Notes from Limit Theorems 2

Mihai Nica

NOTES. These are my notes from the class Limit Theorems 2 taught by Proffesor McKean in Spring 2012. I have tried to carefully go over the bigger theorems from the course and fill in all the details explicitly. There is also a lot of information that is folded in from other sources.

- The section on Martingales is supplemented with some notes from "A First Look at Rigorous Probability Theory" by Jeffrey S. Rosenthal, which has a really nice introduction to Martingales.
- The section of the law of the iterated logarithm is supplemented with some inequalities which I looked up on the internet...mostly wikipedia and PlanetMath.
- In the section on Ergodic theorem, I use a notation I found on wikipedia that I like for continued fractions. In my pen-and-paper notes, there is also a little section about Ergodic theory for geodesics on surfaces, which is really cute. However, I couldn't figure out a good way to draw the pictures so it hasn't been typed up yet.
- The section on Brownian Motion is supplemented by the book Brownian Motion and Martingale's in Analysis by Richard Durret which is really wonderful. Some of the slick results are taken straight from there.
- I also include an appendix with results that I found myself reviewing as I went through this stuff.

Contents

CHAPTER 1

Martingales

1. Definitions and Examples

This section on Martingales contains heavy use of conditional random variables. I do a quick review of this topic from Limit Theorems 1 in the appendix.

Definition 1.1. A sequence of random variables $X_0, X_1, ...$ is called a *martingale* if $\mathbf{E}(|X_n|) < \infty$for all n and with probability 1:

$$\mathbf{E}\left(X_{n+1}|X_0, X_1, ..., X_n\right) = X_n$$

Intuitively, this is says that the average value of $X_{n+1}$is the same as that of X_n, even if we are given the values of $X_0$to X_n. Note that conditioning on $X_0, ..., X_n$is just different notation for conditioning on $\sigma(X_0, ..., X_n)$, which is the sigma algebra generated by preimages of Borel sets through $X_0, ..., X_n$. One can make more general martingales by replacing $\sigma(X_0, ..., X_n)$ with an arbitrary increasing chain of sigma algebras $\mathcal{F}_n$; the results here carry over to that setting too.

Example 1.2. Sometimes martingales are called "fair games". The analogy is that the random variable X_n represents the bankroll of the gambler at time n. The game is fair, because at any point in time the equity of the gambler is constant.

Definition 1.3. A *submartingale* is when $\mathbf{E}\left(X_{n+1}|X_0, X_1, ..., X_n\right) \geq X_n$ (i.e. the capital is increasing) and a *supermartingale* is when $\mathbf{E}\left(X_{n+1}|X_0, X_1, ..., X_n\right) \leq X_n$ (i.e. the capital is decreasing) Most of the theorems for martingales work for submartingales, just change the inequality in the right place. To avoid confusion between sub-, super-, and ordinary martingales, we will sometimes call a martingale a "fair martingale".

Example 1.4. The symmetric random walk, $X_n = Z_0 + Z_1 + ... + Z_n$ with each $Z_n = \pm 1$ with probability $\frac{1}{2}$ is a martingale. In terms of the fair game, this is gambling on the outcome of a fair coin.

Remark. Using the properties of conditional probabilities to see that:

$$\begin{aligned} \mathbf{E}\left(X_{n+2}|X_0, X_1, ..., X_n\right) &= \mathbf{E}\left(\mathbf{E}\left(X_{n+2}|X_0, X_1, ..., X_{n+1}\right)|X_0, ...X_n\right) \\ &= \mathbf{E}\left(X_{n+1}|X_0, ...X_n\right) \\ &= X_n \end{aligned}$$

With a simple argument by induction, we get that in general:

$$\mathbf{E}\left(X_m|X_0, X_1, ..., X_n\right) = X_n$$

In particular then $\mathbf{E}(X_n) = \mathbf{E}(X_0)$ for every n. If τ is a random "time", (a non-negative integer) that is independent of the X_n's, then $\mathbf{E}(X_\tau)$ is a weighted average of $\mathbf{E}(X_n)$'s, so have $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$ still. What if $\upsilon$is dependent on the

$X_n's$? In general we cannot have equality for the example of the simple symmetric random walk (coin-flip betting), with τ =first time that $X_n = -1$ has $\mathbf{E}(X_n) = -1 \neq 0 = \mathbf{E}(X_0)$. The next section gives some conditions where this holds.

2. Stopping times

DEFINITION 2.1. For a martingale $\{X_n\}$, A non-negative integer valued random variable τ is a *stopping time* if it has the property that:

$$\{\tau = n\} \in \sigma(X_1, X_2, \ldots, X_n)$$

Intuitively, this is saying that one can determine if $\tau = n$ just by looking at the first n steps in the martingale.

EXAMPLE 2.2. In the example of the random coin flipping, if we let τ be the first time so that X_n =10, then τ is a stopping time.

EXAMPLE 2.3. We often are interested in X_τ, the value of the martingale at the random time τ. This is precisely defined as $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$. Another handy rewriting is: $X_\tau = \sum X_k \mathbf{1}_{\{\tau=k\}}$.

LEMMA 2.4. *If $\{X_n\}$is a submartingale and $\tau_1, \tau_2$are bounded stopping times so that $\exists M$ s.t. $0 \leq \tau_1 \leq \tau_2 \leq M$ with probability 1, then $\mathbf{E}(X_{\tau_1}) \leq \mathbf{E}(X_{\tau_2})$, with equality for fair martingales.*

PROOF. For fixed k, the event $\{\tau_1 < k \leq \tau_2\}$can be written as $\{\tau_1 < k \leq \tau_2\} = \{\tau_1 \leq k-1\} \cap \{\tau_2 \leq k-1\}^C$ from which we see that the event $\{\tau_1 < k \leq \tau_2\} \in \sigma(X_0, X_1, \ldots, X_{k-1})$ because $\tau_1$and $\tau_2$are both stopping times. We have then the following manipulation using a telescoping series, linearity of the expectation, the fact that $\mathbf{E}(Y\mathbf{1}_A)$= $\mathbf{E}(\mathbf{E}(Y|X_0, X_1, \ldots, X_{k-1})\mathbf{1}_A)$ for events $A \in \sigma(X_0, X_1, \ldots, X_{k-1})$, and finally the fact that $E(X_k|X_0, X_1, \ldots X_{k-1}) - X_{k-1} \geq 0$ since X_n is a (sub)martingale. (with equality for fair martingales):

$$\begin{aligned}
\mathbf{E}(X_{\tau_2}) - \mathbf{E}(X_{\tau_1}) &= \mathbf{E}(X_{\tau_2} - X_{\tau_1}) \\
&= \mathbf{E}(\sum_{k=\tau_1+1}^{\tau_2} X_k - X_{k-1}) \\
&= \mathbf{E}\left(\sum_{k=1}^{M} (X_k - X_{k-1})\mathbf{1}_{\{\tau_1 < k \leq \tau_2\}}\right) \\
&= \mathbf{E}\left(\sum_{k=1}^{M} (\mathbf{E}(X_k|X_0, X_1, \ldots X_{k-1}) - X_{k-1})\mathbf{1}_{\{\tau_1 < k \leq \tau_2\}}\right) \\
&= \sum_{k=1}^{M} \mathbf{E}\left((\mathbf{E}(X_k|X_0, X_1, \ldots X_{k-1}) - X_{k-1})\mathbf{1}_{\{\tau_1 < k \leq \tau_2\}}\right) \\
&\geq \sum_{k=1}^{M} \mathbf{E}\left(0\mathbf{1}_{\{\tau_1 < \mathbf{k} \leq \tau_2\}}\right) \\
&= 0
\end{aligned}$$

Where the inequality is equality in the case of a fair martingale. □

THEOREM 2.5. *Say $\{X_n\}$ is a martingale and τ a bounded stopping time, (that is $\exists M$ s.t. $0 \le \tau \le M$ with probability 1). Then:*

$$\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$$

PROOF. Let υbe the random variable which is constantly 0. This is a stopping time! So by the above lemma, since $0 \le \upsilon \le \tau \le M$, we have that $\mathbf{E}(X_\tau) = \mathbf{E}(X_\upsilon) = \mathbf{E}(X_0)$ □

THEOREM 2.6. *For $\{X_n\}$a martingale and τ a stopping time which is almost surely finite (that is $\mathbf{P}(\tau < \infty) = 1$) we have:*

$$\mathbf{E}(X_\tau) = \mathbf{E}(X_0) \iff \mathbf{E}\left(\lim_{n\to\infty} X_{\min(\tau,n)}\right) = \lim_{n\to\infty} \mathbf{E}\left(X_{\min(\tau,n)}\right)$$

PROOF. It suffices to show that $\mathbf{E}(X_\tau) = \mathbf{E}\left(\lim_{n\to\infty} X_{\min(\tau,n)}\right)and\mathbf{E}(X_0) = \lim_{n\to\infty} \mathbf{E}\left(X_{\min(\tau,n)}\right)$. The first equality holds since $\mathbf{P}(\tau < \infty) = 1$ gives $\mathbf{P}(\lim_{n\to\infty} X_{min(\tau,n)} = X_\tau) = 1$, so they agree almost surely. The second holds by the above theorem concerning bounded stopping times since for any n, $\min(\tau, n)$ is a bounded stopping time, so we have $\mathbf{E}\left(X_{\min(\tau,n)}\right) = \mathbf{E}(X_0)$, so equality holds in the limit too. □

REMARK. The above theorem can be combined with things like monotone convergence theorem or Lebesgue dominated convergence theorem to switch the limits and conclude that $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$. Here are some examples:

EXAMPLE 2.7. If $\{X_n\}$is a martingale and τ a stopping time so that $\mathbf{P}(\tau < \infty) = 1$ and $\mathbf{E}(|X_\tau|) < \infty$, and $\lim_{n\to\infty} \mathbf{E}(X_n \mathbf{1}_{\tau>n}) = 0$, then $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$.

PROOF. For any n we have: $X_{\min(\tau,n)} = X_n \mathbf{1}_{\tau>n} + X_\tau \mathbf{1}_{\tau\le n}$Taking expectation and then the limit as $n \to \infty$, gives:

$$\begin{aligned} \lim_{n\to\infty} \mathbf{E}(X_{\min(\tau,n)}) &= \lim_{n\to\infty} \mathbf{E}(X_n \mathbf{1}_{\tau>n}) + \lim_{n\to\infty} \mathbf{E}(X_\tau \mathbf{1}_{\tau>n}) \\ &= 0 + \mathbf{E}(X_\tau) \end{aligned}$$

Where the first term is 0 by hypothesis, and the second limit is justified since $X_\tau \mathbf{1}_{\tau>n} \to X_\tau$pointwise almost surely since $\mathbf{P}(\tau < \infty) = 1$, and the dominant majorant $\mathbf{E}(|X_\tau|) < \infty$lets us use the Lebesgue dominated convergence theorem to conclude the convergence of the expectation. □

EXAMPLE 2.8. Suppose $\{X_n\}$is a martingale and τ a stopping time so that $\mathbf{E}(\tau) < \infty$ and $|X_{n+1} - X_n| \le M < \infty$for some fixed M and for every n. Then $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$.

PROOF. Let $Y = |X_0| + M\tau$. Then Y can be used as a dominant majorant in a L.D.C.T. very similar to the above example to get the conclusion. □

3. Martingale Convergence Theorem

The proof relies on the famous upcrossing lemma:

LEMMA 3.1. *[The Upcrossing Lemma]. Let $\{X_n\}$be a submartingale. For fixed $\alpha, \beta \in \mathbb{R}$, $\beta > \alpha$,and $M \in \mathbb{N}$ let $U_M^{\alpha,\beta}$be the number of "upcrossings" that the martingale $\{X_n\}$makes of the interval α, β in the time period $1 \le n \le M$. (An upcrossing is when X_n goes from being less than α initially to being more than β*

later. Precisely this is: $U_M^{\alpha,\beta} = \max_k\{k : \exists t_1 < u_1 < \ldots < t_k < u_k \leq M \; s.t. \; X_{t_i} \leq \alpha \text{ and } X_{u_i} \geq \beta \, \forall i\}$ *). Then:*

$$\mathbf{E}(U_M^{\alpha,\beta}) \leq \frac{\mathbf{E}\left(|X_M - X_0|\right)}{\beta - \alpha}$$

Proof. Firstly, we remark that it suffices to prove the result when the submartingale $\{X_n\}$is replaced by $\{\max(X_n, \alpha)\}$, since this is still a submartingale, it has the same number of upcrossings as X_n, and $|\max(X_M, \alpha) - \max(X_0, \alpha)| \leq |X_M - X_0|$, so the equality is only strengthened. In other words, we assume without loss of generality that $X_n \geq \alpha$ for all n. This simplification is used in exactly one spot later on to get the inequality we need.

Let us now carefully nail down where the upcrossings happen. Define $u_0 = v_0 = 0$ and iteratively define:

$$\begin{aligned} u_j &= \min(M, \inf_{k > v_{j-1}} \{k : \; X_k \leq \alpha\} \\ v_j &= \min(M, \inf_{k > u_j} \{k : \; X_k \geq \beta\} \end{aligned}$$

These record the times where the martingale crosses the interval $[\alpha, \beta]$; the u_j's record when it first crosses moving to the left of the interval, and the v_j's record crosses going to the right of the interval. They are also truncated at time M so that they are *bounded stopping times.* Moreover, since these times are strictly increasing until they hit M, it must be the case that $v_M = M$. We have then, using some crafty telescoping sums:

$$\begin{aligned} \mathbf{E}(X_M) &= \mathbf{E}\left(X_{v_M}\right) \\ &= \mathbf{E}\left(X_{v_M} - X_{u_M} + X_{u_M} - X_{v_{M-1}} + X_{v_{M-1}} - \ldots - X_{u_1} + X_{u_1} - X_0 + X_0\right) \\ &= \mathbf{E}\left(X_0\right) + \mathbf{E}\left(\sum_{k=1}^{M} X_{v_k} - X_{u_k}\right) + \sum_{k=1}^{M} \mathbf{E}\left(X_{u_k} - X_{v_{k-1}}\right) \end{aligned}$$

The third term is non-negative! This is because u_k and $v_{k-1}$are both bounded stopping times with $0 \leq v_{k-1} \leq u_k \leq M$, so our theorem about stopping times gives that this expectation is non-negative. (This is subtle! Most of the time (when we haven't hit time M yet) we expect $X_{u_k} < \alpha$while $X_{v_{k-1}} > \beta$, so their difference is negative. However, because of the small probability event where $v_{k-1} < M$ and $u_k = M$, we get a big positive number with small probability which balances the whole expectation. Compare to the example of a simple symmetric random walk with a truncated stopping time for $\tau =$first time that $X_n = -1$.)

Now the second term, has $\mathbf{E}\left(\sum_{k=1}^{M} X_{v_k} - X_{u_k}\right) \geq \mathbf{E}\left((\beta - \alpha) U_M^{\alpha,\beta}\right)$. This is because each upcrossing counted in $U_M^{\alpha,\beta}$ contributes at least $(\beta - \alpha)$ to the sum, null cycles (where $u_k = v_k = M$) contribute nothing, and the possibly one incomplete cycle (where $u_k < M$ but $v_k = M$) must give a non-negative contribution to the sum *by the simplification that* $X_n > \alpha$.

Hence we have:

$$\mathbf{E}(X_M) \geq \mathbf{E}\left(X_0\right) + (\beta - \alpha)\mathbf{E}\left(U_M^{\alpha,\beta}\right) + 0$$

Which gives the desired result. □

THEOREM 3.2. *[Martingale Convergence Theorem] Let* $\{X_n\}$*be a submartingale with* $\sup_n \mathbf{E}(|X_n|) < \infty$. *Then there exists a random variable* X *so that* $X_n \to X$ *almost surely. (That is* $X_n(\omega) = X(\omega)$ *for almost all* $\omega \in \Omega$*).*

PROOF. Firstly, since $\sup_n \mathbf{E}(|X_n|) < \infty$, by Fatou's lemma we have: $\mathbf{E}(\liminf_n |X_n|) \leq \liminf_n \mathbf{E}(|X_n|) \leq \sup_n \mathbf{E}(|X_n|) < \infty$, from which it follows that $\mathbf{P}(|X_n| \to \infty) = 0$. This ensures that the $X_n$cannot "leak away" probability to $\pm\infty$, which would prevent the limiting random variable from being properly normalized.

Now suppose by contradiction that $\mathbf{P}(\liminf X_n < \limsup X_n) > 0$, i.e. there is a non-zero probability of $X_n$not converging. Then, using the density of the rationals and countable subadditivity to find an α and β so that $\mathbf{P}(\liminf\ X_n < \alpha < \beta < \limsup X_n) > 0$. Counting the number of upcrossing X_n makes of $[\alpha, \beta]$,we see that we must have: $\mathbf{P}\left(\lim_{M\to\infty} U_M^{\alpha,\beta} = \infty\right) > \mathbf{P}(\liminf\ X_n < \alpha < \beta < \limsup\ X_n) > 0$. Hence $\mathbf{E}\left(\lim_{M\to\infty} U_M^{\alpha,\beta}\right) = \infty$. By the monotone convergence theorem however, we have that $\lim_{M\to\infty} \mathbf{E}(U_M^{\alpha,\beta}) = \mathbf{E}\left(\lim_{M\to\infty} U_M^{\alpha,\beta}\right) = \infty$.

But now we have reached a contradiction! For by the upcrossing lemma:

$$\lim_{M\to\infty} \mathbf{E}(U_M^{\alpha,\beta}) \leq \frac{\lim_{M\to\infty} \mathbf{E}\left(|X_M - X_0|\right)}{\beta - \alpha} \leq \frac{2\sup_M \mathbf{E}(|X_n|)}{\beta - \alpha} < \infty$$

□

4. Applications

THEOREM 4.1. *[Levy]Suppose* Z *a random variable with* $\mathbf{E}(|Z|) < \infty$*, and that* $\{\mathcal{F}_n\}$ *is a decreasing chain of* $\sigma-$*algebras,* $\mathcal{F}_1 \supset \mathcal{F}_2 \supset \ldots$ *(This is saying that they are getting coarser and coarser). Let* $\mathcal{F}_\infty = \cap \mathcal{F}_n$. *Then we have almost surely:*

$$\lim_{n\to\infty} \mathbf{E}(Z|\mathcal{F}_n) = \mathbf{E}(Z|\mathcal{F}_\infty)$$

PROOF. We first prove that there is an almost sure limit using the martingale convergence theorem, and then we check the defining properties of $\mathbf{E}(Z|\mathcal{F}_\infty)$ to verify that this is indeed the limit.

Firstly, let $X_n = \mathbf{E}(Z|\mathcal{F}_n)$. Then for any fixed $M \in \mathbb{N}$ we have that the sequence $X_M, X_{M-1}, \ldots X_2, X_1$is a martingale (Here we are referring to a slightly more general martingale than in our original definition, the sigma algebra $\sigma(X_1, X_2, \ldots)$ in the definition is replaced by arbitrary increasing sigma algebras $\mathcal{F}_n$. The expectation property of the martingale follows by the fact that $\mathbf{E}(\mathbf{E}(Z|\mathcal{F})|\mathcal{G}) = \mathbf{E}(Z|\mathcal{G})$ when $\mathcal{G} \subset \mathcal{F}$) Notice that we had to reverse the order of the sequence to get the sigma algebras to increase (i.e. get finer and finer), so that we really have a martingale. For this reason, the martingale convergence theorem does not apply directly but the idea of the proof will still work. Suppose by contradiction, as in the proof of the martingale convergence theorem, that $\mathbf{P}(\liminf\ X_n < \limsup\ X_n) > 0$. Then, as before, find $\alpha$and β so that $\mathbf{P}(\liminf X_n < \alpha < \beta < \limsup X_n) > 0$. Since there are infinitely many crossings then of the interval $[\alpha, \beta]$, we can know that the number of downcrossings $D_M^{\alpha,\beta}$has $\mathbf{P}\left(\lim_{M\to\infty} D_M^{\alpha,\beta} = \infty\right) > 0$ and so $\mathbf{E}\left(\lim_{M\to\infty} D_M^{\alpha,\beta}\right) = \infty$. Hence, since $D_M^{\alpha,\beta}$is increasing in M (the number of downcrossings can only increase if we wait longer), we may find an $M_0 \in \mathbb{N}$so that $\mathbf{E}\left(D_{M_0}^{\alpha,\beta}\right) > \frac{2\mathbf{E}(|Z|)}{\beta-\alpha}$.

Taking now the martingale sequence $X_{M_0}, X_{M_0-1}, \dots X_2, X_1$, we have a violation of the upcrossing lemma just as we did in the martingale convergence theorem.

Next, to verify that the limit is indeed $\mathbf{E}(Z|\mathcal{F}_\infty)$ we just need to check the two defining properties, namely that it is $\mathcal{F}_\infty$measurable and that it has the correct expectation value for events in $\mathcal{F}_\infty$. $\lim_{n\to\infty} \mathbf{E}(Z|\mathcal{F}_n)$ is $\mathcal{F}_\infty$measurable, since $\mathcal{F}_\infty \subset \mathcal{F}_n$ for every n, meaning that $\mathbf{E}(Z|\mathcal{F}_n)$ is $\mathcal{F}_\infty$measurable for every n, and so the limit is too.

To see that $\lim_{n\to\infty} \mathbf{E}(Z|\mathcal{F}_n)$ takes the correct expectations for events in $\mathcal{F}$, notice that for any $A \in \mathcal{F}_\infty \subset \mathcal{F}_n$ we have for every n that $\mathbf{E}\left(\mathbf{E}(Z|\mathcal{F}_n)\mathbf{1}_A\right) = \mathbf{E}(Z\mathbf{1}_A)$ since $A \in \mathcal{F}_n$, so in the limit $\lim_{n\to\infty} \mathbf{E}\left(\mathbf{E}(Z|\mathcal{F}_n)\mathbf{1}_A\right) = \mathbf{E}(Z\mathbf{1}_A)$. Hence the problem of proving that $\mathbf{E}\left(\lim_{n\to\infty} \mathbf{E}(Z|\mathcal{F}_n)\mathbf{1}_A\right) = \mathbf{E}(Z\mathbf{1}_A)$ is reduced to an interchange of a limit with an expectation. If Z is bounded, this is justified by the bounded convergence theorem. For Z not bounded, truncating Z by $Z\mathbf{1}_{\{Z\le N\}}$with a bit more work will give the same interchange of limits. □

THEOREM 4.2. *[Levy] Suppose Z a random variable with $\mathbf{E}(|Z|) < \infty$, and that $\{\mathcal{F}_n\}$ is an increasing chain of σalgebras, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ (This is saying that they are getting finer and finer). Let $\mathcal{F}_\infty = \cup \mathcal{F}_n$. Then we have almost surely:*

$$\lim_{n\to\infty} \mathbf{E}(Z|\mathcal{F}_n) = \mathbf{E}(Z|\mathcal{F}_\infty)$$

PROOF. This proof is like the last one. In this case $\mathbf{E}(Z|\mathcal{F}_n)$ really is a martingale (no backwards), so an almost sure limit exists by the martingale convergence theorem. Some more work here is needed....I think you get the desired property by approximation with “tame events” $A \in \mathcal{F}_\infty$, for every $\epsilon > 0$ there exists $A_n \in \mathcal{F}_n$such that $\mathbf{P}(A\Delta A_n) < \epsilon$. □

REMARK. This result is often known as the *“Levy Zero-One Law”* since a common application is to consider an event $A \in \mathcal{F}_\infty$, for which the theorem tells us that:

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}(A|\mathcal{F}_n) &= \lim_{n\to\infty} \mathbf{E}(\mathbf{1}_A|\mathcal{F}_n) \\ &= \mathbf{E}(\mathbf{1}_A|\mathcal{F}_\infty) \\ &= \mathbf{1}_A \end{aligned}$$

Where the last equality holds since A is $\mathcal{F}_\infty$measurable. This says in particular that this probability is either 0 or 1, since these are the only two values taken on by $\mathbf{1}_A$. In this setting, the theorem gives a short proof of the Kolmogorov zero-one law.

THEOREM 4.3. *[Kolmogorov Zero-One law] Let $X_1, X_2, \dots$be an infinite sequence of i.i.d. random variables. Define:*

$$\begin{aligned} \mathcal{F}_n &= \sigma\left(\bigcup_{k=1}^{n} \sigma(X_k)\right) \\ \mathcal{F}_\infty &= \bigcup_{k=1}^{n} \mathcal{F}_n \\ \mathcal{F}_{tail} &= \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{k=n}^{\infty} \sigma(X_k)\right) \end{aligned}$$

Then any event $A \in \mathcal{F}_{tail}$ has either $\mathbf{P}(A) = 0$ or $\mathbf{P}(A) = 1$. These are those events which do not depend on finitely many of the $X_n's$.

Proof. Let $A \in \mathcal{F}_{tail}$. For any $n \in \mathbb{N}$ we have that $\mathbf{P}(A) = \mathbf{P}(A|\mathcal{F}_n) = \mathbf{E}(\mathbf{1}_A|\mathcal{F}_n)$ since $A \in \mathcal{F}_{tail}$ does not depend on the first n variables, so its conditional expectation is a constant. Have then, (as in the above "Levy 0-1" remark):

$$\begin{aligned}\mathbf{P}(A) &= \lim_{n\to\infty} \mathbf{P}(A|\mathcal{F}_n) \\ &= \mathbf{1}_A\end{aligned}$$

Since $A \in \mathcal{F}_\infty$So indeed, the only the values of $\mathbf{P}(A)$ that are possible are 1 and 0. □

Theorem 4.4. *[Strong Law of Large Numbers] Suppose $X_1, X_2, \ldots$ are i.i.d. Then we have almost surely that:*

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mathbf{E}(X_1)$$

Proof. Define $S_n = X_1 + X_2 + \ldots + X_n$, and let $\mathcal{F}_n = \sigma(\bigcup_{k=n}^{\infty} \sigma(S_k))$ be the sigma algebra of the tail $S_n, S_{n+1}, \ldots$. We now claim that:

$$\mathbf{E}(X_1|\mathcal{F}_n) = \frac{S_n}{n}$$

This can be seen in the following slick way. First notice that by symmetry, we must have $\mathbf{E}(X_1|\mathcal{F}_n) = \mathbf{E}(X_2|\mathcal{F}_n) = \ldots = \mathbf{E}(X_n|\mathcal{F}_n)$. By linearity now: $\sum_{k=1}^{n} \mathbf{E}(X_k|\mathcal{F}_n)$=$\mathbf{E}(\Sigma_{k=1}^{n} X_k|\mathcal{F}_n) = \mathbf{E}(S_n|\mathcal{F}_n) = S_n$, since $S_n \in \mathcal{F}_n$. Hence since they are all equal, and sum to S_n, we get $\mathbf{E}(X_1|\mathcal{F}_n) = \frac{S_n}{n}$ as desired. By Levy's theorem now:

$$\begin{aligned}\lim_{n\to\infty} \frac{S_n}{n} &= \lim_{n\to\infty} \mathbf{E}(X_1|\mathcal{F}_n) \\ &= \mathbf{E}\left(X_1|\bigcap_k \mathcal{F}_k\right)\end{aligned}$$

From here, one can use the Hewitt-Savage zero-one law (which says that permutation invariant events have a zero one law), to see that the whole sigma algebra $\bigcap_k \mathcal{F}_k$ must be the trivial one, so then $\mathbf{E}\left(X_1|\bigcap_k \mathcal{F}_k\right) = \mathbf{E}(X_1)$. Alternatively, once we have conclude that such an almost sure limit exists, one could then remark by the Kolmogorov zero that the limit must be a constant (for $\lim_{n\to\infty} \frac{S_n}{n}$ does not depend on finitely many of the $X_n's$ so any type of event $\{\lim \frac{S_n}{n} < \alpha\}$must have probability 0 or 1. By taking a sup, we can find that it must be a constant.) Combining this with the above, using the fact that conditional random variables preserve the expectation, shows the constant is indeed $\mathbf{E}(X_1)$. □

Theorem 4.5. *[Hewitt Savage Zero-One Law] Let $X_1, X_2, \ldots$be an infinite sequence of i.i.d. random variables. Let A be an event which is unchanged under finite permutations of the induces of the $X_i's$. (e.g. for every finite permutation Π,$\omega = (x_1, x_2, \ldots) \in A$ iff $\Pi(\omega) = (x_{\Pi(1)}, x_{\Pi(2)}, \ldots) \in A$ i.e. $\Pi(A) = A$). Then $\mathbf{P}(A) = 0$ or 1.*

PROOF. We call an event "tame" if it only depends on finitely many of the $X_i's$.The proof is a consequence of the fact that for any ϵ, any event A can be approximated by a "tame event" B so that $\mathbf{P}(B \triangle A) < \epsilon$. (This is completely analogous to the fact that for the usual Lebesgue measure on $\mathbb{R}$, one can approximate any measurable set S by a finite union of open intervals I_n so that $\lambda(\cup_{i=1}^{n} I_i \triangle U) < \epsilon$. This comes from the definition of the Lebesgue measure as the inf of the outer measure with open sets, and the fact that every open set is a union of countably many intervals, of which only finitely many are needed to be within $\epsilon/2$. In the same vein, the probability measure on the infinite sequence of events is generated by the outer measure from tame events. This is usually all packaged up in the Caratheodory extension theorem.). Once we have this tame event B, depending only on $X_1, \ldots X_n$ we letΠ be the permutation that permutes $1, \ldots n$ with $n+1, \ldots, 2n$ so that B and $\Pi(B)$ are independent events. Have then:

$$\begin{aligned}
\mathbf{P}(A) &\approx \mathbf{P}(A \cap B) \\
&= \mathbf{P}(\Pi(A) \cap \Pi(B)) \\
&= \mathbf{P}(A \cap \Pi(B)) \\
&\approx \mathbf{P}(B \cap \Pi(B)) \\
&= \mathbf{P}(B)\mathbf{P}(\Pi(B)) \\
&= \mathbf{P}(B)^2 \\
&\approx \mathbf{P}(A)^2
\end{aligned}$$

Where each of the approximations hold within ϵ by the choice of B. Since we can do this for every $\epsilon > 0$, we get $\mathbf{P}(A) = \mathbf{P}(A)^2$ and the result follows. $\square$

CHAPTER 2

The Law of the Iterated Logarithm

We will prove that for a sequence of i.i.d events $X_1, X_2, \ldots$with mean 0 and variance 1 that for $S_n = \sum_{i=1}^n X_n$:

$$\mathbf{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n\log(\log n))}} = \sqrt{2}\right)$$

This result is giving us finer information about these sums than the law of large numbers or the central limit theorem. We need the theory of martingales to get Doob's inequality, and then a bunch of other sneaky tricks, like the Borel Cantelli lemmas, to get the result. We will also need a few analytic type estimates along the way. (Actually, our proof here will only prove the case where the $X_n's$ are ± 1 with probability 1/2 each. The result can be generalized by using even finer estimates)

1. First Half of the Law of the Iterated Logarithm

To start, we will first prove some helpful lemmas.

Lemma 1.1. *[Doob's Inequality] For a submartingale Z_n, we have for any $\alpha > 0$ that:*

$$\mathbf{P}\left(\left(\max_{0\le i\le n} Z_i\right) \ge \alpha\right) \le \frac{\mathbf{E}(|Z_n|)}{\alpha}$$

Proof. (Taken from Rosenthal) Let $A_k$be the event that $\{X_k \ge \alpha,$ but $X_i < \alpha$ for $i < k\}$,i.e. that the process reaches $\alpha$for the first time at time k. These are disjoint events with $A = \cup A_k = \{(\max_{0\le i\le n} Z_i) \ge \alpha\}$ which is the event we want. Now consider:

$$\begin{aligned}
\alpha\mathbf{P}(A) &= \sum_{k=0}^{n} \alpha\mathbf{P}(A_k) \\
&= \sum \mathbf{E}(\alpha \mathbf{1}_{A_k}) \\
&\le \sum \mathbf{E}(X_k \mathbf{1}_{A_k}) \text{ since } X_k \ge \alpha \text{on } A_k \\
&\le \sum \mathbf{E}\left(\mathbf{E}\left(X_n | X_1, X_2, \ldots X_k\right)\mathbf{1}_{A_k}\right) \text{ since it's a submartingale} \\
&= \sum \mathbf{E}(X_n \mathbf{1}_{A_k}) \\
&= \mathbf{E}(X_n \mathbf{1}_A) \\
&\le \mathbf{E}(|X_n|)
\end{aligned}$$

And the result follows. □

Remark. This is a “rich man’s version of Chebyushev-type inequalities”, which are proved using the same trick as in lines 3 and 4 of the inequality train above. The fact that the behavior of the whole martingale can be controlled by the end point of the martingale gives us the little extra oomph we need.

Lemma 1.2. *[Hoeffding’s Inequality] Let Y be a random variable so that $\mathbf{E}(Y) = 0$ and $a, b \in \mathbb{R}$ so that $a \le Y \le b$ almost surely. Then $\mathbf{E}(e^{tY}) \le e^{t^2(b-a)^2/8}$.*

Proof. Write X as a convex combination of a and b: $Y = \alpha b + (1-\alpha)a$ where $\alpha = (Y-a)/(b-a)$. By convexity of $e^{(\,)}$, have then:

$$e^{tY} \le \frac{Y-a}{b-a}e^{tb} + \frac{b-Y}{b-a}e^{ta}$$

Taking expectations (and using $\mathbf{E}(Y) = 0$), have:

$$\mathbf{E}\left(e^{tY}\right) \le \frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta} = e^{g(t(b-a))}$$

For $g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$and $\gamma = -\frac{a}{b-a}$. Notice $g(0) = g'(0) = 0$ and $g''(u) < \frac{1}{4}$for all u. Hence by Taylor’s theorem:

$$\begin{aligned} g(u) &= g(0) + ug'(0) + \frac{u^2}{2}g''(\xi) \\ &\le 0 + 0 + \frac{u^2}{2}\frac{1}{4} = \frac{u^2}{8} \end{aligned}$$

So then $\mathbf{E}\left(e^{tY}\right) \le e^{g(t(b-a))} \le e^{t^2(b-a)^2/8}$ □

Lemma 1.3. *Let $X_1, X_2, \dots$ be i.i.d with $\mathbf{P}(X_1 = \pm 1) = \frac{1}{2}$ and $S_n = \sum_{k=1}^n X_k$. Then $\mathbf{P}(\max_{k \le n} S_k > \lambda) \le e^{-\lambda^2/2n}$.*

Proof. Have, by using Doob’s inequality and Hoeffding’s Inequality, for any $t \in \mathbb{R}$, we have:

$$\begin{aligned} \mathbf{P}(\max_{k \le n} S_k > \lambda) &= \mathbf{P}(\max_{k \le n} e^{tS_k} > e^{t\lambda}) \\ &\le e^{-t\lambda}\mathbf{E}(e^{tS_n}) \\ &= e^{-t\lambda}\mathbf{E}(e^{tX_1})^n \\ &\le e^{-t\lambda}e^{nt^2(b-a)^2/8} \end{aligned}$$

Set $t = 4\lambda/n(b-a)^2$ to get:

$$\begin{aligned} \mathbf{P}(\max_{k \le n} S_k > \lambda) &\le e^{-(4\lambda/n(b-a)^2)\lambda}e^{n(4\lambda/n(b-a)^2)^2(b-a)^2/8} \\ &= e^{-2\lambda^2/n(b-a)^2} \end{aligned}$$

For simple symmetric steps, we have $a = -1$ and $b = 1$, so this gives the result. □

Theorem 1.4. *Let $X_1, X_2, \dots$ be i.i.d with $\mathbf{P}(X_1 = \pm 1) = \frac{1}{2}$ and $S_n = \sum_{k=1}^n X_k$. Then for any $\epsilon > 0$,*

$$\mathbf{P}\left(\limsup_{n \to \infty} \frac{S_n}{\sqrt{n \log(\log n))}} > \sqrt{2+\epsilon}\right) = 0$$

Or in other words, since this holds for any value of $\epsilon > 0$:

$$\mathbf{P}\left(\limsup_{n\to\infty}\frac{S_n}{\sqrt{n\log(\log n))}} \le \sqrt{2}\right) = 1$$

PROOF. Fix some $\theta > 1$ (the choice will be made more precise later). We will show that with the correct choice of θ, the events $A_n = \{S_k > \sqrt{(2+\epsilon)k\log(\log k))}$for some k, $\theta^{n-1} \le k < \theta^n\}$ happens only finitely many times, which will show that the limsup can't be more than $\sqrt{2+\epsilon}$. To do this it suffices to show that $\mathbf{P}(A_n)$ is summable, because then the Borel-Cantelli lemmas will show that A_n happens finitely often with probability 1. We have (using our previous lemma):

$$\begin{aligned}
\mathbf{P}(A_n) &= \mathbf{P}\left(S_k > \sqrt{(2+\epsilon)k\log(\log k))}, \theta^{n-1} \le k < \theta^n\right) \\
&\le \mathbf{P}\left(S_k > \sqrt{(2+\epsilon)\theta^{n-1}\log(\log\theta^{n-1}))}, \theta^{n-1} \le k < \theta^n\right) \\
&\le \mathbf{P}\left(\max_{k\le\theta^n} S_k > \sqrt{(2+\epsilon)\theta^{n-1}\log(\log\theta^{n-1}))}\right) \\
&\le exp\left(-\frac{\sqrt{(2+\epsilon)\theta^{n-1}\log(\log\theta^{n-1}))}^2}{2\theta^n}\right) \\
&= exp\left(-\frac{2+\epsilon}{2}\frac{\theta^{n-1}(\log(n-1)+\log(\log(\theta)))}{\theta^n}\right) \\
&\approx exp\left(-\left(1+\frac{\epsilon}{2}\right)\theta^{-1}\log(n-1)\right) \quad \text{for large } n
\end{aligned}$$

So choosing $\theta < 1 + \frac{\epsilon}{2}$, gives us that $\left(1+\frac{\epsilon}{2}\right)\theta^{-1} > 1$, so this is:

$$\mathbf{P}(A_n) \le (n-1)^{-\left(1+\frac{\epsilon}{2}\right)\theta^{-1}}$$

From which we see that $\mathbf{P}(A_n)$ is summable (it's a p-series!). By using the Borel Cantelli lemma, this means that A_n happens only finitely many times with probability 1, which is the desired result. □

2. Second Half of the Law of the Iterated Logarithm

To prove the other half, we need some more estimates.

LEMMA 2.1. *[Mill's Inequality] This is an estimate concerning the probability density function of a Gaussian:*

$$\frac{\lambda}{\lambda^2+1}e^{-\lambda^2/2} \le \int_\lambda^\infty e^{-y^2/2}dy \le \frac{1}{\lambda}e^{-\lambda^2/2}$$

PROOF. To prove the lower bound, we find a remarkable anti-derivative:

$$\begin{aligned}
\int_\lambda^\infty e^{-y^2/2}dy &\ge \int_\lambda^\infty e^{-y^2/2}\left(\frac{y^4+2y^2-1}{y^2+2y^2+1}\right)dy \\
&= \left[-\frac{y}{y^2+1}e^{-y^2/2}\right]_\lambda^\infty \\
&= \frac{\lambda}{\lambda^2+1}e^{-\lambda^2/2}
\end{aligned}$$

The upper bound is found by using the estimate $y/\lambda > 1$ in the range of integration:

$$\begin{aligned}\int_\lambda^\infty e^{-y^2/2}dy &\leq \int_\lambda^\infty \frac{y}{\lambda}e^{-y^2/2}dy \\ &= \frac{1}{\lambda}\left[-e^{-y^2/2}\right]_\lambda^\infty \\ &= \frac{1}{\lambda}e^{-\lambda^2/2}\end{aligned}$$

□

THEOREM 2.2. *Let $X_1, X_2, \ldots$ be i.i.d with $\mathbf{P}(X_1 = \pm 1) = \frac{1}{2}$ and $S_n = \sum_{k=1}^n X_k$. Then for any $\epsilon > 0$,*

$$\mathbf{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n\log(\log n))}} \geq \sqrt{2-2\epsilon}\right) = 1$$

Or in other words, since this holds for any value of $\epsilon > 0$:

$$\mathbf{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n\log(\log n))}} \geq \sqrt{2}\right) = 1$$

PROOF. As in the proof of the other half of the law, the idea is to prove that the appropriate events happen infinitely often using the Borel-Cantelli lemmas. Fix $\theta > 1$ (the choice will be made precise later). Let $B_n = \left\{S_{\theta^n} - S_{\theta^{n-1}} \geq \sqrt{(2-\epsilon)\theta^n \log(\log(\theta^n))}\right\}$. We will show that these occur infinitely often and then show why this gives the result. Notice that the $B_n's$ are independent, as each B_n depends only on the value of X_k for $\theta^{n-1} \leq k \leq \theta^n$, so to prove that $B_n$happens i.o. it suffices to show, via the Borel Cantelli lemma, that $\mathbf{P}(B_n)$ is not summable. Consider:

$$\begin{aligned}\mathbf{P}(B_n) &= \mathbf{P}\left(S_{\theta^n} - S_{\theta^{n-1}} \geq \sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))}\right) \\ &= \mathbf{P}\left(S_{\theta^n - \theta^{n-1}} \geq \sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{\frac{\sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))}}{\sqrt{\theta^n-\theta^{n-1}}}}^\infty e^{-y^2/2}dy\end{aligned}$$

Where the first equality holds using the Markov property of the sums (equivalently, look at the definition as sums of $X_i's$ and the fact the $X_i's$ are i.i.d.), and the second equality is coming asymptotically as $\theta^n - \theta^{n-1} \to \infty$ from the central limit theorem. Now, let $\lambda = \frac{\sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))}}{\sqrt{\theta^n-\theta^{n-1}}}$ be the lower bound of the integral and use Mill's inequality to get:

$$\begin{aligned}\mathbf{P}(B_n) &\geq \frac{1}{\sqrt{2\pi}}\frac{\lambda}{\lambda^2+1}e^{-\lambda^2/2} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\lambda+\lambda^{-1}}e^{-\lambda^2/2}\end{aligned}$$

But now notice that $\lambda = \frac{\sqrt{(2-\epsilon)\theta^n \log(\log(\theta^n))}}{\sqrt{\theta^n - \theta^{n-1}}} \approx \sqrt{2-\epsilon}\frac{\sqrt{\log n}}{\sqrt{1-\theta^{-1}}}$, so $\lambda^2 \approx (2-\epsilon)\frac{\log n}{1-\theta^{-1}}$. So our estimate is:

$$\begin{aligned}\mathbf{P}(B_n) &\geq C\frac{1}{\sqrt{\log n}+\sqrt{\log n}^{-1}}\exp\left(\frac{2-\epsilon}{2(1-\theta^{-1})}\log n\right)\\ &\geq Cn^{-\left(\frac{1-\epsilon/2}{1-\theta^{-1}}\right)}\log n^{-1/2}\end{aligned}$$

Where C's are some constants. By choosing θlarge enough, $\frac{1-\epsilon/2}{1-\theta^{-1}} < 1$ and this will not be summable! Have then B_n occurs infinitely often.

Now, we will show that these events $B_n$occurring infinitely often will be enough to see that $S_{\theta^n} \geq \sqrt{(2-2\epsilon)\theta^n \log(\log(\theta^n))}$ infinitely often too. To do this we will use the first half of the law of the iterated logarithm we already proved, namely that for any $\eta > 0$, the events $\{S_k > \sqrt{(2+\eta)k\log(\log k))}\}$happen only finitely often with probability 1. By symmetry, we'll have the events $\{S_k < -\sqrt{(2+\eta)k\log(\log k))}\}$happen only finitely often too. Hence, the events $A_n = \left\{S_{\theta^{n-1}} < -\sqrt{(2+\eta)\theta^{n-1}\log(\log\theta^{n-1}))}\right\}$ happens only finitely often with probability 1. Now, since the $B_n's$ occur infinitely often with probability 1, and the $A_n's$ occur only finitely often with probability 1, the events $B_n \cap A_n^c$ will occur infinitely often with probability 1 too. This will give us the infinite sequence we need, for on the event $B_n \cap A_n^c$ we have the inequalities:

$$\begin{aligned}S_{\theta^n} - S_{\theta^{n-1}} &\geq \sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))}\\ S_{\theta^{n-1}} &\geq -\sqrt{(2+\eta)\theta^{n-1}\log(\log\theta^{n-1}))}\end{aligned}$$

Hence, with probability 1, we have that for infinitely many values of n:

$$\begin{aligned}S_{\theta^n} &\geq \sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))} + S_{\theta^{n-1}}\\ &\geq \sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))} - \sqrt{(2+\eta)\theta^{n-1}\log(\log\theta^{n-1}))}\\ &\geq \sqrt{(2-\epsilon)\theta^n\log(\log(\theta^n))} - \sqrt{\frac{(2+\eta)}{\theta}\theta^n\log(\log\theta^n))}\\ &= \left(\sqrt{2-\epsilon} - \sqrt{\frac{2+\eta}{\theta}}\right)\sqrt{\theta^n\log(\log(\theta^n))}\end{aligned}$$

So by fixing η, (any choice will do) and then choosing θ large enough we can make the coefficient $\left(\sqrt{2-\epsilon} - \sqrt{\frac{2+\eta}{\theta}}\right) \geq \sqrt{2-2\epsilon}$. (Note that this doesn't disrupt our choice of θ previously because that too was a choice to make θ large, so we can always find θso big to suit both our needs.) We have then that for infinitely many n:

$$\frac{S_{\theta^n}}{\sqrt{\theta^n\log(\log(\theta^n))}} \geq \sqrt{2-2\epsilon}$$

So then:

$$\mathbf{P}\left(\limsup_{n\to\infty}\frac{S_n}{\sqrt{n\log(\log n))}} \geq \sqrt{2-2\epsilon}\right) = 1$$

□

The two halves of the law of the iterated logarithm give the full result:

$$\mathbf{P}\left(\limsup_{n\to\infty}\frac{S_n}{\sqrt{n\log(\log n)}}=\sqrt{2}\right)=1$$

CHAPTER 3

Ergodic Theorem

1. Motivation

The study of Ergodic Theory was first motivated by statistical mechanics. Here, one is interested in the long term average of systems. For example, say we have some particles with position $Q(t)$ at time t, and momentum $P(t)$ at time t. Let f be a function on this state space, for example f might be the pressure/temperature/some other macroscopic variable. Can we find a distribution G so that:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(Q(s),P(s))ds=\int f dG$$

Gibbs et al. worked on this problem and it turns out that $G=\frac{1}{Z}e^{-H/kT}$ with Z the partition function, H the Hamiltonian, T temperature, and k Boltzmann's constant has this! These types of long term averaging things can be useful. We will start with a simple example.

Example 1.1. Let $\Omega=[0,1)=\{\theta:\ 0\le\theta<1\}$ where we think of Ω as a circle with perimeter 1 (and $\theta$the position on the circle). For some fixed angle ω, let $T:\Omega\to\Omega$ be rotation by ω, that is $T(\theta)=\theta+\omega \mod 1$. This is clearly measure preserving in the sense that for any set B we have that $m(B)=m(T^{-1}(B))$ where m is the usual Lebesgue measure. Could it be that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}f\left(T^n x\right)=\int_0^1 f(s)ds$$

If $\omega$is rational, this doesn't have a chance, because T^neventually cycles back to the identity, so $T^n x$ will only sample finitely many points. However, if $\omega$is irrational, this is true! We can prove it in this case using Fourier analysis. When $f(x)=e^{2\pi imx}$, for $m\in\mathbb{N}$, we have the geometric series:

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}f\left(T^n x\right) &= \frac{1}{N}\sum_{n=0}^{N-1}e^{2\pi im(x+n\omega)}\\ &= \frac{1}{N}e^{2\pi mx}\frac{e^{2\pi imN\omega}-1}{e^{2\pi im\omega}-1}\\ &\to 0\\ &= \int_0^1 f(s)ds\end{aligned}$$

Where the fact that ωirrational ensures that $e^{2\pi im\omega}-1\neq 0$. In the case $m=0$, f is constant, so of course the result holds. Now for any $f\in C^2(\Omega)$, we can expand

f as a Fourier series to see the result holds. This lets us calculate for example:

$$\lim_{N\to\infty}\frac{\#\{k\leq N:\ x+k\omega\in(a,b)\}}{N}=b-a$$

For if $f=\mathbf{1}_{(a,b)}$ notice that $\frac{\#\{k\leq N:\,x+k\omega\in(a,b)\}}{N}=\frac{1}{N}\sum_{n=0}^{N-1}f\left(T^{n}x\right)$. By approximating f by C^2functions (in the L^1sense) from above and below, and applying the limit calculated above, we get the result.

Is there away we can do this kind of thing using probability methods (rather than Fourier)? The next result is a nice theorem in this direction.

2. Birkhoff's Theorem

Theorem 2.1. *[Birkoff-Khinchin Ergodic Theorem] Say $(\Omega,\mathcal{F},\mathbf{P})$ is a probability space. Suppose $T:\Omega\to\Omega$ is a measure preserving map, in the sense that $\mathbf{P}(T^{-1}(B))=\mathbf{P}(B)$ for all $B\in\mathcal{F}$. Let $\mathcal{F}_0=\{A\in\mathcal{F}:T^{-1}A=A\ a.e.\}$ be the field of T invariant events. For $f:\Omega\to\mathbb{R}$ a random variable with $\mathbf{E}(|f|)<\infty$, we have almost surely:*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}f\left(T^{n}x\right)=\mathbf{E}\left(f|\mathcal{F}_0\right)$$

Corollary 2.2. In the case that $\mathcal{F}_0$ is the trivial field, $\mathbf{E}\left(f|\mathcal{F}_0\right)=\mathbf{E}(f)$ is a constant, so this is exactly the thing we had above. This happens precisely when $T^{-1}A=A\Rightarrow\mathbf{P}(A)=0$ or 1. In this case we say that the map T is "ergodic".

The proof of this theorem relies on the following lemma.

Lemma 2.3. *[Maximal Ergodic Lemma] Say $(\Omega,\mathcal{F},\mathbf{P})$ is a probability space. Suppose $T:\Omega\to\Omega$ is a measure preserving map, in the sense that $\mathbf{P}(T^{-1}(B))=\mathbf{P}(B)$ for all $B\in\mathcal{F}$. Say $f:\Omega\to\mathbb{R}$ a random variable with $\mathbf{E}(|f|)<\infty$. Let $S_n=\sum_{k=1}^{n-1}f(T^kx)$ and let $A=\{\sup_{n\geq1}S_n>0\}$ be the event that this is positive at some point. Then:*

$$\mathbf{E}\left(f\mathbf{1}_A\right)=\int_A fd\mathbf{P}>0$$

Proof. Define $f^+(x)=f(Tx)$ and let $m_n=\max\{0,S_1,S_2,\dots S_n\}$, and m_n^+in the same way, replacing f by f^+ in the definition of S_k. Notice that by this definition the m_n's are non-decreasing. Notice that the event $A=\{\sup_{n\geq1}S_n>0\}$ is the same as saying $m_n>0$ for n large enough. For this reason, it will be enough to restrict our attention to the events $\{m_n>0\}$. Notice that if we are in the event $\{m_n>0\}$ then we have:

$$\begin{aligned}
S_1+m_n^+ &= S_1+\max\{0,S_1^+,S_2^+,\dots S_n^+\}\\
&= S_1+\max\{0,S_2-S_1,S_3-S_1,\dots S_{n+1}-S_1\}\\
&= \max\{S_1,S_2,\dots S_{n+1}\}\\
&= m_{n+1}
\end{aligned}$$

Where we used that we're on the event$\{m_n>0\}$ in the last step to see the last equality, and we used $S_n^+=\sum_0^{n-1}f(T^kTx)=\sum_1^n f(Tx)=S_{n+1}-S_1$in the second equality. We have then:

$$\begin{aligned}\mathbf{E}\left(f\mathbf{1}_{\{m_n>0\}}\right) &= \mathbf{E}\left(S_1\mathbf{1}_{\{m_n>0\}}\right)\\ &= \mathbf{E}\left((m_{n+1}-m_n^+)\mathbf{1}_{\{m_n>0\}}\right)\\ &= \mathbf{E}\left(m_{n+1}\mathbf{1}_{\{m_n>0\}}\right)-\mathbf{E}\left(m_n^+\mathbf{1}_{\{m_n>0\}}\right)\\ &\geq \mathbf{E}\left(m_{n+1}\mathbf{1}_{\{m_n>0\}}\right)-\mathbf{E}\left(m_n^+\right)\end{aligned}$$

The last inequality holds since on the event $\{m_n = 0\}$,we have $S_1 \leq 0$, so $m_n^+ = m_{n+1} - S_1 \geq m_{n+1} \geq 0$, so $\mathbf{E}\left(m_n^+\mathbf{1}_{\{m_n=0\}}\right) \geq 0$. Hence $\mathbf{E}\left(m_n^+\right) = \mathbf{E}\left(m_n^+\mathbf{1}_{\{m_n>0\}}\right) + \mathbf{E}\left(m_n^+\mathbf{1}_{\{m_n=0\}}\right) \geq \mathbf{E}\left(m_n^+\mathbf{1}_{\{m_n>0\}}\right)$. From here, we note that $\mathbf{E}(m_n^+) = \mathbf{E}(m_n)$ since the map T is measure preserving, and the only difference between $m_n^+$and m_n is the map $x \to Tx$. Have then:

$$\begin{aligned}\mathbf{E}\left(f\mathbf{1}_{\{m_n>0\}}\right) &\geq \mathbf{E}\left(m_{n+1}\mathbf{1}_{\{m_n>0\}}\right)-\mathbf{E}\left(m_n\right)\\ &= \mathbf{E}\left(m_{n+1}\mathbf{1}_{\{m_n>0\}}\right)-\mathbf{E}\left(m_n\mathbf{1}_{\{m_n>0\}}\right)\\ &= \mathbf{E}\left((m_{n+1}-m_n)\mathbf{1}_{\{m_n>0\}}\right)\\ &\geq 0\end{aligned}$$

The second equality holds since $m_n \geq 0$ always holds, and the last inequality holds since the $m_n's$ are non-increasing. Finally, to get the result, notice that $\{m_n > 0\}$ is increasing to $\{\sup S_n > 0\}$, so by a monotone convergence theorem result, we have:

$$\mathbf{E}\left(f\mathbf{1}_{\{\sup S_n>0\}}\right) = \lim_{n\to\infty}\mathbf{E}\left(f\mathbf{1}_{\{m_n>0\}}\right) \geq 0$$

□

With this in hand, we can prove Birkhoff's theorem:

THEOREM 2.4. *[Birkoff-Khinchin Ergodic Theorem] Say* $(\Omega, \mathcal{F}, \mathbf{P})$ *is a probability space. Suppose* $T : \Omega \to \Omega$ *is a measure preserving map, in the sense that* $\mathbf{P}(T^{-1}(B)) = \mathbf{P}(B)$ *for all* $B \in \mathcal{F}$. *Let* $\mathcal{F}_0 = \{A \in \mathcal{F} : T^{-1}A = A \text{ a.e.}\}$ *be the field of* T*-invariant events. For* $f : \Omega \to \mathbb{R}$ *a random variable with* $\mathbf{E}(|f|) < \infty$, *we have almost surely:*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right) = \mathbf{E}\left(f|\mathcal{F}_0\right)$$

PROOF. Firstly, we will argue that $\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right)$ converges a.s. to some random variables, and then we (as usual) check that it has the two defining properties of conditional expectation.

Define $S_N = \sum_{n=0}^{N-1} f(T^n x)$ as before, so that we are interested in the sum S_n/n. Suppose by contradiction that $\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right)$ does not converge a.s.. By the usual trick with rational numbers then, we can find $a, b \in \mathbb{R}$ so that the even $A = \left\{\liminf \frac{S_n}{n} \leq a < b \leq \limsup \frac{S_n}{n}\right\}$ has$\mathbf{P}\left(A\right) > 0$. Notice moreover, that A is a T-invariant event, i.e. $x \in A \Rightarrow Tx \in A$, since applying T shifts the terms in S_n by one, which does not affect the limsup or liminf of S_n/n. (Indeed, these don't depend on finitely many of the terms!). For this reason, we may define a new probability measure on the set A, namely we think of $(A, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ as a new probability space, with $\tilde{\mathcal{F}}$=$\{A \cap B : B \in \mathcal{F}\}$and $\tilde{\mathbf{P}}(E) = \mathbf{P}(E)/\mathbf{P}(A)$. The fact that A is T-invariant means that $T^n x \in A$ whenever $x \in A$ so we can still talk about $S_n$and so on on this space. The fact that $\mathbf{P}(A) > 0$ means that there is no problem re-normalizing like this. So we have now $\tilde{\mathbf{P}}(A) = 1$ is the whole space. With this new space as our framework, we let $f'(\omega) = f(\omega) - b$, then we get new sums S'_n with $\frac{S'_n}{n} = \frac{S_n}{n} - b$ and then $A = \left\{\liminf \frac{S'_n}{n} \leq a - b < 0 \leq \limsup \frac{S'_n}{n}\right\}$. Notice then that $\tilde{\mathbf{P}}(\limsup \frac{S_n}{n} \geq 0) \geq \tilde{\mathbf{P}}(A) = 1$ so then $\tilde{\mathbf{P}}(\{\sup S'_n > 0\}) = 1$ is the whole space A. Have then by the maximal ergodic lemma that:

$$0 < \tilde{\mathbf{E}}(f'\mathbf{1}_{\{\sup S'_n > 0\}}) = \tilde{\mathbf{E}}(f') = \tilde{\mathbf{E}}(f) - b$$

The same argument on $f'''(\omega) = a - f(\omega)$ gives:

$$0 < \tilde{\mathbf{E}}(f''\mathbf{1}_{\{\sup S''_n > 0\}}) = a - \tilde{\mathbf{E}}(f)$$

But this is a contradiction now, for we have:

$$a > \tilde{\mathbf{E}}(f) > b$$

Which is impossible since $a < b$. This contradiction means that its impossible to separate the liminf and the limsup like this, in other words we have almost sure convergence.

Next it remains only to see that the random variable that this converges to is $\mathbf{E}(f|\mathcal{F}_0)$. Let us denote Firstly, notice that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$ by $\bar{f}$. We must show $\bar{f}$ is $\mathcal{F}_0$ measurable and that $\mathbf{E}(\bar{f}\mathbf{1}_A) = \mathbf{E}(f\mathbf{1}_A)$ for all $A \in \mathcal{F}_0$. Notice that applying $x \to Tx$ does not change $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$ as it only effects finitely many terms. This shows that $\bar{f}(x) = \bar{f}(Tx)$ This is the reason why $\bar{f}$ is $\mathcal{F}_0$ measurable. More formally, to see that $\bar{f}^{-1}(B)$ is T-invariant for any Borel set B, just write out the definitions:

$$\begin{aligned}
T(\bar{f}^{-1}(B)) &= \left\{Tx \in \Omega : \bar{f}(x) \in B\right\} \\
&= \left\{Tx \in \Omega : \bar{f}(Tx) \in B\right\} \\
&= \left\{y \in \Omega : \bar{f}(y) \in B\right\} \\
&= \bar{f}^{-1}(B)
\end{aligned}$$

So indeed, $\bar{f}^{-1}(B) \subset \mathcal{F}_0$ means $\bar{f}$ is $\mathcal{F}_0$ measurable. To see that $\bar{f}$ has the right expectation values, we first see prove the result for indicator functions and then use the "ladder" of integration to get the result we need. Consider that for sets $A \in \mathcal{F}_0$ and $B \in \mathcal{F}$ we have:

$$\begin{aligned}
\int_A \mathbf{1}_B(x) d\mathbf{P} &= \int \mathbf{1}_A(x)\mathbf{1}_B(x) d\mathbf{P} \\
&= \int \mathbf{1}_A(Tx)\mathbf{1}_B(Tx) d\mathbf{P} \\
&= \int \mathbf{1}_A(x)\mathbf{1}_B(Tx) d\mathbf{P} \\
&= \int_A \mathbf{1}_B(Tx) d\mathbf{P}
\end{aligned}$$

Where the second equality is using the fact that $\mathbf{P}$ is T-invariant and the third equality uses the fact that $A \in \mathcal{F}_0 \Rightarrow \mathbf{1}_A(x) = \mathbf{1}_A(Tx)$. Since$\int_A \mathbf{1}_B(\mathrm{x})\mathrm{d}\mathbf{P}$=$\int_A \mathbf{1}_B(\mathrm{Tx})\mathrm{d}\mathbf{P}$, by following along with the construction of the Lebesgue integral starting from indicator functions, we conclude that $\int_A f(x)d\mathbf{P} = \int_A f(Tx)d\mathbf{P}$ for any integrable f. Applying this inductively, we see that for any $N \in \mathbb{N}$ that:

$$\int\limits_A \frac{1}{N}\sum_{k=0}^{N-1} f(T^k x)d\mathbf{P} = \int\limits_A f(x)d\mathbf{P}$$

When $\bar{f}$ is bounded, we can take the limit as $N \to \infty$ and use the bounded convergence theorem to conclude:

$$\begin{aligned} \int\limits_A \bar{f}d\mathbf{P} &= \lim_{N\to\infty}\int\limits_A \frac{1}{N}\sum_{k=0}^{N-1} f(T^k x)d\mathbf{P} \\ &= \int\limits_A f(x)d\mathbf{P} \end{aligned}$$

For general $\bar{f}$, we can use a truncation argument and the monotone convergence theorem to get finish the result. □

Example 2.5. If we look at our first example of rotation by an angle ω, we concluded (using Fourier analysis) that when $\omega$is irrational and f has a Fourier series that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right) = \int\limits_0^1 f(s)ds$$

By Birkhoff's theorem, we know that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right) = \mathbf{E}(f|\mathcal{F}_0)$$

So we conclude that: $\int_0^1 f(s)ds = \mathbf{E}(f|\mathcal{F}_0)$. Since this holds for every f, it must be that $\mathcal{F}_0$ is the trivial field. Notice that this improves our result a little bit, since we may now apply it to any f integrable, not just f which are C^2.

Example 2.6. In the first example, we were essentially looking at $\frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i m(x+n\omega)}$. Now lets ask about the series: $\frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i m(2^n x)}$. This is harder to handle with Fourier techniques, but we can still use Birkhoff's theorem. Again take $\Omega = [0,1)$ to be our space, but instead of thinking of this as a circle, think of this as binary sequence (which are the binary expansions of each number between 0 and 1), $\Omega = \{0.e_1e_2\ldots :\ e_i = \pm 1\}$. Let $T : \Omega \to \Omega$ by $T(0.e_1e_2e_3\ldots) = 0.e_2e_3\ldots$. This translates to $T(x) = 2x \mod 1$ (this is the reason that applying it N times gives $2^N x$). It's not hard to verify that this is measure preserving. By the Kolmogorov Zero-One law, the field $\mathcal{F}_0$of T-invariant events must be the trivial field, for by applying T N times, we see that an event $A \in \mathcal{F}_0$cannot depend on the first N digits $e_1, e_2, \ldots e_N$. Since this works for any N, this is a subset of the tail field, which by K-0-1 is trivial. Hence, by Birkhoff's Theorem, we have:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right) = \mathbf{E}(f|\mathcal{F}_0) = \mathbf{E}(f) = \int_0^1 f d\mathbf{P}$$

For the Fourier basis function $f(x) = e^{2\pi i m x}$, this is saying that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i m(2^n x)} = 0$$

Example 2.7. We can use Birkhoff's theorem to give yet another proof of the strong law of large numbers. Let $(X_1, X_2, \ldots)$ be a sequence of i.i.d. random variables with finite mean and let Ω be the probability space for these sequences. Define $T : \Omega \to \Omega$ by $T(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots)$. Notice that since the $X's$ are i.i.d. that this is measure preserving. As in example 2, the Kolmogorov zero one law tells us the field $\mathcal{F}_0$ of T-invariant is trivial. Let $f(x_1, x_2, \ldots) = x_1$. By Birkhoff's theorem:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} x_n = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right) = \mathbf{E}(f|\mathcal{F}_0) = \mathbf{E}(f) = \mathbf{E}(X_1)$$

Which is exactly the strong law of large numbers.

3. Continued Fractions

One way to specify a number in $x \in [0, 1)$ is the binary expansion. Each binary digit tells you "which half" of the number line x is in. e.g. first digits says if its in $\left[0, \frac{1}{2}\right]$or $\left[\frac{1}{2}, 1\right]$, and then we treat that interval like $[0, 1)$ and start over again for the next digit. Another way to do this game would be to draw the harmonic series $\frac{1}{n}$ on the number line, and then specify which interval $[\frac{1}{n+1}, \frac{1}{n})$ the number is in. Call this first number n_1, and we'll have then that $\frac{1}{n_{1+1}} \le x < \frac{1}{n_1}$. From this we may conclude that:

$$x = \frac{1}{n_1 + \epsilon_1}$$

For some $\epsilon_1 \in [0, 1)$. Play the same game again for ϵ_1, and we get:

$$x = \frac{1}{n_1 + \frac{1}{n_2+\epsilon_2}}$$

Continuing this indefinitely gives us the "continued fraction expansion" for x. Since this is hard to write, we will adopt the convention that $x = [n_1; n_2; n_3; \ldots]$ to mean the continued fraction expansion n_1 and then n_2 and so on.

Proposition 3.1. *If the sequence* $[n_1; n_2; n_3; \ldots]$ *is cyclic (that is it repeats after some finite number of steps), then* $x = [n_1; n_2; n_3; \ldots]$ *is algebraic.*

PROOF. The easiest way to see this is an example. Suppose we look at $x = [1; 1; 1; \ldots]$. Then:

$$x = \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}$$

So then:

$$\frac{1}{x} = 1 + x$$

But then $x^2 - x + 1 = 0$, so x is the root of a quadratic equation. In this case $x = \frac{\sqrt{5}-1}{2}$ is the golden section. In general, if the continued fraction expansion is periodic after N steps, then x will be the root of an $N+1$ order polynomial. □

DEFINITION 3.2. We write $x = [n_1; n_2; n_3; \ldots]$ to mean:

$$x = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \ldots}}}$$

PROBLEM 3.3. Let $T : (0,1) \to (0,1)$ by $T([n_1; n_2; \ldots]) = [n_2; n_3; \ldots]$. This is the map $T(x) = \frac{1}{x} \mod 1$. Is there a probability density $\mathbf{P}$ we can put on $(0,1)$ so that T will be measure preserving?

PROOF. [Gauss] The probability density $d\mathbf{P} = \frac{1}{\log 2} \frac{1}{1+x} dx$ will do the trick! Indeed, just notice that by the definition of T that:

$$T^{-1}(a,b) = \bigcup_{n=1}^{\infty} \left(\frac{1}{b+n}, \frac{1}{a+n} \right)$$

So then the requirement $\mathbf{P}(T^{-1}(a,b)) = \mathbf{P}(a,b)$ gives (using ρ as a probability density function):

$$\int_a^b \rho(x) dx = \sum_{n=1}^{\infty} \int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \rho(x) dx$$

Taking the derivative w.r.t. b here gives:

$$\rho(x) = \sum_{n=1}^{\infty} \rho\left(\frac{1}{x+n} \right) \frac{1}{(x+n)^2}$$

This is hard to solve, but its easy to verify that $\rho(x) = \frac{1}{1+x}$ works, since the LHS is $\frac{1}{1+x}$ while the RHS is:

$$\begin{aligned}\sum_{n=1}^{\infty} \rho\left(\frac{1}{x+n}\right)\frac{1}{(x+n)^2} &= \sum_{n=1}^{\infty} \frac{1}{1+\frac{1}{x+n}}\frac{1}{(x+n)^2} \\ &= \sum_{n=1}^{\infty} \frac{x+n}{1+(x+n)}\frac{1}{(x+n)^2} \\ &= \sum_{n=1}^{\infty} \frac{1}{(x+n+1)(x+n)} \\ &= \sum_{n=1}^{\infty} \frac{1}{x+n} - \frac{1}{x+n+1} \\ &= \frac{1}{x+1}\end{aligned}$$

Which is a telescoping sum so we can evaluate it exactly. The factor of $\frac{1}{\log 2}$ normalizes ρ so that $\int_0^1 \rho(x)dx = 1$. Indeed:

$$\int_0^1 \frac{1}{\log 2}\frac{1}{x+1}dx = \frac{1}{\log 2}\left[\log(1+x)\right]_0^1 = \frac{\log 2 - \log 1}{\log 2} = 1$$

□

THEOREM 3.4. *The shift function $T : [0,1] \to [0,1]$ given by $T([n_1; n_2; , \ldots]) = [n_2; n_3; \ldots]$is ergodic.*

PROOF. Fix $N \in \mathbb{N}$ and a list of integers $n_1, n_2, \ldots, n_N$. Now define:

$$n(x) := \frac{1}{n_1 + \frac{1}{n_2 + \ldots + \frac{1}{n_N + x}}}$$

For each choice of $n_1, n_2, \ldots, n_N$, the image of $[0,1]$ through $n(x)$ is an interval whose endpoints are $n(0)$ and $n(1)$. As N increases, the interval $[n(0), n(1)]$ gets smaller and smaller. An easy proof by induction shows that $n(x)$ can be written as:

$$n(x) = \frac{Ax+B}{Cx+D}$$

For $A, B, C, D \in \mathbb{R}$with $0 \leq A \leq B$ and $1 \leq C \leq D$ and with $AD - BC = \pm 1$ where the sign depends on the parity of N. Now, let $I = [n(0), n(1)]$ and let $J = (a, b)$ be an arbitarty interval.

CLAIM. $|I \cap T^{-N}(J)| \geq \frac{1}{2}|I||J|$ holds for all $N \in \mathbb{N}$.

PROOF. Take $x \in I \cap T^{-N}(J)$. Notice that $x \in I$ if and only if $x = n(y)$ for some $y \in [0,1]$ by definition of I. So we can write x as a continued fraction $x = [n_1; n_2; \ldots; n_{N-1}; n_N + y]$. On the other hand, $x \in T^{-N}(J)$ if and only if $T^N x \in J$. But $T^N x = T^N([n_1; n_2; \ldots; n_{N-1}; n_N; y]) = y$ by definition of T. This shows that $x \in T^{-N}(J)$ if and only if $y \in J$.

Have then, using the the observation that n is a fractional linear transformation, that:

$$I \cap T^{-N}(J) = \{n(y) : y \in J\} = [n(a), n(b)]$$

This shows:

$$\begin{aligned}
|I \cap T^{-N}(J)| &= |n(b) - n(a)| \\
&= \left| \frac{Ab+B}{Cb+D} - \frac{Aa+B}{Ca+D} \right| \\
&= \left| \frac{b-a}{(Ca+D)(Cb+D)} \right| \\
&\geq \frac{|b-a|}{(C+D)^2} \text{ since } a,b < 1 \\
&\geq \frac{1}{2}|b-a||I| \\
&= \frac{1}{2}|J||I|
\end{aligned}$$

The last inequality holds by writing out $|I|$and using $AD - BC = \pm 1$ and the fact that $1 \leq C \leq D$ so that $C + D \leq 2D$:

$$\begin{aligned}
|I| &= |n(0) - n(1)| \\
&= \left| \frac{A+B}{C+D} - \frac{B}{D} \right| \\
&= |AD - BC| \frac{1}{D(C+D)} \\
&= \frac{1}{D(C+D)} \\
&\leq \frac{2}{(C+D)^2}
\end{aligned}$$

□

Finally, to see that T is ergodic, take any Borel set $B \in \mathcal{F}$. By approximating B by intervals, the inequality from the claim still holds:

$$\left| I \cap T^{-N} B \right| \geq \frac{1}{2}|I||B|$$

Take any set A now. Again, by approximting A by intervals I, we can use the above inequality to get:

$$\left| A \cap T^{-N} B \right| \geq \frac{1}{2}|A||B|$$

This gives what we want, for if B is $T-$invariant, we have $T^{-N}B = B$ for every N. The choice $A = B^c$ in the above gives:

$$\begin{aligned}
\frac{1}{2}|B|B^c| &\leq |B^c \cap T^{-N}B| \\
&= |B^c \cap B| \\
&= 0
\end{aligned}$$

So $|B||B^c| = 0$, which is only possible if $|B| = 1$ or $|B| = 0$. This is saying all T invarant sets are either measure zero or full measure. In other words, T is ergodic. □

CHAPTER 4

Brownian Motion

1. Motivation

Our aim is to discuss a stochastic process on $[0,1]$ (that is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a collection of random variables $B_t(\omega)$, for $t \in [0,1]$) which has the following properties:

- $B_0(\omega) = 0$ for every $\omega \in \Omega$
- Fix a $T \in [0,1]$, and define for $t > T, B_t^+ = B_{T+t} - B_t$. We want $B_t^+$to look statistically identical to B_t. (This says the process has some sort of "time homogenous" property.)
- We want B_t^+ as defined above to be independent of B_t. (This says that the process has some sort of Markov property)
- $\mathbf{E}(B_t^2) < \infty$
- $\mathbf{E}(B_t) = 0$
- $B_t(\omega)$ is continuous for every (or almost every) $\omega \in \Omega$.

This process is supposed to describe something like a piece of dust that you can see sometimes wiggling about in a sunbeam. Notice that the time homogenous and Markov property together means we can write:

$$B_T = \sum_{k=1}^{N} B_{\frac{kT}{N}} - B_{\frac{(k-1)T}{N}}$$

Which is a sum of many independent increments. By the central limit theorem, this is suggesting $B_t \sim N(0, \sigma^2)$ is normally distributed (to get this more rigorously would take a bit more work, since the above set up is not exactly the set up for the central limit theorem). This is often taken as an "axiom" :

- $B_t \sim N(0, \sigma^2)$

A quick calculation shows that $\sigma^2 \propto t$. Let $f(t) = \sigma^2$ be the variance for B_t. Then:

$$\begin{aligned} f(t+s) &= \mathbf{E}\left((B_{t+s})^2\right) \\ &= \mathbf{E}\left((B_{t+s} - B_s + B_s)^2\right) \\ &= \mathbf{E}\left((B_{t+s} - B_s)^2\right) + \mathbf{E}\left(B_s^2\right) + 2\mathbf{E}\left((B_{t+s} - B_s)B_s\right) \\ &= f(t) + f(s) + 2 \cdot 0 \end{aligned}$$

Where we used the time homogenous property and the Markov property. This functional relation means that $f(t)$ must be linear! $f(0) = 0$ holds since B_0 is known exactly. Hence $f(t) = c \cdot t$. It doesn't hurt to take $c = 1$, since anything we get can be rescaled for other values of c if need be. Sometimes this is taken as the "axiom":

(1) $B_t \sim N(0, t)$

The following resulting property also turns out to be very useful:

PROPOSITION 1.1. $\mathbf{E}(B_a B_b) = \min(a, b)$

PROOF. Suppose W.O.L.O.G. $a < b$. Then: $\mathbf{E}(B_a B_b) = \mathbf{E}(B_a(B_b - B_a + B_a)) = \mathbf{E}(B_a(B_b - B_a)) + \mathbf{E}(B_a^2) = 0 + a = \min(a, b)$ □

It remains to see that such a process really exists. The main difficulty is proving that the process is *continuous*. There is more than one way to skin the cat for this; each method is useful because it gives a different insight into what is going on.

2. Levy's Construction

We will construct Brownian motion on $t \in [0,1]$ as a uniform limit of continuous functions B_t^N, as $N \to \infty$. Each B_t^N will be an approximation of the Brownian motion that is piecewise linear between the dyadic rationals of the form $\frac{a}{2^N}$. The real trick in the construction is the remarkable observation that the corrections from B_t^N to B_t^{N+1} are independent of the construction so far up to level N, which is the crucial fact that makes the construction so nice and allows it to converge. The crucial fact about Brownian motion that makes this possible is captured in the below proposition:

PROPOSITION 2.1. *Let B_t be a Brownian path and $0 < a < b < 1$. Consider the line segment joining B_a and B_b: $l(t) = B_a + (t-a)\frac{B_b - B_a}{b-a}$. Consider the value of the Brownian path at the midpoint time $B_{\frac{a+b}{2}}$. The difference from this point to the line $l(t)$ is independent of B_b and B_a. That is to say: $X = B_{\frac{a+b}{2}} - l(\frac{a+b}{2}) = B_{\frac{a+b}{2}} - \frac{1}{2}B_a - \frac{1}{2}B_b$, is independent of B_a and B_b. Moreover, X is normally distributed $X \sim N(0, \frac{1}{4}(b-a))$.*

PROOF. Firstly, we notice that the random variables X, B_a,and B_b are have a joint normal distribution. This can be seen without much difficulty by expanding the definition of X to write any linear combination of $X, B_a$and B_b as a linear combination of $B_{\frac{a+b}{2}}, B_a$,and B_b. From here, rewrite as a linear combination of $B_a, B_{\frac{a+b}{2}} - B_a$, and $B_b - B_{\frac{a+b}{2}}$. By the hypothesis on our Brownian motion, each of these are independent Gaussian variables, so any linear combination of them is again Gaussian. Hence any linear combination of X, B_a and B_b is Gaussian. This property is a characterization of the joint Gaussian distribution. The observation that X, B_a and B_b are jointly normal substantially simplifies the verification of their independence, as for jointly normal distributions they are independent if and only if they are uncorrelated. From here we calculate (with the help of the useful covariance relation):

$$\begin{aligned}
\mathbf{E}(B_a X) &= \mathbf{E}\left(B_a(B_{\frac{a+b}{2}} - \frac{1}{2}B_a - \frac{1}{2}B_b)\right) \\
&= \mathbf{E}\left(B_a B_{\frac{a+b}{2}}\right) - \frac{1}{2}\mathbf{E}\left(B_a^2\right) - \frac{1}{2}\mathbf{E}(B_a B_b) \\
&= a - \frac{1}{2}a - \frac{1}{2}a \\
&= 0
\end{aligned}$$

A similar calculation holds for $\mathbf{E}(B_b X)$. Since these are uncorrelated and jointly normal, they are independent. A quick calculation using the covariance relation again gives $X \sim N(0, \frac{1}{4}(b-a))$ □

This remarkable fact gives us a nice idea to construct Brownian motion starting with an infinite sequence of standard $\mathbf{E}(Z)=0, \mathbf{E}(Z^2)=1$ i.i.d Gaussian variables $(Z_0, Z_1, Z_2, \ldots)$. The idea is to first construct $B_0=0, B_1=Z_0$. Then, once B_0, and B_1 are constructed by the above proposition, we know that $B_{1/2}-\frac{1}{2}B_0-\frac{1}{2}B_1$ can be modeled by $\frac{1}{4}Z_1$, so set $B_{1/2}=\frac{1}{2}B_1+\sqrt{\frac{1}{4}}Z_1$. Once $B_0, B_{1/2}, B_1$ are constructed, the above proposition gives us a way to get $B_{\frac{1}{4}}$and $B_{\frac{3}{4}}$using two more normal variables $\sqrt{\frac{1}{8}}Z_2$ and $\sqrt{\frac{1}{8}}Z_3$ and so on.

The above proposition and paragraph is the basic idea. It becomes a bit of a mouthful to write it all down. A confused reader should focus on understanding the construction above before digesting the below details.

To formalize the process, we let B_t^N be the construction at the $N-th$ level of construction, which will have the correct values at points of the form $\frac{a}{2^N}, 0 \le a \le 2^N$. We make fill in in between these points with a piecewise continuous function. After some bookkeeping, the easiest way to write this down is as follows. First define some "tent" functions which make little peaks in the interval $\left[\frac{2k}{2^n}, \frac{2(k+1)}{2^n}\right]$ of unit height:

$$T_{n,k}=\begin{cases} 2^n\left(t-(2k)\right) & t\in\left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right] \\ 2^n\left((2k+2)-t\right) & t\in\left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right] \\ 0 & t\notin\left[\frac{2k}{2^n}, \frac{2(k+1)}{2^n}\right] \end{cases}$$

Notice that for every level n, $0\le k\le 2^{n-1}-1$ means there are 2^{n-1}tents, and notice that these tents are disjoint and of unit height.

Now, at every level of the construction we make sure that B_t^N has the right value at points of the form $\frac{a}{2^N}$ by adding in the right tents with heights distributed by scaled normal functions:

$$B_t^N=Z_0t+\sum_{n=1}^{N}\sum_{k=0}^{2^{n-1}-1}\sqrt{\frac{1}{2^{n+1}}}Z_{n,k}T_{n,k}(t)$$

Explanation of this formula: The "Z_0t" is the initial level 0 construction. The sum $0\le n\le N$ sums over the N levels of construction, and the sum $0\le k\le 2^{n-1}-1$ is over the 2^{n-1} tents that get added on at the $n-th$ level. Each tent has a height distributed like $\sqrt{\frac{1}{2^{n+1}}}Z\sim N(0,\frac{1}{2^{n+1}})$, where $Z\sim N(0,1)$(This is the content of the proposition above!) For convenience, we label the infinite sequence of normal variables so that $Z_{n,k}$ is controlling the height of the $k-th$ tent on the $n-th$ level.

Finally we get the Brownian motion as $B_t=\lim_{N\to\infty}B_t^N$, which puts the Brownian motion on the same probability space as the infinite sequence of normal variables. To see that this is continuous, we show that the convergence is uniform almost surely. Since each B_t^N is continuous, and a uniform limit of continuous functions is continuous, this gives that B_t is continuous.

Proposition 2.2. *The family of functions B_t^N is converging uniformly almost surely.*

Proof. As you might suspect, the trick is to use the right summable sequence with a clever application of the Borel Cantelli lemma. Let $H_n=\max_{t\in[0,1]}\left|\sum_{k=0}^{2^{n-1}-1}\sqrt{\frac{1}{2^{n+1}}}Z_{n,k}T_{n,k}(t)\right|$

be the maximum height contribution to B_t at level n. Since the tent functions $T_{n,k}(t)$ are disjoint, this is $H_n = \sqrt{\frac{1}{2^{n+1}}} \max_{0\le k\le 2^{n-1}-1}(|Z_{n,k}|)$. We now make the following estimate:

$$\begin{aligned}
\mathbf{P}(H_n > 2^{-\frac{n}{2}}\sqrt{2n}) &= \mathbf{P}\left(\max_{0\le k\le 2^{n-1}-1}(|Z_{n,k}|) > 2^{-\frac{n}{2}}2^{\frac{n+1}{2}}2^{\frac{1}{2}}\sqrt{n}\right) \\
&\le 2^{n-1}\mathbf{P}\left(|Z| > 2\sqrt{n}\right) \\
&= 2^{n}\mathbf{P}\left(Z > 2\sqrt{n}\right) \\
&= \frac{2^n}{\sqrt{2\pi}}\int_{2\sqrt{n}}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx \\
&\le \frac{2^n}{\sqrt{2\pi}}\frac{1}{2\sqrt{n}}\exp\left(-\frac{(2\sqrt{n})^2}{2}\right) \text{ (this is Mill's ratio)} \\
&= C\cdot\frac{1}{\sqrt{n}}\cdot\left(\frac{2}{e^2}\right)^n
\end{aligned}$$

Which is a summable sequence! Hence, we know by the Borel Cantelli lemma that this happens only finely often almost surely. That is to say, for almost every $\omega\in\Omega$, we can find $N\in\mathbb{N}$ so that $H_n(\omega)\le 2^{-\frac{n}{2}}\sqrt{2n}$ for all $n > N$. But then we have that for all $p, q > N$and any $t\in[0,1]$:

$$\begin{aligned}
|B_t^p - B_t^q| &= \left|\sum_{n=p+1}^{q}\sum_{k=0}^{2^{n-1}-1}\sqrt{\frac{1}{2^{n+1}}}Z_{n,k}T_{n,k}(t)\right| \\
&\le \sum_{n=p+1}^{q}|H_n| \\
&\le \sum_{n=p+1}^{q}2^{-\frac{n}{2}}\sqrt{2n} \\
&\le \sum_{n=N}^{\infty}2^{-\frac{n}{2}}\sqrt{2n}
\end{aligned}$$

But since $2^{-\frac{n}{2}}\sqrt{2n}$ is summable, this can be made arbitrarily small, and we see then that B_t^N is Cauchy in the uniform norm. Since this holds for almost every $\omega\in\Omega$, we indeed have uniform convergence almost surely. □

Finally, to see that the limiting process is really what we want, we just verify that $\mathbf{E}\left((B_t - B_s)^2\right) = |t-s|$, from which it's easy to check the properties we want. To see this, we just use the density of the dyadic rationals in $[0,1]$. The above construction fixes points of the form $\frac{a}{2^n}$ at step n, that is to say $B_t(\frac{a}{2^n}) = B_t^n(\frac{a}{2^n})$. Hence for t, s dyadic rationals, we have $\mathbf{E}\left((B_t - B_s)^2\right) = \mathbf{E}\left((B_t^n - B_s^n)^2\right) = |t-s|$ which is easily checked by the construction above/the earlier proposition.

For arbitrary t now, but s still taken to be a dyadic rational, we take a sequence of dyadic rationals $t_n \to t$. We have then using Fatou's lemma:

$$\begin{aligned}\mathbf{E}\left((B_t - B_s)^2\right) &= \mathbf{E}\left(\lim_{n\to\infty}(B_{t_n} - B_s)^2\right) \\ &\leq \lim_{n\to\infty}\mathbf{E}\left((B_{t_n} - B_s)^2\right) \\ &= \lim_{n\to\infty}|t_n - s| \\ &= |t-s|\end{aligned}$$

Now consider, for any $n \in \mathbb{N}$:

$$\begin{aligned}\mathbf{E}\left((B_t - B_s)^2\right) &= \mathbf{E}\left((B_t - B_{t_n} - B_s + B_{t_n})^2\right) \\ &= \mathbf{E}\left((B_t - B_{t_n})^2\right) + \mathbf{E}\left((B_s - B_{t_n})^2\right) + 2\mathbf{E}\left((B_t - B_{t_n})(B_s - B_{t_n})\right)\end{aligned}$$

Since this holds for any $n \in \mathbb{N}$, we get:

$$\begin{aligned}\mathbf{E}\left((B_t - B_s)^2\right) &= \lim_{n\to\infty}\left(\mathbf{E}\left((B_t - B_{t_n})^2\right) + \mathbf{E}\left((B_s - B_{t_n})^2\right) + 2\mathbf{E}\left((B_t - B_{t_n})(B_s - B_{t_n})\right)\right) \\ &= 0 + \lim_{n\to\infty}|t_n - s| + 0 \\ &= |t-s|\end{aligned}$$

Where we have observed that the two limits on either side are 0 by using $\mathbf{E}\left((B_t - B_s)^2\right) \leq |t-s|$ in a clever way. First:$\lim_{n\to\infty}\mathbf{E}\left((B_t - B_{t_n})^2\right) \leq \lim_{n\to\infty}|t-t_n| = 0$ and secondly with the help of Holder:

$$\begin{aligned}\lim_{n\to\infty}\left|\mathbf{E}\left((B_t - B_{t_n})(B_s - B_{t_n})\right)\right| &\leq \lim_{n\to\infty}\sqrt{\mathbf{E}((B_t - B_{t_n})^2)}\sqrt{\mathbf{E}((B_s - B_{t_n})^2)} \\ &\leq \lim_{n\to\infty}\sqrt{|t-t_n|}\sqrt{|s-t_n|} \\ &= 0\end{aligned}$$

Once we have $\mathbf{E}\left((B_t - B_s)^2\right) = |t-s|$ for arbitrary t and dyadic s, the same argument repeated again will show that $\mathbf{E}\left((B_t - B_s)^2\right) = |t-s|$ works when both t and s are arbitrary.

3. Construction from Durret's Book

(I call this "Durret's construction" since I read it out of Durret's book: "Brownian Motion and Martingale's in Analysis")

The above construction is pretty elementary and gives all the desired properties. The following construction is a bit more technical, in particular it uses a few extension results like Caratheodory and Kolmogorov. However, it gives immediately that not only is the Brownian motion continuous, but it is Holder continuous for exponents $\gamma < \frac{1}{2}$. This construction uses a few "extension theorems", which are gone over briefly in the appendix.

Definition 3.1. (Constructing Brownian Motion with the Kolmogorov Extension Theorem)

The Kolmogorov Extension Theorem gives us a quick way to define a measure on the space of functions. However, since the space of functions $\{f : T \to \mathbb{R}\}$ is so large, this theorem often gives us a very unwieldy space to work with, one in which we can't get our hands on the properties we want. The construction of Brownian motion below is a great example, constructing with the Kolmogorov theorem is bad, while if we take more care and construct it on only countably many points, we get what we want.

Let $\mathbf{P}_{t_1,t_2,\ldots t_n}(A_1\times A_2\times\ldots\times A_n)=\int\limits_{A_1}dx_1\int\limits_{A_2}dx_2\ldots\int\limits_{A_n}dx_n\Pi_{k=1}^n p_{t_i-t_{i-1}}(x_{i-1},x_i)$, where $p_t(x,y)=\sqrt{2\pi t}^{-1}\exp(-\frac{|y-x|^2}{2t})$. This is naively what you get as the distribution of $B_{t_1},B_{t_{2,\ldots}},B_{t_n}$ if you use the Markov property and normal distribution of the Brownian motion. By Kolmogorov, we get a measure $\mathbf{P}$on the entire space of function $\{f:[0,1]\to\mathbb{R}\}$. This defines the Brownian motion!

Proposition 3.2. *With the above description of* $\mathbf{P}$, *it will be impossible to see that the Brownian motion is almost surely continuous because the continuous functions* $\mathcal{C}\subset\{f:[0,1]\to\mathbb{R}\}$ *are not even measurable.*

Proof. Suppose by contradiction $\mathcal{C}$ is measurable. Then we can find a sequence $t_1,t_2,\ldots$ of times and Borel sets $B_1,B_2,\ldots$ so that $\mathcal{C}=\{f:(f(t_i)\in B_i\}$ (The proof of this fact comes by showing that sets of the form $\{f:(f(t_i)\in B_i\}$ are a sigma-algebra which contain the cylinder sets used to define $\Omega=\sigma(\mathcal{A})$). Take any continuous function f now, and alter its value at a single point $t\notin\{t_1,t_2,\ldots\}$ to get a function $\hat{f}$ which agrees with f at $\{t_1,t_2,\ldots\}$ but is not continuous. But then $\hat{f}\in\mathcal{C}=\{f:(f(t_i)\in B_i\}$ since it agrees with f at $\{t_1,t_2,\ldots\}$ is a contradiction. □

This result means that our construction is not good. It is better to construct the B_t as follows:

Definition 3.3. (Constructing Brownian Motion with Uniform Continuity)

Step 1. (Define on dyadic rationals). Let $\mathbf{P}_{t_1,\ldots t_n}$ as above. Use the countable Kolmogorov Extension Theorem to get a measure $\mathbf{P}$ on the set of functions $\Omega=\{f:[0,1]\cap D_2\to\mathbb{R}\}$ from the dyadic rationals to $\mathbb{R}$.

Step 2. Check that functions in Ω are almost surely Holder continuous. i.e. for almost all $f\in\Omega$, $|f(t)-f(s)|\le C|t-s|^\gamma$

Step 3. Conclude that for almost every $f\in\Omega$,there is a unique way to extend f to a function $f:[0,1]\to\mathbb{R}$ since the dyadic rationals are dense in $\mathbb{R}$.

Step 1 is pretty simple, but step 2 requires some verification and is the real heart of the problem:

Proposition 3.4. *Fix* $\gamma<\frac{1}{2}$. *For almost every* $f\in\Omega$, *there is a constant* C *so that* $|f(t)-f(s)|\le C|t-s|^\gamma$

We first prove a lemma.

Lemma 3.5. *Fix* $\gamma<\frac{1}{2}$. *Then there exists* $\delta>0$,*so that for almost every* $f\in\Omega$, *there is an* $N\in\mathbb{N}$ *(which depends on* f*) so that for* $n\ge N$ *we have:*

$$|f(x)-f(y)|\le|x-y|^\gamma$$

Whenever $x=i2^{-n},y=j2^{-n}$ *and* $|x-y|\le\left(\frac{1}{2}\right)^{n(1-\delta)}$

Proof. Take $m\in\mathbb{N}$ so large so that $m>\frac{1}{1-2\gamma}$. We use the inequality $\mathbf{E}\,|f(t)-f(s)|^{2m}\le C_m|t-s|^m$ with $C_m=\mathbf{E}|f(1)|^{2m}$ (This follows by the property that $f(t)-f(s)\sim f(s)+N(0,t-s)$). For any $n\in\mathbb{N}$ now, consider now the following estimates:

$$\mathbf{P}\left(|f(x)-f(y)| > |x-y|^\gamma \text{ for some } x = i2^{-n},\ y = j2^{-n} \text{ and } |x-y| \le \left(\frac{1}{2}\right)^{n(1-\delta)}\right)$$
$$\le \sum |x-y|^{-2m\gamma}\mathbf{E}\left(|f(x)-f(y)|^{2m}\right)$$

Where the sum on the right hand side is taken over all the possible x, y that satisfy the inequality $|x-y| \le \left(\frac{1}{2}\right)^{n(1-\delta)}$ (There are finitely many, since we are restricting ourselves to dyadic rationals $x = i2^{-n}, y = j2^{-n}$). We have used the Chebyshev inequality $\mathbf{P}(|X| > a) \le a^{-m}\mathbf{E}(|X|^m)$ here. Now, by the above inequality, we have:

$$\begin{aligned} LHS &\le C_m \sum |x-y|^{-2m\gamma}|x-y|^m \\ &= C_m \sum |x-y|^{-2m\gamma+m} \\ &\le C_m 2^n 2^{n\delta}(2^{-n(1-\delta)})^{-2m\gamma+m} \\ &= C_m 2^{-n(-(1+\delta)+(1-\delta)(-2m\gamma+m))} \end{aligned}$$

The last bound comes in because $|x-y| \le 2^{-n(1-\delta)}$ for x, y in our sum, and there are at most 2^n choices for x and $2^{n\delta}$ choices for y once x has been fixed (remember, they are all n-th level dyadic rationals). Now, the term that appears in the exponent is:

$$\epsilon = -(1+\delta) + (1-\delta)(-2m\gamma + m)$$

Since m is so large so that $-2m\gamma + m > 1$, we can choose δ so small so that $\epsilon > 0$. We will have then that

$$LHS \le 2^{-n\epsilon}$$

Which is a summable sequence! By the Borel Cantelli lemma, it must be the case that for almost every $f \in \Omega$ the event here happens only finitely many times. This is exactly the statement of the lemma which we wanted to prove. $\square$

PROPOSITION 3.6. *Fix $\gamma < \frac{1}{2}$. For almost every $f \in \Omega$, there is a constant C so that $|f(t) - f(s)| \le C|t-s|^\gamma$*

PROOF. For almost every $f \in \Omega$, find $\delta > 0$, $N \in \mathbb{N}$ as in the lemma. Take any $t, s \in D_2 \cap [0,1]$ with $t - s < 2^{-N(1-\delta)}$.Choose $m > N$ now so that $2^{-(m+1)(1-\delta)} \le t - s \le 2^{-m(1-\delta)}$.Write now $t = i2^{-m} - 2^{-q_1} - 2^{-q_2} - \ldots 2^{-q_k} < (i-1)2^{-m}$, and $s = j2^{-m} + 2^{-r_1} + \ldots + 2^{-r_l} < (j+1)2^m$ for some choice of $q's$ and $r's$ so that $m < q_1 < \ldots < q_k$ and $m < r_1 < \ldots < r_l$. Since $t - r < 2^{-m(1-\delta)}$, we have $i2^{-m} - j2^{-m} < t - s < 2^{-m(1-\delta)}$ so we can apply the result from the lemma to conclude:

$$\begin{aligned} |f(i2^{-m}) - f(j2^{-m})| &\le ((2^{m\delta})2^{-m})^\gamma \\ &= 2^{-m(1-\delta)\gamma} \end{aligned}$$

Now, we use the result of the lemma again many times to see that (using our clever rewriting of t):

$$\begin{aligned} |f(t)-f(i2^{-m})| &\leq |f(i2^{-m}-2^{-q_1})-f(i2^{-m})|+|f(i2^{-m}-2^{-q_1}-2^{-q2})-f(i2^{-m}-2^{-q_1})|+\ldots+|f(i2^{-m} \\ &\leq |2^{-q_1}|^\gamma+\ldots+|2^{-q_k}|^\gamma \\ &\leq \sum_{j=m+1}^{\infty}(2^{-j})^\gamma \\ &\leq C2^{-\gamma m} \end{aligned}$$

Since $m<q_p$ for each p, and where we used Jensen's inequality to bound the sum. We similarly get a bound on $|f(s)-f(j2^{-m})|$.Finally then:

$$\begin{aligned} |f(t)-f(s)| &\leq C2^{-\gamma m(1-\delta)}+C2^{-\gamma m}+C2^{-\gamma m} \\ &\leq C2^{-\gamma m(1-\delta)} \\ &= C2^{\gamma(1-\delta)}\left(2^{-(m+1)(1-\delta)}\right)^\gamma \\ &\leq C2^{\gamma(1-\delta)}|t-s|^\gamma \end{aligned}$$

By the choice of m so that $2^{-(m+1)(1-\delta)}\leq t-s$. □

So from here we see that the Brownian motion is almost surely Holder continuous for exponents $\gamma<\frac{1}{2}$. This result lets us find a unique extension of $f(t)$ from the dyadic rationals to all of $[0,1]$ which is not only continuous, but moreover its Holder continuous for exponents $\gamma<\frac{1}{2}$, which is a stronger result than our first construction. For ease of notation now, we will change our notation now a little bit. We will refer to $\omega\in\Omega$now instead of f and we now have a family of random variables $B_t(\omega)=\omega(t)$. What we have just proven is that for fixed ω, the map $t\to B_t(\omega)$ is indeed a Holder continuous path for exponents $\gamma<\frac{1}{2}$.

4. Some Properties

The following slick result shows that the Brownian motion is nowhere Holder continuous for $\gamma>\frac{1}{2}$, which in particular shows that it is nowhere differentiable.

Proposition 4.1. *For $\gamma>\frac{1}{2}$, the set of functions which are Holder continuous with exponent γ at some point is a null set. In other words, the Brownian motion is almost surely nowhere Holder continuous for exponents $\gamma>\frac{1}{2}$.*

Proof. Fix a $\gamma>\frac{1}{2}$ and $C\in\mathbb{R}$. Choose $m\in\mathbb{N}$ so large so that γ¿$\frac{m+1}{2m}$. Define the events, starting at $n>m$:

$$A_n=\left\{\omega:\exists s\in[0,1]\text{ such that }|B_t-B_s|\leq C|t-s|^\gamma\forall t\in[s-\frac{m}{n},s+\frac{m}{n}]\right\}$$

Define the random variable:

$$Y_{n,k}(\omega)=\max_{j=0,1,\ldots 2m}\left|B\left(\frac{k+j}{n}\right)-B\left(\frac{k+j-1}{n}\right)\right|$$

And finally, the events:

$$B_n=\left\{\text{at least one of the }Y_{n,k}\leq 2C\left(\frac{m}{n}\right)^\gamma\right\}$$

We now claim that $A_n\subset B_n$, since for $\omega\in A_n$, we find an s so that $|B_t-B_s|\leq C|t-s|^\gamma\forall t\in[s-\frac{m}{n},s+\frac{m}{n}]$. In particular, $|B_t-B_s|\leq C\left(\frac{m}{n}\right)^\gamma$ By the pigeonhole principle, inside this interval we can find k so that $\{\frac{k}{n}, \frac{k+1}{n}, \frac{k+2}{n}, \ldots \frac{k+2m}{n}\} \subset [s - \frac{m}{n}, s + \frac{m}{n}]$. But then, for this k, we have:

$$\begin{aligned} Y_{n,k}(\omega) &= \max_{j=0,1,\ldots 2m} \left| B\left(\frac{k+j}{n}\right) - B\left(\frac{k+j-1}{n}\right) \right| \\ &\leq \max_{j=0\ldots 2m} \left| B\left(\frac{k+j}{n}\right) - B(s) \right| + \left| B(s) - B\left(\frac{k+j-1}{n}\right) \right| \\ &\leq 2C\left(\frac{m}{n}\right)^{\gamma} \end{aligned}$$

So $\omega \in B_n$ by definition.Now consider that:

$$\begin{aligned} \mathbf{P}(A_n) &\leq \mathbf{P}(B_n) \\ &\leq \sum_{k=0..n-m} \mathbf{P}\left(Y_{n,k} \leq 2C\left(\frac{m}{n}\right)^{\gamma}\right) \\ &\leq \sum_{k=0..n-m} \mathbf{P}\left(\left|B_{\frac{k+j}{n}} - B_{\frac{k+j-1}{n}}\right| \leq 2C\left(\frac{m}{n}\right)^{\gamma} \text{ for each } j = 0, 1, ..2m\right) \\ &\leq n\mathbf{P}\left(|B_{\frac{1}{n}} - B_0| < 2C\left(\frac{m}{n}\right)^{\gamma}\right)^{2m} \\ &= n\mathbf{P}\left(|B_1 - B_0| < 2C\left(\frac{m}{n}\right)^{\gamma}\sqrt{n}\right)^{2m} \\ &\leq n\left(\frac{2}{\sqrt{2\pi}} 2C\left(\frac{m}{n}\right)^{\gamma}\sqrt{n}\right)^{2m} \\ &= Dn^{(\frac{1}{2}-\gamma)2m+1} = Dn^{m+1-2m\gamma} \to 0 \end{aligned}$$

Where we used the independence property of disjoint intervals of the Brownian motion, the scaling relation $\mathbf{P}(B_t > a) = \mathbf{P}(B_{ct} > \sqrt{c}a)$, and the easy inequality $\mathbf{P}(N(0,1) > \lambda) \leq 2\lambda$which comes from integrating the p.d.f.. Finally, by the choice of m so that $\gamma > \frac{m+1}{2m}$, we know that $m + 1 - 2m\gamma < 0$ so this probability does indeed go to zero. But then, as the events A_n are increasing, this means that A_n are all zero probability events, which is the result we wanted. □

CHAPTER 5

Appendix

1. Conditional Random Variables

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ random variables. $\mathcal{B}$ is the Borel sigma algebra of $\mathbb{R}$.

Definition 1.1. We define $\sigma(X) \subset \mathcal{F}$ to be the sigma-algebra generated by the preimages of Borel sets through $\mathcal{F}$. That is:

$$\sigma(X) = \sigma(\{X^{-1}(B) : B \in \mathcal{B}\})$$

Remark. The sub-algebra $\sigma(X)$ is in coarser than all of $\mathcal{F}$. Intuitively, the random variable X can only "detect" up to sets in $\sigma(X)$.

Definition 1.2. Let $\Sigma \subset \mathcal{F}$ be a subalgebra of $\mathcal{F}$. We say a random variable $X : \Omega \to \mathbb{R}$ is $\Sigma-$**measurable** if $X^{-1}(B) \in \Sigma$ for all $B \in \mathcal{B}$. Equivalently, if $\sigma(X) \subset \Sigma$.

Example 1.3. Every random variable is always $\mathcal{F}$ measurable, since $\sigma(X) \subset \mathcal{F}$.

Definition 1.4. Given X and Y, we can define a new random variable $Z = \mathbf{E}(Y|X)$ to be the unique random variable with the following two properties:
1. Z is $\sigma(X)$ measurable.
2. For any $B \in \mathcal{B}$ we have $\mathbf{E}\left(Z\mathbf{1}_{X\in B}\right) = \mathbf{E}\left(Y\mathbf{1}_{X\in B}\right)$

Remark. The existence of this random variable is proven by restricting the Radon-Nikodym derivative of Y with respect to the probability space to just the sigma field $\sigma(X)$.

Remark. There is no problem with picking any subalgebra $\Sigma \subset \mathcal{F}$ instead of $\sigma(X)$. The second condition is simply that for any $S \in \Sigma$ we have $\mathbf{E}\left(Z\mathbf{1}_S\right) = \mathbf{E}\left(Y\mathbf{1}_S\right)$, which is really the condition above with $\Sigma = \sigma(X)$.

Remark. $Z = \mathbf{E}(Y|X)$ is a random variable $Z : \Omega \to \mathbb{R}$, but it is often thought of as a function $Z : \mathbb{R} \to \mathbb{R}$, whose input is the random variable X. This works because Z is $\sigma(X)$ measurable. The following two little results clear this up a bit:

Proposition 1.5. *If $f : \mathbb{R} \to \mathbb{R}$ is measurable, and $Z : \Omega \to \mathbb{R}$is Σ-measurable, then the random variable $f \circ Z$ is Σ-measurable too.*

Proof. For any $B \in \mathcal{B}$ we have $(f \circ Z)^{-1}(B) = Z^{-1}(f^{-1}(B)) \in \Sigma$since $f^{-1}(B) \in \mathcal{B}$ as f is measurable and Z is Σ- measurable. □

Proposition 1.6. *If Z is $\sigma(X)$-measurable random variable, then we may think of Z as a function $Z : \mathbb{R} \to \mathbb{R}$ whose input is X.*

Proof. Define $\tilde{Z} : \mathbb{R} \to \mathbb{R}$ by $\tilde{Z}(x) = Z(\omega)$ for any representative $\omega \in X^{-1}(\{x\})$. We must justify why this value is independent of the choice of $\omega \in X^{-1}(\{x\})$. Indeed for $\omega_1, \omega_2 \in X^{-1}(\{x\})$, let $z = Z(\omega_1)$.Since Z is $\sigma(X)$ measurable, we have that:

$$\begin{aligned} Z^{-1}(\{z\}) &\in \sigma(X) \\ \Rightarrow Z^{-1}(\{z\}) &= X^{-1}(B) \text{ for some } B \in \mathcal{B} \end{aligned}$$

But then $\omega_1 \in Z^{-1}(\{z\}) = X^{-1}(B)$, so that $X(\omega_1) \in B$. Since $X(\omega_1) = X(\omega_2) = x$, we have then $\omega_2 \in X^{-1}(B) = Z^{-1}(\{z\})$, which means that $Z(\omega_1) = Z(\omega_2) = z$, as desired. Hence $\tilde{Z}$ is well defined! With this definition of $\tilde{Z}$, we see that $Z = \tilde{Z} \circ X$. We often conflate Z with $\tilde{Z}$ in practice. □

2. Extension Theorems

Theorem 2.1. *[Caratheodory Extension Theorem]*

Fix some $(\Omega, \mathcal{A}, \mathbf{P}_0)$, where Ω is a set, $\mathcal{A}$ is an algebra of sets (aka a field of sets), and $\mathbf{P}_0$is a finitely additive probability measure on $\mathcal{A}$. If we have the additional property that:

For sequences of sets $A_1, A_2, \ldots \in \mathcal{A}$ which are pairwise disjoint with the property that $\cup A_n \in \mathcal{A}$ too, then we necessarily have $\mathbf{P}_0(\cup A_n) = \sum \mathbf{P}_0(A_n)$.

Then there is a unique extension to a probability space $(\Omega, \sigma(\mathcal{A}), \mathbf{P})$ so that $\mathbf{P}$ and $\mathbf{P}_0$ agree on $\mathcal{A}$.

Proof. [sketch] The idea is exactly the same as the construction of the Lebesgue measure on $[0,1]$ from the premeasure generated by $\mu((a,b)) = b - a$ on the algebra of open sets. Define an outer measure: $\mathbf{P}(E) := \inf_{E \subset \cup A_n} \sum \mathbf{P}_0(A_n)$. From here you check that $\mathbf{P}$ is indeed a probability measure. Countable subadditivity and monotonicity are easy. To get that $\mathbf{P}(A) = \mathbf{P}_0(A)$ for $A \in \mathcal{A}$ requires the special property we are given above. Once this is done, you can define measurable sets a-la Caratheodory: E measurable iff for all $A \in \mathcal{A}$ we have $\mathbf{P}(A) = \mathbf{P}(A \cap E) + \mathbf{P}(A \cap E^c)$. Then you verify that $\sigma(\mathcal{A})$ is a subset of these measurable sets, and declare $\mathbf{P} = \mathbf{P}$to be the measure on $\sigma(\mathcal{A})$. □

Remark. The above condition needed in the theorem can be replaced with "Continuity from above at $\emptyset$":

For $A_1, A_2, \ldots \in \mathcal{A}$ which are decreasing down to $\emptyset$, then we necessarily have that $\mathbf{P}_0(A_n) \to 0$ too.

The equivalence of these two conditions is not too difficult. The first condition is more intuitive, while this second condition is sometimes easier to verify in practice.

Theorem 2.2. *[Countable Kolmogorov Extension Theorem]*

Suppose for every $n \geq 1$, we have a probability measure $\mathbf{P}_n$ on $\mathbb{R}^n$. Suppose also that these probability measure's satisfy the following consistency condition for every Borel set $E \in \mathbb{R}^n$:

$$\mathbf{P}_{n+k}(E \times \mathbb{R}^k) = \mathbf{P}_n(E)$$

Then there exists a unique measure $\mathbf{P}$on the infinite product measure $\mathbb{R}^\infty$of sequences, so that for every Borel set $E \in \mathbb{R}^n$ $\mathbf{P}(E \times \mathbb{R} \times \mathbb{R} \times \ldots) = \mathbf{P}_n(E)$.

PROOF. [sketch] Take $\Omega = \mathbb{R}^{\infty}$be real-valued sequences. Define the field of cylinder sets to be:

$$\mathcal{A} = \{E \times \mathbb{R} \times \mathbb{R} \times \ldots : E \in \mathbb{R}^n \text{ is Borel}\}$$

With finitely additive measure $\mathbf{P}_0(E \times \mathbb{R} \times \mathbb{R} \times \ldots) := \mathbf{P}_n(E)$. The given condition on the $\mathbf{P}'_n s$ shows this is well defined. To see continuity from above at $\emptyset$, notice that if $A_k \downarrow \emptyset$, then we must have $A_k = E_k \times \mathbb{R} \times \mathbb{R} \times \ldots$ for some sets $E_k \in \mathbb{R}^n$ with $E_k \downarrow \emptyset$. But then of course, since $\mathbf{P}_n$ is a probability measure, we have $\mathbf{P}_0(A_k) = \mathbf{P}_n(E_k) \to 0$. By application of the Caratheodory extension theorem, we get the desired measure! □

THEOREM 2.3. *[Kolmogorov Extension Theorem]*
Let T be any interval $T \subset \mathbb{R}$. Suppose we have a family of probability measure's $\mathbf{P}_{t_1,t_2,\ldots t_n}$ on $\mathbb{R}^n$ whenever $t_1, t_2, \ldots t_n$ is a finite number of points in T. Suppose also that these probability measure's satisfy the following consistency condition:

$$\mathbf{P}_{t_1,t_2,\ldots t_n,\hat{t}_1,\hat{t}_2,\ldots \hat{t}_m}(E \times \mathbb{R}^m) = \mathbf{P}_{t_1,t_2,\ldots t_n}(E)$$

Then there exists a unique measure $\mathbf{P}$ on the set of functions $\{f : T \to \mathbb{R}\}$so that:

$$\mathbf{P}\left(\{f : (f(t_1), f(t_2), \ldots f(t_n)) \in E\}\right) = \mathbf{P}_{t_1,t_2,\ldots t_n}(E)$$

REMARK. This is very similar to the countable version, but requires some more work to make it work out. However, since the space of functions $\{f : T \to \mathbb{R}\}$ is so large, this theorem often gives us a very unwieldy space to work with, one in which we can't get our hands on the properties we want. The construction of Brownian motion below is a great example, constructing with the uncountable Kolmogorov theorem is bad, while with the countable one is a good.