LECTURE NOTES ON LIMIT THEOREMS IN PROBABILITY

DAVID STEINSALTZ

1. Review: Laws of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. (independent, identically distributed) random variables with expectation μ.

Theorem 1.1 (Weak Law of Large Numbers (WLLN)).

$$\lim_{n\to\infty} \frac{1}{n}\left(X_1 + \cdots + X_n\right) = \mu \text{ in probability.}$$

That is, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{1}{n}\left(X_1 + \cdots + X_n\right) - \mu \right| > \epsilon \right\} = 0.$$

Theorem 1.2 (Strong Law of Large Numbers (SLLN)).

$$\lim_{n\to\infty} \frac{1}{n}\left(X_1 + \cdots + X_n\right) = \mu \text{ almost surely.}$$

That is,

$$\mathbb{P}\left\{ \lim_{n\to\infty} \frac{1}{n}\left(X_1 + \cdots + X_n\right) = \mu \right\} = 1.$$

Elementary treatments usually assume that the random variables have a finite variance, or even (for the SLLN) a finite fourth moment. This is not required. Is independence required? Clearly we can't dispense with it entirely: If the random variables are all the same, there can't be convergence to a deterministic limit.

One elementary proof of the WLLN goes as follows: Suppose $X_1, X_2, \ldots$ are random variables with expectation μ and variance bounded by S. Let $\bar{X}_n := n^{-1}\left(X_1 + \cdots + X_n\right)$. Then

$$\begin{aligned}
\operatorname{Var}(\bar{X}_n) &= n^{-2}\left(\sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{1\le i<j\le n} \operatorname{Cov}\left(X_i, X_j\right) \right) \\
&\le \frac{S}{n} + \frac{2}{n^2} \sum_{1\le i<j\le n} \operatorname{Cov}\left(X_i, X_j\right).
\end{aligned}$$

When the random variables are independent, the covariances are all 0, so the variance is bounded by S/n, and by Chebyshev's Inequality

$$\mathbb{P}\left\{\left|\bar{X}-\mu\right| \geq \epsilon\right\} \leq \frac{S}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

for any $\epsilon > 0$. This is also true, of course, if the random variables are merely uncorrelated. But the covariances don't have to be exactly 0. All this proof requires is that the variance of $\bar{X}_n$ go to 0 as $n \to \infty$. Suppose the random variables approach being uncorrelated as they get further apart in the sequence. That is, suppose there are constants a_k with $\lim_{k\to\infty} a_k = 0$ such that

$$\text{Cov}\left(X_i, X_j\right) \leq a_{|i-j|}.$$

Then

$$\begin{aligned}\text{Var}(\bar{X}_n) &\leq \frac{2}{n^2}\sum_{k=0}^{n-1}(n-k)a_k \\ &\leq \frac{2}{n}\sum_{k=1}^{n} a_k \\ &\xrightarrow{n\to\infty} 0.\end{aligned}$$

The WLLN also has another strengthened form: The Central Limit Theorem (CLT). Dividing the partial sums by n is too much: It crushes out all the randomness. The CLT says that the randomness is actually on the order of $\sqrt{n}$, and that it converges to a Gaussian distribution:

Theorem 1.3.

$$n^{-1/2}\left(X_1 + \cdots + X_n - \mu n\right) \xrightarrow{n\to\infty}_d \mathcal{N}(0, \sigma^2).$$

I have not put explicit conditions on these theorems, because there are a lot of options for how to formulate the conditions. There are three kinds of assumptions that go into these limit theorems: moment conditions, independence conditions, and identical distribution conditions. Since in real applications these conditions might be hard to check, we'd like to know what the minimum required is. The brief answers are:

Moments These can be reduced to the obvious minimum required for the results to make sense. The laws of large numbers are about convergence to a mean, so they require that the random variables have an expectation; in fact, that is all that is required. The CLT is about convergence of distributions to a normal distribution with the same variance, so the X_i must have finite variance. This suffices, though there is a certain tradeoff between moments and identical distribution: If we allow the X_i to have different distributions, we need to at least assume that $\mathbb{E}[|X_i|^{2+\delta}]$ is universally bounded for some $\delta > 0$.

Independence If the random variables are identically distributed in a strong sense, called *stationary*, then we can reduce the independence requirement to a very weak form, called *ergodicity*, meaning that there is no **infinitely long-term** dependence. A description of this **ergodic theory** will be the main topic of these lectures.

Identical distribution The proof of the WLLN doesn't use anything about the distributions except their mean and a bound on their variance. It turns out that we can dispense with identical distribution for the SLLN and the CLT as well, as long as we strengthen the independence to a sort of average independence (still weaker than assuming actual independence) called the *martingale property.* We will briefly discuss this at the end of the second lecture.

2. Ergodic theory

2.1. **Some definitions.** These definitions may be found in [6, 1]. Since we are looking at infinite sequences of random variables, a rigorous treatment requires measure-theoretic probability, which we will not use. Instead, we will make some definitions whose consistency will have to be accepted, and proceed from there. An accessible introduction to these concepts may be found in [5].

Let $X_1, X_2, \ldots$ be a one-sided infinite discrete-time stochastic process taking values in a state space $\mathcal{X}$. We assume the process is *stationary*, meaning that the distribution of $(X_1, X_2, \ldots)$ is the same as the distribution of $(X_n, X_{n+1}, \ldots)$ for any n. It will be convenient to think of the sequence as being two-sided: $(\ldots, X_{-1}, X_0, X_1, X_2, \ldots)$.

Theorem 2.1. *Any one-sided stationary sequence may be embedded in a two-sided stationary sequence.*

Proof. Clearly we may define a stationary sequence $(X_{-n}, X_{-n+1}, \ldots)$ for any n. The Kolmogorov Extension Theorem allows this to be extended to an infinite sequence. □

Our sample space is $\Omega = \mathcal{X}^{\mathbb{Z}}$. For $m \le n$ let $\mathcal{B}_m^n$ be the set of bounded $(X_i)_m^n$*-measurable* random variables, by which we mean bounded random variables that can be written as a function of the sequence $(X_m, X_{m+1}, \ldots, X_n)$; we write $\mathcal{B}_m$ for $\mathcal{B}_{m,\infty}$ and $\mathcal{B}^n$ for $\mathcal{B}_{-\infty,n}$. We will also say that an event B is in $\mathcal{B}_m^n$ if its indicator is. We define the shift operator $\tau : \Omega \to \Omega$ by shifting every coordinate one to the left. That is, if $X := (\ldots, X_{-1}, X_0, X_1, \ldots)$ then

$$(\tau X)_i = X_{i+1}.$$

If $Y = f(X_m, \ldots, X_n) \in \mathcal{B}_m^n$ then $Y \circ \tau^k = f(X_{m+k}, \ldots, X_{n+k}) \in \mathcal{B}_{m+k}^{n+k}$. Note that if $B \subset \Omega$ is any event then

$$\mathbf{1}_{\tau^{-k}B} = \tau^k \mathbf{1}_B.$$

We say the stochastic process is *ergodic*

An important interpretation of *ergodic* is that an ergodic process can't be split into a mixture of processes with different distributions.

Theorem 2.2. *The following are equivalent:*

(i) *If A and B are any events with nonzero probability then* $\mathbb{P}(A \cap \tau^{-n}B) > 0$ *for some n;*
(ii) *Any shift-invariant event — that is, an event $A \subset \Omega$ such that* $\mathbb{P}\left(A \triangle \tau^{-1}A\right) = 0$ *— has probability 0 or 1;*
(iii) *Any shift-invariant function of $(X_0, X_1, \dots)$ — that is, a function f such that $f(X_0, X_1, \dots) = f(X_1, X_2, \dots)$ almost surely — is almost surely constant;*
(iv) *There is no way to represent (X_i) by defining two different stationary stochastic processes (Y_i) and (Z_i) with distinct distributions, and an independent Bernoulli(p) random variable ξ (with $0 < p < 1$), such that for all $i = 1, 2, \dots$*

$$X_i = \begin{cases} Y_i & \text{if } \xi = 0, \\ Z_i & \text{if } \xi = 1. \end{cases}$$

Proof. (i)$\Longrightarrow$(iii) Let f be a shift-invariant function. For any x let $A_x := \{f(X_0, X_1, \dots) > x\}$. By shift invariance

$$\tau^{-n}A_x = \left\{X \,:\, f(X_n, X_{n+1}, \dots) > x\right\},$$

so

$$\tau^{-n}A_x \cap A_x^{\mathsf{c}} \subset \left\{f(X_0, X_1, \dots) \neq f(X_n, X_{n+1}, \dots)\right\},$$

which has probability 0 for all n. It follows that A or A^{c} must have probability 0. Since this is true for any x, it follows that f is almost-surely constant.

(iii)$\Longrightarrow$(ii)

(ii)$\Longrightarrow$(i) Consider any sets A and B with nonzero probability such that $\mathbb{P}(A \cap \tau^{-n}B) = 0$ for all n. Then

$$C := \bigcup_{n=0}^{\infty} \tau^{-n}B$$

must have probability strictly between 0 and 1. Also, $\tau^{-1}C \subset C$. Since $\mathbb{P}(\tau^{-1}C) = \mathbb{P}(C)$ (since the process is stationary) it follows that C is invariant. Since $\mathbb{P}(C) \geq \mathbb{P}(A) > 0$, it must be that $\mathbb{P}(C) = 1$. But then

$$0 < \mathbb{P}\left(A \cap C\right) = \mathbb{P}\left(\bigcup_{n=0}^{\infty} A \cap \tau^{-n}B\right).$$

The equivalence to (iv) is left as an exercise. □

We say the process is ergodic if it satisfies these equivalent conditions.

2.2. **Mixing conditions.** The stochastic process is *mixing* (or *strongly mixing*) if for any events A, B

$$\lim_{n\to\infty} \mathbb{P}\left(A \cap \tau^{-n} B\right) = \mathbb{P}(A)\mathbb{P}(B). \tag{1}$$

Finally, the stochastic process is ϕ*-mixing* if for any k, and any

$$\phi(n) := \sup_{A\in\mathcal{B}_{-\infty,m},\, B\in\mathcal{B}_{m+n,\infty}} \left|\mathbb{P}\left(B \mid A\right) - \mathbb{P}\left(B\right)\right| \xrightarrow{n\to\infty} 0. \tag{2}$$

Suppose the process is mixing, and let A be any shift-invariant event. Then

$$\mathbb{P}(A) = \mathbb{P}\left(A \cap \tau^{-n} A\right),$$

so, taking the limit as $n \to \infty$, we have $\mathbb{P}(A) = \mathbb{P}(A)^2$, implying $\mathbb{P}(A)$ is 0 or 1. Thus

$$\phi\text{–mixing} \implies \text{mixing} \implies \text{ergodic}\ .$$

The main result about ergodic processes is that they satisfy the equivalent of the Strong Law of Large Numbers, Birkhoff's Ergodic Theorem, stated in section 2.4. It follows that ergodicity is equivalent to

$$\lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \mathbb{P}\left(A \cap \tau^{-n} B\right) = \mathbb{P}(A)\mathbb{P}(B). \tag{3}$$

for any events A, B.

The definition in terms of the map τ may be hard to understand at first glance. The definition of mixing is essentially a generalisation of the statement that X_0 and X_n are asymptotically independent; that is, for any $A, B \subset \mathcal{X}$

$$\lim_{n\to\infty} \mathbb{P}\left\{X_0 \in A,\, X_n \in B\right\} = \mathbb{P}\left\{X_0 \in A\right\} \cdot \mathbb{P}\left\{X_n \in B\right\}.$$

A thorough discussion of various mixing conditions may be found at `https://www.encyclopediaofmath.org/index.php/Strong_mixing_conditions`

2.3. **Examples.**

2.3.1. *i.i.d. sequences.* Any i.i.d. sequence is mixing, hence ergodic. An immediate consequence is the Kolmogorov zero–one law: We define a *tail event* for an i.i.d. sequence to be an event whose occurrence may be inferred from $X_n, X_{n+1}, \ldots$ for n arbitrarily large. The zero–one law says that all tail events have probability 0 or 1. For example, the event $\{n^{-1}(X_1 + \cdots + X_n)$ converges to a limit$\}$ is a tail event, so must have probability 0 or 1. (The SLLN states that the probability is 1.)

2.3.2. *Rotations of the circle.* Fix $\alpha \in (0,1)$. Let X_0 be uniform on $[0,1)$, and let $X_n = X_0 + \alpha n (\text{mod } 1)$. This sequence is ergodic iff α is irrational. Suppose α is irrational, and let f be a shift-invariant function, which may be taken to be $f(X_0)$. We know that there is a unique sequence of Fourier coefficients

$$a_k = \int_0^1 e^{-2\pi i k x} f(x)\,dx$$

such that

$$\sum_{k=-K}^{K} a_k e^{2\pi i k x} \xrightarrow{K\to\infty} f(x).$$

Since $f \circ \tau(x) = f(x+\alpha)$, we have

$$\sum_{k=-K}^{K} \left(a_k e^{2\pi i k \alpha} \right) e^{2\pi i k x} \xrightarrow{K\to\infty} f(x+\alpha).$$

Since f is shift-invariant this must be the same as $f(x)$, so we have an alternative Fourier series $a_k e^{2\pi i k \alpha}$ for the same function. Since the Fourier transform is unique, we must have for all k

$$a_k e^{2\pi i k \alpha} = a_k.$$

But this happens only if $k\alpha$ is an integer or $a_k = 0$. Thus $a_k = 0$ for all $k \neq 0$, implying that f is constant.

We leave it as an exercise to show the sequence is not ergodic when α is rational, and that it is never mixing.

2.3.3. *Markov chains.* Let $\mathcal{X}$ be finite, and let $X_0, X_1, X_2, \dots$ be a Markov chain with transition probabilities $P(i,j)$, and X_0 in a stationary distribution π_i satisfying $\pi_j = \sum_{i \in \mathcal{X}} \pi_i P(i,j)$. Then (X_i) is a stationary sequence. It is ergodic if and only if it is irreducible and aperiodic. If $\mathcal{X}$ is countably infinite it must be recurrent.

2.4. **The Birkhoff Pointwise Ergodic Theorem.**

Theorem 2.3 (Birkhoff pointwise ergodic theorem)**.** *Let $X_0, X_1, \dots$ be an ergodic stationary sequence, and $f : \mathcal{X} \to \mathbb{R}$ any function such that $\mu := \mathbb{E}[f(X_0)]$ is defined. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = \mu \text{ almost surely.} \tag{4}$$

Proof. This proof comes from [4]. Let

$$\begin{aligned} S_n &:= n^{-1} \sum_{i=0}^{n-1} f(X_i), \\ S_N^* &:= \max_{1 \le n \le N} S_n, \\ S^* &:= \sup_N S_N^*, \\ \bar{S} &:= \limsup_{n\to\infty} S_n, \qquad \underline{\mathrm{S}} := \liminf_{n\to\infty} S_n. \end{aligned}$$

Since $\bar{S}$ and $\underline{\mathrm{S}}$ are invariant, they must be constant. Since $\bar{S} \ge \mathbb{E}[X_0] \ge \underline{\mathrm{S}}$, all we need to show is that $\bar{S} = \underline{\mathrm{S}}$.

Suppose f is bounded, and fix a natural number N and some $\epsilon > 0$. Let E_N be the event $\{S_N^* > \bar{S} - \epsilon\}$. Note that if $(X_i) \notin E_N$ then in particular $f(X_0) \le \lambda$, so we always have

$$\big(f(X_0) - \lambda\big)\,\mathbf{1}_{E_N} \ge f(X_0) - \lambda.$$

Note as well that $\mathbf{1}_{E_N}$ are nondecreasing in N, with

$$\lim_{N\to\infty} \mathbf{1}_{E_N} = \mathbf{1}_{S^* \ge \bar{S} - \epsilon} = 1.$$

Since $\mathbb{P}(E_N) > 0$, there is almost surely some k such that $\tau^k(X) = (X_k, X_{k+1}, \dots) \in E_N$. Following that, there is some $n \le N$ such that $\sum_{i=k}^{k+n-1} f(X_i) \ge n\lambda$. Thus, if we look at the sum

$$\sum_{i=0}^{m-1} \big(f(X_i) - \lambda\big)\,\mathbf{1}_{E_N}(X_i, X_{i+1}, \dots)$$

for some large m, it breaks up into sequences of summands that are 0 alternating with sums of $\le N$ terms that total to something positive. The sum can only be negative if it ends in the middle of one of the latter blocks, and then the smallest it can be is $-N(|f|_\infty + \lambda^+)$. Computing the expectation, and using stationarity, we have for all m

$$m\mathbb{E}\left[\big(f(X_0) - \bar{S} + \epsilon\big)\,\mathbf{1}_{E_N}\right] \ge -N\left(|f|_\infty + \bar{S}\right).$$

Dividing by m and sending $m \to \infty$ we then have

$$\mathbb{E}\left[\big(f(X_0) - \bar{S} + \epsilon\big)\,\mathbf{1}_{E_N}\right] \ge 0.$$

We can then send $N \to \infty$ and conclude[1] that

$$\mathbb{E}\left[f(X_0)\right] \ge \bar{S} - \epsilon.$$

[1] The fact that the limit of the expectation is the expectation of the limit in this case is an example of the Dominated Convergence Theorem, which may be found in any text on measure-theoretic probability.

Since this is true for any $\epsilon > 0$, it follows that $\mathbb{E}\left[f(X_0)\right] \geq \bar{S}$. Applying the same result to the function $-f$, since $\limsup(-S_n) = -\liminf S_n$, we have

$$-\mathbb{E}\left[f(X_0)\right] \geq -\underline{S}.$$

Thus

$$\bar{S} \geq \underline{S} \geq \mathbb{E}\left[f(X_0)\right] \geq \bar{S},$$

so $\bar{S} = \underline{S}$.

We have now completed the proof for f bounded. We extend to general f by the usual expedient of truncating at some constant K, which we then let go to ∞. □

2.5. **Recurrence.** If $X = (X_0, X_1, \dots)$ is a stationary ergodic process, then it makes sense to suppose that any state that it has nonzero probability of being in will be returned to infinitely often. This is the content of the Poincaré Recurrence Theorem. The Kac Recurrence Theorem tells us, in addition, that the return will happen at a frequency proportional to the probability.

Theorem 2.4 (Poincaré Recurrence Theorem)**.** *Let $X = (X_0, X_1, \dots)$ be a stationary process. Let $A \subset \Omega$ be any event, then almost every point of A is recurrent. That is,*

$$\mathbb{P}\left\{X \in A,\, \{n : \tau^n X \in A\} \text{ is finite}\right\} = 0.$$

If the process is ergodic and $\mathbb{P}(A) > 0$ then almost every point is recurrent to A. That is,

$$\mathbb{P}\left\{\{n : \tau^n X \in A\} \text{ is finite}\right\} = 0.$$

The event A may involve the entire future of the process. In the special case where $A = \{\omega : \omega_0 \in E\}$ for some $E \subset \mathcal{X}$, it states that $X_n \in E$ infinitely often with probability 1.

Definition 2.4.1. *For an event $A \subset \Omega$ we define the first recurrence time*

$$R_A(\omega) := \min\{n > 0 : \tau^n \omega \in A\}$$

The Poincaré Recurrence Theorem tells us that $R_A(\omega) < \infty$ for almost every $\omega \in A$, and for almost every $\omega \in \Omega$ when the process is ergodic.

Theorem 2.5 (Kac Recurrence Theorem)**.** *If (X_n) is ergodic, for any A with $\mathbb{P}(A) > 0$,*

$$\mathbb{E}\left[R_A(X) \;\middle|\; X \in A\right] = \frac{1}{\mathbb{P}(A)}.$$

When $X_0, X_1, \dots$ is a stationary ergodic Markov chain on a finite state space $\mathcal{X}$, we may take $A = \{\omega : \omega_0 = i\}$ for any $i \in \mathcal{X}$. Then $R_A(X) = \min\{n \geq 1 : X_n = i\}$, and the Kac Recurrence Theorem tells us that the expected time to return to i for a process started at i is $1/\pi_i$.

Consider the simple random walk $S_n = X_1 + \cdots + X_n$ with $X_i = \pm 1$ with probability $\frac{1}{2}$. The SLLN tells us that $S_n/n \to 0$, and the CLT tells us that S_n is approximately normal with mean 0 and variance n. But we also know that S_n is recurrent; that is, S_n returns to 0 infinitely often. We may also want to know how often a random walk returns to 0.

2.6. **The subadditive ergodic theorem.**

Theorem 2.6 (Subadditive ergodic theorem)**.** *Suppose* $(X_{m,n})_{m=0}^{n-1}$, $n = 1, 2, \ldots$ *is a triangular array of random variables satisfying*

(i) $X_{0,m} + X_{m,n} \geq X_{0,n}$;
(ii) *For each* $k \geq 1$, *the sequence* $Y_n := X_{nk,nk+k}$ *for* $n = 1, 2, dots$ *is stationary and ergodic;*
(iii) *The distribution of* $\left(X_{m,m+k}\right)_{m=1}^{\infty}$ *does not depend on* m*;*
(iv) $\mathbb{E}\left[X_{0,1}^{+}\right] < \infty$ *and* $\mathbb{E}\left[X_{0,n}\right] \geq -\gamma_0 n$ *for some finite* γ_0.

Then

$$\lim_{n\to\infty} \frac{X_{0,n}}{n}$$

exists almost surely, and is almost surely equal to

$$\gamma := \inf_m \frac{\mathbb{E}[X_{0,m}]}{m}.$$

Note: This theorem is originally due to Kingman. Following [1] we give the stronger version — that is, with weaker conditions — due to Liggett. In the following diagram, the coloured variables are the starts of stationary sequences described in condition (ii):

$$\begin{array}{cccccc}
X_{0,1} & & & & & \\
X_{0,2} & X_{1,2} & & & & \\
X_{0,3} & X_{1,3} & X_{2,3} & & & \\
X_{0,4} & X_{1,4} & X_{2,4} & X_{3,4} & & \\
X_{0,5} & X_{1,5} & X_{2,5} & X_{3,5} & X_{4,5} & \\
X_{0,6} & X_{1,6} & X_{2,6} & X_{3,6} & X_{4,6} & X_{5,6} \\
\vdots & & & & & \vdots \\
X_{0,n} & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}.$$

The columns all have the same distribution, as described in condition (iii), but are not necessarily stationary.

To understand why the definitions are this way, consider the case where $\xi_1, \xi_2, \ldots$ is a stationary ergodic sequence with finite expectation, and $X_{m,n} := \xi_{m+1} + \cdots + \xi_n$. Then the sequence is additive, rather than merely subadditive as required for (i). The sequences defined in (ii) are stationary because they involve non-overlapping successive partial sums from the stationary sequences.

Example: Range of a random walk: Let $\xi_1, \xi_2, \ldots$ be a stationary sequence, and $S_n := \xi_1 + \cdots + \xi_n$. Let $X_{m,n} = \left|\{S_{m+1}, S_{m+2}, \ldots, S_n\}\right|$ be the number of points ever hit by the random walk. This satisfies all the conditions, so we may infer that the range increases linearly at rate γ asymptotically, though this does not immediately tell us anything about how to compute γ.

Products of random matrices: Suppose $M_1, M_2, \dots$ are a stationary sequence of $k \times k$ random positive matrices satisfying

$$\mathbb{E}\left[\|M_i\|\right] < \infty.$$

Let

$$X_{m,n} := \log \|M_n \cdot \dots \cdot M_{m+1}\|$$

This satisfies the subadditivity condition $X_{0,m} + X_{m,n} \geq X_{0,n}$. We may conclude that

$$\lambda := \lim_{n\to\infty} n^{-1} \log \|M_n \cdot \dots \cdot M_1\|$$

exists and is constant almost surely.

In fact, this result holds in significantly greater generality: The matrices do not need to be positive, and the Multiplicative Ergodic Theorem tells us that there are k deterministic Lyapunov exponents $\lambda = \lambda_k \geq lambda_{k-1} \geq \dots \geq \lambda_1$ such that there is a (random) decomposition of $\mathbb{R}^k$ into $j-dimensional$ subspaces E_j such that $E_j \setminus E_{j-1}$ comprises points x such that

$$\lim_{n\to\infty} n^{-1} \log \|M_n \cdots M_1 x\| = \lambda_j.$$

These are the stochastic analogue of the decomposition of $\mathbb{R}^k$ into eigenspaces of some fixed matrix M.

2.7. **Central Limit Theorem.** Let $X_0, X_1, \dots$ be a stationary sequence and f a function from $\mathcal{X}$ to $\mathbb{R}$. Let $S_n = X_0 + \dots + X_{n-1}$ and $\mu = \mathbb{E}[X_i]$. The variance of S_n is given by

$$\sigma_n^2 := \mathrm{Var}(S_n) = n \,\mathrm{Var}\,(X_0) + 2\sum_{i=1}^{n-1} (n-i)\,\mathrm{Cov}\,(X_0, X_i)\,.$$

The following theorem is due to Ibragimov [3]:

Theorem 2.7. *Suppose the stationary process $X_0, X_1, \dots$ is ϕ-mixing, $\lim_{n\to\infty} \sigma_n = \infty$, and for some $\delta > 0$*

$$\mathbb{E}\left[|X_i|^{2+\delta}\right] < \infty.$$

Then $S_n/\sigma_n \xrightarrow{n\to\infty}_d \mathcal{N}(0,1)$.

3. Martingales

An excellent source for all of what is presented here (and much more) is [7].

3.1. **Definitions.** A *martingale* is a mathematical formalisation of the notion of a "fair bet": A process that on average stays constant, regardless of any information about the past behaviour.

A stochastic process $X_0, X_1, X_2, \dots$ is a *martingale difference sequence* if the expectations are finite and for any $m < n$

$$\mathbb{E}\left[X_n \;\middle|\; X_0, \dots, X_m\right] = 0. \tag{5}$$

The partial-sum sequence $M_n := X_0 + \cdots + X_n$ is then called a *martingale*. If the condition '$= 0$' is replaced by '≥ 0' then these are called *submartingale differences* and *submartingale* respectively. If replaced by '≤ 0' then they are *supermartingale differences* and *supermartingale*. Note that a martingale is also a submartingale and a supermartingale.

Examples:

(i) **Symmetric random walk**. Given $X_1, X_2, \dots$ independent with $X_i = \pm 1$ with probability $1/2$, this is a martingale difference sequence, and $S_n = X_1 + \cdots + X_n$ is a martingale.

(ii) **Exponential martingale**. Given $X_1, X_2, \dots$ i.i.d. with moment generating function $e^{m(t)}$ — that is, $\log \mathbb{E}[e^{tX_i}] = m(t)$ — for any t, $M_n := \exp\{tS_n - nm(t)\}$ is a martingale.

(iii) **Polya's Urn**. We have an urn that starts with R_0 red and B_0 blue balls. We pick a ball at random, and then replace it in the urn with k balls of the same colour. Let X_n be the fraction of red balls after the k-th draw. Then $X_0, X_1, \dots$ is a martingale.

Definition 3.0.1. *A* stopping time *is a random time T such that at time T it can be determined that T has occurred based on $X_0, X_1, \dots, X_T$.*

Definition 3.0.2. *A stochastic process $X_0, X_1, \dots$ is* uniformly integrable *if*

$$\lim_{K \to \infty} \sup_i \mathbb{E}\left[|X_i| \mathbf{1}_{\{|X_i| > K\}}\right] = 0.$$

Examples: The first entry time $T = \min\{n \geq 0 : X_n \in E\}$ into a set $E \subset \mathcal{X}$ is a stopping time. The last exit time $T = \max\{n \geq 0 : X_n \in E\}$ is not.

The two key results about martingales are the **Optional Stopping Theorem** and the **Martingale Convergence Theorem**. The Optional Stopping Theorem tells us that a martingale stopped at a random stopping time is still a martingale; in other words, when playing a succession of fair games there is no way to make money on average by choosing some gambling system that tells you when to stop, unless you can look into the future.

Lemma 3.0.1. *Let X_n be a submartingale and T a stopping time. Suppose that one of these conditions holds:*

(i) $\mathbb{E}[T] < \infty$ *and there is a constant C such that*

$$\mathbb{E}[|X_{n+1} - X_n| \,|\, X_1, \dots, X_n] \leq C \textit{almost surely};$$

(ii) $X_1, X_2, \ldots$ *is uniformly integrable and* T *is almost surely finite;*
(iii) T *is bounded.*

Then the process $X_{n\wedge T}$ *is uniformly integrable.*

Theorem 3.1 (Optional Stopping)**.** *If* (X_n) *is a submartingale then* $(X_{n\wedge T})$ *is a submartingale. If* $X_{n\wedge T}$ *is uniformly integrable then*

$$\mathbb{E}\left[X_T \mid X_0\right] \geq X_0.$$

For a supermartingale the same holds with reversed inequalities. And if (X_n) *is a martingale with* $X_{n\wedge T}$ *uniformly integrable then*

$$\mathbb{E}\left[X_T \mid X_0\right] = X_0.$$

Note that uniform integrability is definitely required. Consider, for example, the simple random walk started at 0, stopped at $T = \min\{n : S_n = -1\}$. It is a martingale, and T is almost surely finite, but $\mathbb{E}[X_T] = -1 < 0 = X_0$. The reason is that $\{S_n\}$ are not uniformly integrable, and T has infinite expectation.

Theorem 3.2 (Martingale convergence Theorem)**.** *Suppose* $X_0, X_1, \ldots$ *is a submartingale with* $\mathbb{E}[X_n^+] < K$ *for some fixed* K*. Then* $X_\infty := \lim_{n\to\infty} X_n$ *exists almost surely, and* $\mathbb{E}[|X_\infty|] < \infty$*.*

Example: Let $X_1, X_2, \ldots$ be an i.i.d. sequence with $\mathbb{E}[X_i] = 0$ and nonzero variance, with $\mathbb{P}\{X_i \leq K\} = 1$ Let $S_n = X_1 + \cdots + X_n$. Let $T = \min\{n : S_n \geq 0\}$. $S_{T\wedge n}$ is a martingale (hence also submartingale), and $S_{T\wedge n} \leq K$ for all n. Then $S_{T\wedge n}$ converges. Since S_n cannot converge to any finite value (since it is changing by fixed increments at each step), this implies that $S_{n\wedge T}$ is stopped by hitting T. Hence $T < \infty$ almost surely.

4. Martingale Central Limit Theorem

It is clear that the Martingale CLT cannot be quite as simple as the CLT for identically distributed random variables, since the variance of the successive martingale differences not only could easily be different, but they could be random, depending on past random variables. There are numerous versions, described at length in [2], which is available online at `http://www.stat.yale.edu/~mjk56/MartingaleLimitTheoryAndItsApplication.pdf`

One standard approach is to take a random number of terms to make up a fixed variance.

Theorem 4.1 (Martingale Central Limit Theorem)**.** *Let* $\xi_1, \xi_2, \ldots$ *be a mean-0 martingale difference sequence with* ξ_i *bounded, and* $X_n = \xi_1 + \cdots + \xi_n$*. Let*

$$\sigma_n^2 := \mathbb{E}\left[\xi_{n+1}^2 \mid \xi_1, \ldots, \xi_n\right]$$

be the n*-th conditional variance. For each* $v > 0$ *let*

$$\tau_v := \min\{n : \sum_{i=0}^{n} \sigma_i^2 > v\}.$$

Then if τ_v is almost-surely finite for all v,

$$\frac{X_{\tau_v}}{\sqrt{v}} \xrightarrow{v\to\infty}_d \mathcal{N}(0,1).$$

References

[1] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[2] Peter Hall and Christopher C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, London, 1980.

[3] I. A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability and its Applications*, 7:349–82, 1962.

[4] Michael Keane and Karl Petersen. Easy and nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. *Lecture Notes–Monograph Series: Dynamics & Stochastics*, 48:248–51, 2006.

[5] Jeffrey S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, second edition, 2006.

[6] Gennady Samorodnitsky. *Stochastic processes and long range dependence*. SV, 2016.

[7] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.