Dynkin's π-system lemma. Consider the π-system

$$\mathcal{C}_{\mathrm{cyl}} = \{\pi_{t_1}(A_1) \cap \cdots \cap \pi_{t_n}(A_n) : t_1, \ldots, t_n \in [0, \infty) \text{ and } A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})\}.$$

---

[2] Two random variables having the same law means that they induce the same probability measure on the measurable space.

Now, by definition, the finite-dimensional distributions of $W_\cdot$ and $W'_\cdot$ are equal, so the laws coincide on $\mathcal{C}_{\mathrm{cyl}}$. The required follows on using Dynkin's π-system lemma. ∎

Next, how do we show existence? First of all, note that it would suffice to construct a Wiener process on $[0,1]$ alone. We can then iterate to get it on the succeeding intervals. That is,

Lemma 1.13. Let $\{W_t : t \in [0,1]\}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$ that satisfies Definition 1.13. Then there exists a stochastic process $\{W'_t : t \in [0, \infty)\}$ on a probability space $(\Omega', \mathcal{F}', P')$ that satisfies *Definition* 1.13 for all t.

The proof follows by setting $\Omega' = \Omega \times \Omega \times \cdots$, $\mathcal{F}' = \mathcal{F} \times \mathcal{F} \times \cdots$, and $P' = P \times P \times \cdots$ with each Ω carrying iid $\{W^n_t : t \in [0,1]\}$, then checking that $W'_t = \sum_{k=1}^{\lfloor t \rfloor} W^k_1 + W^{\lfloor t \rfloor + 1}_{t - \lfloor t \rfloor}$ satisfies the required.

So, to show existence on $[0,1]$, the basic idea is to define a sequence W^n_t of random walks with continuous sample paths, such that $\sum_n \sup_{t \in [0,1]} |W^n_t - W^{n+1}_t| < \infty$ almost surely. This would imply that they almost surely uniformly converge to some stochastic process W_t and further, this W_t has continuous sample paths.
If this is the case, the structure of the finite dimensional distributions is then easy to see.

So how do we construct these random walks?
The random walk W^n_t consists of 2^n points with the adjacent ones connected by straight lines. To go from W^n_t to W^{n+1}_t, we insert 2^n more points between the old points. That is, we keep adding detail to make the curve finer and finer. The question is: how do we add these points to make the limiting curve have the required characteristics?
By our construction, the points of $W^n_{k2^{-n}}$ are already as in Lemma 1.10. In particular, $W^n_{(k+1)2^{-n}} - W^n_{k2^{-n}}$ is independent of $W_{k2^{-n}}$ for any k. Now, given W^n_t, let

$$Y_0 = W^{n+1}_{k2^{-n}} = W^n_{k2^{-n}} \text{ and } Y_1 = W^{n+1}_{(k+1)2^{-n}} = W^n_{(k+1)2^{-n}}.$$

We wish to choose a $X = W^{n+1}_{(2k+1)2^{-(n+1)}}$ such that $Y_1 - X$ and $X - Y_0$ are Gaussian with mean 0 and variance 2^{-n}, and $Y_1 - X$, $X - Y_0$, and Y_0 are independent. It is not too difficult to check that $(Y_0 + Y_1)/2 + 2^{-(n+1)/2}\xi$ does the job, where ξ is standard normal and independent of Y_0 and Y_1.

Now, let us make the recursion less explicit. These tent-like interpolations we have performed are known as *Schauder functions*.
For $n = 0, 1, \ldots$ and $k = 1, 3, \ldots, 2^n - 1$, define the *Haar wavelet* $H_{n,k}(t)$ as

$$H_{0,1}(t) = 1 \qquad H_{n,k}(t) = \begin{cases} 2^{(n-1)/2}, & (k-1)2^{-n} < t \le k2^{-n}, \\ -2^{(n+1)/2}, & k2^{-n} < t \le (k+1)2^{-n}, \\ 0, & \text{otherwise.} \end{cases}$$

The Schauder functions are then defined as

$$S_{n,k}(t) = \int_0^t H_{n,k}(s)\,\mathrm{d}s.$$

The Nth random walk is then defined as

$$W^N_t = \sum_{n=0}^N \sum_{k=1,3,\ldots,2^n-1} \xi_{n,k} S_{n,k}(t),$$

where the $\xi_{n,k}$ are iid standard normal.

Now, we must show that these converge uniformly to prove the required. We have

$$\Pr\left[\sup_{t\in[0,1]}|W_t^n - W_t^{n-1}| > \varepsilon_n\right] = \Pr\left[\sup_{k=1,3,\ldots,2^{n-1}}|\xi_{n,k}| > 2^{(n+1)/2}\varepsilon_n\right]$$

$$\leq \sum_{k=1,3,\ldots,2^{n-1}}\Pr\left[|\xi_{n,k}| > 2^{(n+1)/2}\varepsilon_n\right]$$

$$= \sum_{k=1,3,\ldots,2^{n-1}}\Pr\left[|\xi_{0,1}| > 2^{(n+1)/2}\varepsilon_n\right]$$

$$= \sum_{k=1,3,\ldots,2^{n-1}}2\Pr\left[e^{\xi_{0,1}} > \exp\left(2^{(n+1)/2}\varepsilon_n\right)\right]$$

$$\leq \sum_{k=1,3,\ldots,2^{n-1}}2\exp(-2^{(n+1)/2}\varepsilon_n)\mathbf{E}[e^{\xi_{0,1}}]$$

$$= \exp\left(n\log 2 + 1/2 - 2^{(n+1)/2}\varepsilon_n\right).$$

Setting $\varepsilon_n = n^{-2}$,

$$\sum_{n=1}^{\infty}\Pr\left[\sup_{t\in[0,1]}|W_t^n - W_t^{n-1}| > n^{-2}\right] < \infty.$$

Using the Borel-Cantelli lemma, we infer that

$$\sup_{t\in[0,1]}|W_t^n - W_t^{n-1}| \leq n^{-2}$$

almost surely for sufficiently large n. Therefore,

$$\sum_{n=1}^{\infty}\sup_{t\in[0,1]}|W_t^n - W_t^{n+1}| < \infty$$

almost surely. Setting the sample paths of a null set to 0 is an indistinguishable change, so the W_t^n converge uniformly to W_t, which is continuous. Finally, we must check that W_t has the correct finite-dimensional distributions. This is equivalent to showing that for any $t > s > r$, $W_t - W_s$ and W_r are independent.
To do this, a result states that it suffices to show that

$$\mathbf{E}[e^{i\alpha W_r + i\beta(W_t - W_s)}] = e^{\alpha^2 r/2 - \beta^2(t-s)/2}.$$

Showing this however, is direct by considering a sequence of dyadic rationals – numbers of the form $k2^{-n}$ that converge to s and t, then using dominated convergence and the continuity of W_t.

1.6.3. Some Properties

Definition 1.14. Let $\mathcal{F}_t$ be a filtration. A Wiener process W_t is said to be a $\mathcal{F}_t$-*Wiener process* if W_t is $\mathcal{F}_t$-adapted and $W_t - W_s$ is independent of $\mathcal{F}_s$ for any $t > s$.
Given a Wiener process $\mathcal{F}_t$, we also define its *natural filtration* $\mathcal{F}_t^W = \sigma\{W_s : s \leq t\}$.

It is not too difficult to show that a $\mathcal{F}_t$-Wiener process is a $\mathcal{F}_t$-martingale.

Definition 1.15. A $\mathcal{F}_t$-adapted process X_t is called a $\mathcal{F}_t$-*Markov process* if $\mathbf{E}[f(X_t) \mid \mathcal{F}_s] = \mathbf{E}[f(X_t) \mid X_s]$ for all $t \geq s$ and all bounded measurable functions f.

Lemma 1.14. A $\mathcal{F}_t$-Wiener process is a $\mathcal{F}_t$-Markov process.

Intuitively, the sample paths of the Wiener process should be very irregular due to the randomness. This is stated better as:

Lemma 1.15. With unit probability, the sample paths of a Wiener process are non-differentiable at any rational time t.

Proof. Suppose that W_t is differentiable at some t. Then for sufficiently small h, $(W_{t+h} - W_t)/h < M$ for some finite M. This implies that $\sup_{n \geq 1} n|W_{t+n^{-1}} - W_t| < \infty$. Now,

$$\Pr[\sup_{n \geq 1} n|W_{t+n^{-1}} - W_t| < \infty] \leq \Pr\left[\bigcup_{M \geq 1} \bigcap_{n \geq 1} \{n|W_{t+n^{-1}} - W_t| < M\}\right]$$

$$\leq \lim_{M \to \infty} \inf_{n \geq 1} \Pr[n|W_{t+n^{-1}} - W_t| < M].$$

However, $W_{t+n^{-1}} - W_t$ is Gaussian with mean 0 and variance n^{-1}. As a result,

$$\inf_{n \geq 1} \Pr[n|W_{t+n^{-1}} - W_t| < M] = \inf_{n \geq 1} \Pr[|\xi| < Mn^{-1/2}] = 0,$$

where ξ is standard normal. Therefore, W_t is almost surely not differentiable at t. Since the set of rational numbers is countable, the result follows. ∎

For any real-valued function f, define the *total variation* of f on the interval $[a, b]$ by

$$\text{TV}(f, a, b) = \sup_{k \geq 0} \sup_{(t_i) \in P(k,a,b)} \sum_{i=0}^{k} |f(t_{i+1}) - f(t_i)|,$$

where $P(k, a, b)$ is the set of partitions $t_0 = a < t_1 < \cdots < t_{k+1} = b$ of $[a, b]$.
This is essentially the "distance" travelled to go from a to b along the curve.

Lemma 1.16. With unit probability, $\text{TV}(W., a, b) = \infty$ for any $a < b$.

The main goal of this is to give meaning to the stochastic integral, of the form $\int_0^t f_s \, dW_s$. However, the above makes this highly problematic. Indeed, it may be shown that if a function g is of infinite total variation, there is a continuous function f such that the usual Riemann integral of f with respect to g does not exist. Since a reasonable thing to desire is that continuous functions are integrable, it seems that this integral will not do the job.
How do we fix this then? The idea in fact arises by considering the total *squared* variation instead. Note that

$$\lim_{n \to \infty} \sum_{t_i \in \pi_n} (W_{t_{i+1}} - W_{t_i})^2 = b - a$$

almost surely for some sequence of partitions π_n, which is finite.

Finally, it is interesting to know that the $x_t(N)$ described in Equation (1.1) do in fact converge in law to a Wiener process.

§2. The Itô Integral and Stochastic Differential Equations