

1 Analysis of Markov Chains

1.1 Martingales

Martingales are certain sequences of dependent random variables which have found many applications in probability theory. In order to introduce them it is useful to first re-examine the notion of conditional probability. Recall that we have a probability space Ω on which random variables are defined. For simplicity we may assume the random variables assume only a countable set of values, however, the results are valid for random variables taking values in $\mathbf{R}$. To obtain this greater generality requires a refinement of the notion of conditional probability and expectation as discussed below. The proofs remain valid without any essential change after replacing the word partition (see below) by the appropriate σ -algebra. The assumption that random variables take only countably many values dispenses with the need to introduce the concept of σ -algebras. For random variables X and Y conditional probability $P[X = a|Y]$ is itself a random variable. For every $b \in \mathbf{Z}$ such that $P[Y = b] \neq 0$ we have

$$P[X = a \mid Y = b] = \frac{P[X = a \text{ and } Y = b]}{P[Y = b]}.$$

Similarly conditional expectation $E[X|Y]$ is a random variable which takes different values for different b 's. We can put this in a slightly different language by saying that Y defines a partition of the probability space Ω ;

$$\Omega = \bigcup_{b \in \mathbf{Z}} A_b, \quad (\text{disjoint union}) \tag{1.1.1}$$

where $A_b = \{\omega \in \Omega \mid Y(\omega) = b\}$. Thus conditional expectation is a random variable which is constant on each piece A_b of the partition defined by Y . With this picture in mind we redefine the notion of conditioning by making use of partitionings of the probability space. So assume we are given a partition of the probability space Ω as in (1.1.1), however we do not require that this partition be defined by any specifically given random variable. It is just a partition which somehow has been specified. Each subset A_b is an event and if A_b has non-zero probability, then $P[X|A_b]$ and $E[X|A_b]$ make sense. It is convenient to introduce a succinct notation for a partition. Generally we use $\mathcal{A}$ or $\mathcal{A}_n$ to denote a partition or sequence of partitions which we shall

encounter shortly. Notice that each $\mathcal{A}_n$ is a partition of the probability space and the subscript n does not refer to the subsets comprising the partition $\mathcal{A}$. By $\mathcal{A} \prec \mathcal{A}'$ we mean every subset of Ω defined by the partition $\mathcal{A}$ is a union of subsets of Ω defined by the partition $\mathcal{A}'$. In such a case we say $\mathcal{A}'$ is a *refinement* of $\mathcal{A}$. For example if Y and Z are random variables, we define the partition $\mathcal{A}_Y$ as

$$\Omega = \bigcup A_b^Y, \text{ where } A_b^Y = \{\omega | Y(\omega) = b\},$$

and similarly for $\mathcal{A}_Z$. Then the collection of intersections $A_b^Y \cap A_c^Z$ defines a partition $\mathcal{A}'$ which is a refinement of both $\mathcal{A}^Y$ and $\mathcal{A}^Z$. The set $A_b^Y \cap A_c^Z$ consists of all $\omega \in \Omega$ such that $Y(\omega) = b$ and $Z(\omega) = c$. For a random variable X notion of $P[X = a | \mathcal{A}]$ simply means that for every subset of positive probability defined by the partition $\mathcal{A}$ we have a number which is the conditional probability of $X = a$ given that subset. Thus $P[X = a | \mathcal{A}]$ itself is a random variable which is constant on each subset A_b of the partition defined by $\mathcal{A}$. Similarly, the conditional expectation $E[X | \mathcal{A}]$ is a random variable which is constant on each subset defined by the partition $\mathcal{A}$.

Given a partition $\mathcal{A}$ we say a random variable X is $\mathcal{A}$ -*admissible* if X is constant on every subset defined by $\mathcal{A}$. For example, the random variables $P[X = a | \mathcal{A}]$ and $E[X | \mathcal{A}]$ are $\mathcal{A}$ -admissible. For an $\mathcal{A}$ -admissible random variable X clearly we have

$$E[X | \mathcal{A}] = X. \tag{1.1.2}$$

If $\mathcal{A}'$ is a refinement of the partition $\mathcal{A}$, then

$$E[E[X | \mathcal{A}'] | \mathcal{A}] = E[X | \mathcal{A}]. \tag{1.1.3}$$

This is a generalization of the statment $E[E[X|Y]] = E[X]$ and its proof is again just a re-arrangement of a series.

With the preliminaries out of the way we can now define martingale. A sequence $(X_l, \mathcal{A}_l)$ consisting of random variables X_l and partitions $\mathcal{A}_l$ such that $\mathcal{A}_l \prec \mathcal{A}_{l+1}$ is called a *martingale* if

1. $E[|X_l|] < \infty$.
2. $E[X_{l+1} | \mathcal{A}_l] = X_l$.

If condition (2) is replaced by $E[X_{l+1} \mid \mathcal{A}_l] \geq X_l$, (resp. $E[X_{l+1} \mid \mathcal{A}_l] \leq X_l$) then we refer to $(X_l, \mathcal{A}_l)$ as a *submartingale* (resp. *supermartingale*). The statement (X_l) is a martingale means $(X_l, \mathcal{A}_l)$ is a martingale where $\mathcal{A}_l$ is the partition defined by the random variables $X_0, X_1, \dots, X_l$. The first condition is simply a necessary technical requirement since manipulations involving expectations are problematic unless we assume finiteness! The second condition is the essential feature. To motivate this requirement and why such sequences are called martingales we look at the following basic example:

Example 1.1.1 Suppose we are playing a fair game such as receiving (resp. paying) \$1 for every appearance of an H (resp. a T) in tosses of a fair coin. Let X_n denote the total winnings after n trials. Let $\mathcal{A}_n$ denote the partition of the probability space according to the random variables $X_1, \dots, X_n$. This means for every set of integers $(a_1, \dots, a_n)$ we define the subset $A_{(a_1, \dots, a_n)}$ as the set of $\omega \in \Omega$ (i.e., paths) such that $X_j = a_j$. In this case of course $A_{(a_1, \dots, a_n)}$ is either empty or consists of a single path. Since the coin is assumed to be fair, the expected winnings after $(n+1)^{\text{st}}$ trial is the same as the actual winnings after n^{th} trial. This is just the second requirement. There was a misconception that in a fair game if one player doubles his/her last bet after every loss, then he/she will surely win. This strategy is called a martingale in gambling circles. One of the basic theorems about martingales is that under certain conditions no strategy will change the expected winnings of a fair game which is zero. Intuitively, the flaw in the application of the martingale strategy to gambling is that while the player may win many times, the actual amount won every time is small. On the other hand, since no player has infinite amount of money, there is positive probability of losing everything. On balance the expected winnings do not change. Many have lost considerable sums of money by relying on the appearance of winning strategies which do not exist. ♠

The significance of martingales is based on two fundamental theorems which state and apply in this subsection and prove in the next. The idea behind the first theorem is making mathematics out of the fact that changing strategies does not affect the the fairness of a game. To do so we need some definitions. Assume we are given a sequence of partitions $\mathcal{A}_n$ with $\mathcal{A}_n \prec \mathcal{A}_{n+1}$. A random variable $T : \Omega \rightarrow \mathbf{Z}_+$ such that the set $\{\omega \in \Omega \mid T(\omega) = n\}$ is

a union of subsets of $\mathcal{A}_n$ is called a *Markov time*. If in addition $P[T < \infty] = 1$, then T is called a *stopping time*. There is an important technical condition which appears in connection with many theorems involving martingales. It is called uniform integrability. Let X_n be a sequence of random variables. For a positive real number c let

$$\Omega^{(n,c)} = \{\omega \in \Omega \mid |X_n(\omega)| > c\},$$

and

$$I_{|X_n|>c}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega^{(n,c)}; \\ 0 & \text{otherwise.} \end{cases}$$

That is, $I_{|X_n|>c}$ is the indicator function of the set $\Omega^{(n,c)}$. *Uniform integrability* of the sequence X_n means

$$\sup_n \mathbb{E}[|X_n| I_{|X_n|>c}] \rightarrow 0 \quad \text{as } c \rightarrow \infty.$$

We will discuss the meaning and significance of this condition later. For the time being we will not dwell on uniform integrability and try to understand the implications of martingale property assuming that this technical condition is fulfilled. The first fundamental result about martingales is a pair of related propositions known as the Optional Stopping Time and Optional Sampling Time theorems.

Proposition 1.1.1 *Let $(X_n, \mathcal{A}_n)$ be a martingale, T a stopping time, and assume*

1. $\mathbb{E}[|X_T|] < \infty$;
2. $\mathbb{E}[X_n \mid T > n]P[T > n] \rightarrow 0$ as $n \rightarrow \infty$.

Then

$$\mathbb{E}[X_T] = \mathbb{E}[X_1].$$

If $(X_n, \mathcal{A}_n)$ is a submartingale and the same hypotheses hold then $\mathbb{E}[X_T] \geq \mathbb{E}[X_1]$.

Proposition 1.1.2 *Let $(X_n, \mathcal{A}_n)$ be a uniformly integrable martingale, and X_∞ be its limit, then for any stopping time T_1 we have*

$$\mathbb{E}[X_\infty \mid \mathcal{A}_{T_1}] = X_{T_1}.$$

If T_2 is any stopping time and $P[T_2 \geq T_1] = 1$, then

$$\mathbb{E}[X_{T_2} \mid \mathcal{A}_{T_1}] = \mathbb{E}[X_{T_1}].$$

If $(X_n, \mathcal{A}_n)$ is a uniformly integrable submartingale, and the same hypotheses hold, then the same assertions are valid after replacing $=$ by $\geq$.

To understand the meaning of these results in the context of games, note that T (the stopping time) is the mathematical expression of a strategy in a game. Proposition 1.1.1 asserts that the expected winnings $\mathbb{E}[X_T]$ is in fact independent of the strategy and equal to $\mathbb{E}[X_0]$. To better appreciate the significance of these results and to maintain the continuity of the treatment we consider some examples.

Example 1.1.2 To appreciate the significance of the condition of proposition 1.1.1 let us assume a person is playing a fair game and his/her rather bizarre strategy is to stop playing once he/she loses \$100. Obviously with this strategy the assertion of theorem $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ is no longer valid. However this does not violate proposition 1.1.1. Mathematically, the given strategy assigns to each path the time it reaches $-\$100$. It is a simple exercise to show that T is a stopping time, the second condition in proposition 1.1.1 is not fulfilled. The purpose of this example was to show that some hypothesis is necessary to make the remarkable conclusion of propositions 1.1.1 or 1.1.2 valid. This motivates the technical condition of uniform integrability which will be introduced later. ♠

Example 1.1.3 Consider the simple random walk $S_0 = k$, $S_l = X_l + S_{l-1}$ on $\mathbf{Z}$ which moves one unit to the right with probability p and one unit to the left with probability $q = 1 - p$. Assume $p \neq q$ and set

$$Y_l = \left(\frac{q}{p}\right)^{S_l}$$

Let $\mathcal{A}_l$ be the partition defined by the random variables $X_1, \dots, X_l$. Then $(Y_l, \mathcal{A}_l)$ is a martingale. In fact we have

$$\begin{aligned}
\mathbb{E}[Y_{l+1} \mid \mathcal{A}_l] &= \mathbb{E}\left[\left(\frac{q}{p}\right)^{X_{l+1}+S_l} \mid \mathcal{A}_l\right] \\
&= \left(\frac{q}{p}\right)^{S_l} \mathbb{E}\left[\left(\frac{q}{p}\right)^{X_{l+1}} \mid \mathcal{A}_l\right] \\
&= \left(\frac{q}{p}\right)^{S_l} [p \frac{q}{p} + q \frac{p}{q}] \\
&= Y_l.
\end{aligned}$$

Now assume $0 < k < N$ and T denote the time the random walk hits 0 or N . Then proposition 1.1.1 implies

$$\mathbb{E}\left[\left(\frac{q}{p}\right)^{S_T}\right] = \mathbb{E}[Y_0] = \left(\frac{q}{p}\right)^k.$$

Let

$$p_k = P[\text{Absorption at } 0 \mid S_0 = k], \quad 1 - p_k = P[\text{Absorption at } N \mid S_0 = k].$$

Then

$$\mathbb{E}[Y_T] = \left(\frac{q}{p}\right)^0 p_k + \left(\frac{q}{p}\right)^N (1 - p_k) = \left(\frac{q}{p}\right)^k.$$

Solving for p_k we obtain

$$p_k = \frac{\alpha^k - \alpha^N}{1 - \alpha^N}, \quad \text{where } \alpha = \frac{q}{p}.$$

We could have derived this familiar result for gambler's ruin problem by more elementary means but the use of martingales simplified the treatment and demonstrates how martingales are applied. ♠

Example 1.1.4 Let us consider a coin tossing experiment where H 's appear with probability p and T 's with probability $q = 1 - p$. We want to calculate the expected time of the appearance of a given pattern. Let us begin with a very simple pattern, say TT . First we have to construct a martingale which allows us to make the calculation. Assume there is a casino and one player enters the casino at the passage of each unit of time. Let us assume that every player bets \$1 on T at first. If H appears the player loses \$1. If T appears the player collects $\$(1 + \frac{p}{q})$ to make the game fair. Since we are interested in the appearance of the pattern TT , we make the stipulation that if T appears, then the player bets the entire $\$(1 + \frac{p}{q})$ on the next bet. It should be emphasized that a new player enters the casino after the passage of a unit of time regardless of whether H or T appears. Let S_l denote the

total winnings of the casino after passage of l units of time, and T be the time of the appearance of TT . If $T = N$ then

$$S_{N-1} = N - 2 - \frac{p}{q}, \quad S_N = N - 2 - \frac{p}{q} - \frac{p}{q}\left(1 + \frac{p}{q}\right) - \frac{p}{q}.$$

On the other hand, proposition 1.1.1 tells us $\mathbb{E}[S_T] = \mathbb{E}[S_0] = 0$. Therefore $T = N$ yields, after a simple calculation,

$$\mathbb{E}[T] = \frac{1}{q} + \frac{1}{q^2}$$

as the expected time of the first appearance of TT . Of course we had calculated this quantity by using generating functions (see example ??). It should be pointed out that in this example and in the application of martingales one should judiciously construct a sequence of random variables which form a martingale and that it should be possible to make the calculation of the relevant quantity (S_T in this case) easily. This is often the tricky step in the application of martingales. The fact that we chose a very simple pattern TT is not essential, and the argument works for complex patterns as well. ♠

An important and intuitive corollary to proposition 1.1.1 is

Corollary 1.1.1 (Wald Identity) *Let $X_1, X_2, \dots$ be a sequence of iid random variables with $\mathbb{E}[|X_i|] < \infty$. Let T be a stopping time for the sequence X_i with $\mathbb{E}[T] < \infty$. Then*

$$\mathbb{E}[X_1 + \dots + X_T] = \mathbb{E}[X_1]\mathbb{E}[T].$$

If furthermore $\mathbb{E}[X_i^2] < \infty$, then

$$\mathbb{E}[(X_1 + \dots + X_T) - T\mathbb{E}[X_1]]^2 = \mathbb{E}[T]\text{Var}[X_1].$$

Proof - Let $\mathcal{A}_n$ be the partition defined by the random variables $X_1, \dots, X_n$. Set

$$Y_l = X_1 + \dots + X_l - l\mathbb{E}[X_1].$$

It is clear that the sequence $(Y_l, \mathcal{A}_l)$ is a martingale with $\mathbb{E}[Y_l] = 0$. Therefore

$$0 = \mathbb{E}[Y_T] = \mathbb{E}[X_1 + \dots + X_T] - \mathbb{E}[T]\mathbb{E}[X_1],$$

whence the first assertion. Set

$$Z_l = (X_1 + \cdots + X_l - l\mathbb{E}[X_1])^2 - l\text{Var}[X_1]$$

Then $(Z_l, \mathcal{A}_l)$ is a martingale, and applying theorem to proposition 1.1.1 we obtain the second assertion. ♣

Exercise ?? shows that Wald's identity is not valid if T were not a stopping time.

The second basic theorem on martingales is about convergence. It is useful to clarify the meaning of some of the notions of convergence which are commonly used in probability and understand their relationship. Let $X_1, X_2, \dots$ be a sequence of random variables (defined on a probability space Ω).

1. (*L^p Convergence*) - X_n converges to a random variable X (defined on Ω) in L^p if

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

We will be mainly interested in $p = 1, 2$. The Cauchy-Schwartz inequality in this context can be stated as

$$|\mathbb{E}[WZ]| \leq \sqrt{\mathbb{E}[W^2]\mathbb{E}[Z^2]},$$

where W and Z are real-valued random variables. Substituting $W = |X_n - X|$ and $Z = 1$ we deduce that convergence in L^2 implies convergence in L^1 . Note that if $X_n \rightarrow X$ in L^1 then $\mathbb{E}[X_n]$ converges to $\mathbb{E}[X]$.

2. (*Pointwise Convergence*) - X_n converges to X pointwise if for every $\omega \in \Omega$ the sequence of numbers $X_n(\omega)$ converges to $X(\omega)$. Pointwise convergence does not imply convergence in L^1 and $\mathbb{E}[X_n]$ may not converge to $\mathbb{E}[X]$ as shown in example ?. It is often more convenient to relax the notion of pointwise convergence to that of *almost pointwise convergence*. This means there is a subset $\Omega_o \subset \Omega$ such that $P[\Omega_o] = 1$ and $X_n \rightarrow X$ pointwise on Ω_o . We have already seen how deleting a set of probability zero makes it possible to make precise statements about the behavior of a sequence of random variables. For example,

with probability 1 (i.e., in the complement of a set of paths of probability zero) a transient state is visited only finitely many times. Or the (strong) law of large numbers states that for an iid sequence of random variables $X_1, X_2, \dots$ with mean $\mu = \mathbf{E}[X_j]$, we have almost pointwise convergence of the sequence $\frac{X_1 + \dots + X_n}{n}$ to μ . Related to almost pointwise convergence is the weaker notion of convergence probability which, for example, in the case of the Law of Large Numbers leads to the Weak Law of Large Numbers.

3. (*Convergence in Distribution*) - Let F_n be the distribution function of X_n and F that of X . Assume F is a continuous function. X_n converges to X in distribution means for every $x \in \mathbf{R}$ we have

$$P[X_n \leq x] = F_n(x) \longrightarrow F(x) = P[X \leq x].$$

The standard statement of the Central Limit Theorem is about convergence in distribution. Convergence in distribution does not imply almost pointwise convergence, but almost pointwise convergence implies convergence in distribution.

The essential point for our immediate application is that convergence in L^1 implies convergence of expectations, however, weaker notions of convergence do not.

Recall the statement of the submartingale convergence theorem:

Theorem 1.1.1 *Let $(X_n, \mathcal{A}_n)$ be a submartingale and assume*

$$\sup_n \mathbf{E}[X_{n+}] < \infty, \quad |\mathbf{E}[X_1]| < \infty.$$

Then X_n converges almost everywhere to an integrable random variable X_∞ . If the submartingale $(X_n, \mathcal{A}_n)$ is uniformly integrable, then convergence is also in L^1 .

The following example clarifies the need for the condition of uniform integrability.

Example 1.1.5 Consider the sequence of functions $\{f_n\}$ defined on $[0, 1]$ defined as (draw a picture to see a sequence of *spike* functions)

$$f_n(x) = \begin{cases} 2^{2n}x & \text{if } 0 \leq x \leq 2^{-n} \\ -2^{2n}x + 2^{n+1} & \text{if } 2^{-n} < x \leq 2^{-n+1} \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that

$$\int_0^1 f_n(x)dx = 1,$$

and the sequence f_n converges to the zero function everywhere on $[0, 1]$. Therefore

$$\lim_{n \rightarrow \infty} \int_0^1 f_n(x)dx \neq \int_0^1 \lim_{n \rightarrow \infty} f_n(x)dx.$$

This is obviously an undesirable situation which should be avoided since when dealing with limiting values we want the integrals (e.g. expectations) to converge to the *right* values. The assumption of uniform integrability eliminates cases like this when we cannot interchange limit and integral. It is easy to see that the uniform integrability condition is not satisfied for the sequence $\{f_n\}$. ♠

Example 1.1.6 Let Ω be a probability space, $\mathcal{A}$ a σ -algebra. and X an $\mathcal{A}$ -admissible random variable with the property $E[|X|] < \infty$. Let $\mathcal{A}_n$ be a family of σ -algebras with $\mathcal{A}_n \prec \mathcal{A}_{n+1} \prec \mathcal{A}$ for all n . Let $Y_n = E[X | \mathcal{A}_n]$. We show that $(Y_l, \mathcal{A}_l)$ is a uniformly integrable martingale. We have

$$\begin{aligned} E[|Y_l|] &= E[|E[X | \mathcal{A}_l]|] \\ &= \leq E[E[|X| | \mathcal{A}_l]] \\ &= E[|X|] < \infty; \end{aligned}$$

and

$$\begin{aligned} E[Y_{l+1} | \mathcal{A}_l] &= E[E[X | \mathcal{A}_{l+1}] | \mathcal{A}_l] \\ &= E[X | \mathcal{A}_l] \\ &= Y_l, \end{aligned}$$

proving Martingale property. The martingale $(Y_l, \mathcal{A}_l)$ is uniformly integrable. To prove this set $Z_n = E[|Y_n| | \mathcal{A}_n]$ and note that

$$E[(|X| - Z_n)I_{[Z_n \geq a]}] = 0,$$

where I_A denotes the indicator function of the set A . Therefore

$$\begin{aligned} E[|X|I_{[Z_n \geq a]}] &= E[E[|X| | \mathcal{A}_n]I_{[Z_n \geq a]}] \\ &\geq E[E[X | \mathcal{A}_n]I_{[Z_n \geq a]}] \\ &= E[Y_n I_{[Z_n \geq a]}] \end{aligned}$$

By Markov inequality

$$P[Z_n \geq a] \leq \frac{E[Z_n]}{a} = \frac{E[|X|]}{a},$$

and therefore $P[Z_n \geq 0] \rightarrow 0$ as $a \rightarrow \infty$ uniformly in n . Since $|X|$ is integrable, $E[|X|I_{[Z_n \geq a]}] \rightarrow 0$ uniformly in n which implies uniform integrability of Y_n . ♠

Example 1.1.7 A special case of Doob's martingale is when $\Omega = [0, 1]$, X is an integrable function on $[0, 1]$ and $\mathcal{A}_n$ is the partition of $[0, 1]$ into 2^n intervals of length $\frac{1}{2^n}$. Then X_n is the approximation by the step function to X which on each interval of length $\frac{1}{2^n}$ of the partition is integral of X on that interval. In this fashion Doob's martingale gives a sequence of approximations to X where by increasing n the level of detail increases. ♠

It is a remarkable fact about uniformly integrable martingales that they are necessarily of the form of Doob's martingale in the sense described below:

Corollary 1.1.2 *Let $(X_n, \mathcal{A}_n)$ be a uniformly integrable martingale, then $X_n \rightarrow X_\infty$ almost everywhere and in L^1 and $X_n = E[X_\infty | \mathcal{A}_n]$ almost everywhere.*

Proof - In view of theorem 1.1.1 it only remains to show $X_n = E[X_\infty | \mathcal{A}_n]$. It suffices to show that for every $A \in \mathcal{A}_n$ of positive probability, we have

$$X_n = E[X_n | A] = E[X_\infty | A]$$

In view of martingale property for $m \geq n$ we have $E[X_n | A] = E[X_m | A]$. Now

$$|E[X_m | A] - E[X_\infty | A]| \leq E[|X_m - X_\infty| | A] \leq \frac{E[|X_m - X_\infty|]}{P[A]}$$

which tends to 0 as $m \rightarrow \infty$. The required result follows. ♣

At this point it is instructive to prove Kolmogorov's zero-one law. The proof elegantly uses of some of the concepts we have introduced.

Corollary 1.1.3 *Let $X_1, X_2, \dots$ be independent random variables. Let $\mathcal{B}_n$ be the partition defined by the random variables $X_n, X_{n+1}, \dots$ and $\mathcal{T} = \cap \mathcal{B}_n$. Let $A \in \mathcal{T}$, then $P[A] = 0$ or 1.*

Proof - Let $\mathcal{A}_n$ be the partition defined by the random variables $X_1, \dots, X_n$, and $\mathcal{A}_\infty = \lim_{n \rightarrow \infty} \mathcal{A}_n$. It is clear from the hypotheses that $A \in \mathcal{A}_\infty$. Let I_A

denote the indicator function of the set A , and set $Y_n = \mathbb{E}[I_A \mid \mathcal{A}_n]$. From $A \in \mathcal{A}_\infty$ it follows that

$$Y_n \longrightarrow \mathbb{E}[I_A \mid \mathcal{A}_\infty] = I_A \quad (1.1.4)$$

almost surely and in L^1 . Since the random variables are independent and $A \in \mathcal{T} = \cap \mathcal{B}_n$, A is independent of $\mathcal{A}_n$ for every n . Therefore

$$Y_n = \mathbb{E}[I_A \mid \mathcal{A}_n] = P[A]. \quad (1.1.5)$$

From (1.1.4) and (1.1.5) it follows that $P[A] = 0$ or 1 . ♣

It is customary to call a event $A \in \mathcal{T}$ a *tail event* and a and a $\mathcal{T}$ -admissible random variable a *tail function*. An equivalent way of stating Kolmorove's 0-1 law is

Corollary 1.1.4 *A tail function is constant almost everywhere.*

Example 1.1.8 Let us give a simple example of a tail event. Let $a_1, a_2, \dots$ be random numbers chosen from $[0, 1]$ independently and according to the uniform distribution. Whether or not the series $\sum a_n e^{i2\pi n\theta}$ converges is a tail event since it does not depend on any finite number of a_n 's. Therefore with probability one either all such series converge or diverge. It is not difficult to convince oneself that the latter alternative is the case. ♠

As another applications of martingales we derive the Strong Law of Large Numbers (SLLN). This requires introducing the notion of *reverse martingale* which is a sequence $X_1, X_2, \dots$ of random variables. together with partitions $\mathcal{A}_n \succ \mathcal{A}_{n+1}$ such that

$$\mathbb{E}[X_n \mid \mathcal{A}_{n+1}] = X_{n+1}.$$

Note that for reverse martingales the partition $\mathcal{A}_n$ is finer than $\mathcal{A}_{n+1}$ as opposed to the case of martingales where the opposite is true.

Example 1.1.9 For our applications the reverse martingale which is of interest is the analogue of Doob's martingale. In fact assume that X is an integrable random variable and $\mathcal{A}_n \succ \mathcal{A}_{n+1}$ are partitions where $n = 1, 2, \dots$. Then $Y_n = \mathbb{E}[X \mid \mathcal{A}_n]$ together with the partitions $\mathcal{A}_n$ define a reverse martingale. Let $\mathcal{A}_\infty = \cap \mathcal{A}_n$. Then

$$Y_n \longrightarrow \mathbb{E}[X \mid \mathcal{A}_\infty]$$

almost everywhere and in L^1 . The proof of this fact is similar to that in case of Doob's martingale and will not be repeated. ♠

Corollary 1.1.5 *Let $X_1, X_2, \dots$ be iid random variables with $E[X_k] < \infty$, and $S_n = X_1 + \dots + X_n$. Then $\frac{S_n}{n}$ converges to μ almost everywhere and in L^1 .*

Proof - It is clear that for $k \leq n$

$$E[X_1] = \frac{S_n}{n} \quad \text{almost everywhere.} \quad (1.1.6)$$

Independence of random variables X_1 and $X_n, X_{n+1}, X_{n+2}, \dots$ and (1.1.6) imply

$$\begin{aligned} \frac{S_n}{n} &= E[X_1 \mid S_n, X_{n+1}, X_{n+2}, \dots] \\ &= E[X_1 \mid S_n, S_{n+1}, S_{n+2}, \dots]. \end{aligned}$$

Let $\mathcal{B}_n$ denote the partition generated by the $S_n, S_{n+1}, S_{n+2}, \dots$ and $\mathcal{T} = \cap \mathcal{B}_n$. Since $\lim_{n \rightarrow \infty} E[X_1 \mid \mathcal{B}_n]$ is a tail function, it is a constant c by corollary 1.1.4. From example 1.1.9 convergence of $E[X_1 \mid \mathcal{B}_n]$ to c is almost everywhere and in L^1 . The constant c is necessarily μ since $E[\frac{S_n}{n}] = \mu$. ♣

We now give some examples of martingales which directly involve Markov chains. Let $X_0, X_1, \dots$ be a Markov chain with transition matrix P . Linear combinations of the states of the Markov chain are given as row vectors and functions on state space are described as column vectors. Thus the action of the Markov chain on functions on the state space is by $f \rightarrow Pf$. A function f on the state space is called *harmonic* (resp. *subharmonic*) if $Pf = f$ (resp. $Pf \geq f$).

Example 1.1.10 In this example we show how a Markov chain and a bounded function f on the state space give rise to a martingale. Let P denote the transition matrix of the Markov chain $X_0, X_1, \dots$. Define

$$Y_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (P - I)f(X_k).$$

Here $(P - I)f(X_k)$ is the obvious matrix/vector multiplication. The boundedness condition on f is necessary to ensure existence of infinite sums $\sum_i P_{ij}f_j$

and integrability of the random variables Y_n^f when the state space is infinite. Let $\mathcal{A}_n$ be the partition defined by $X_0, \dots, X_n$. It is easy to see that $(Y_n^f, \mathcal{A}_n)$ is a martingale. In fact, we have

$$Y_{n+1}^f - Y_n^f = f(X_{n+1}) - Pf(X_n).$$

Since $\mathbb{E}[f(X_{n+1})|\mathcal{A}_n] = Pf(X_n)$ we have

$$\mathbb{E}[Y_{n+1}^f|\mathcal{A}_n] - Y_n^f = \mathbb{E}[Y_{n+1}^f - Y_n^f|\mathcal{A}_n] = 0$$

proving the martingale property. This martingale is often called *Lévy's martingale*. ♠

To a Markov chain $X_0, X_1, \dots$ with transition matrix P we associated the operator $P - I$ defined on functions on the state space. This operator in some ways resembles the Laplace operator on a domain in $\mathbf{R}^n$ or a manifold, and the analogy has led to a number of interesting results on Markov chains. Assume the transition matrix has the property that $P_{ij} \neq 0$ if and only if $P_{ji} \neq 0$. Then to the Markov chain we assign a graph whose vertices are the set of states and two vertices i and j are connected by an edge if $P_{ij} \neq 0$. For our purposes here the graph and the assumption on the transition matrix are not essential and are noted only for better pictorial representation of the Markov chain. A subset $U \subset S$ is called a *domain* and its complement is denoted by U' . The boundary ∂U of U consists of the set of vertices (states) $k \in U'$ such that there is $j \in U$ with $P_{jk} \neq 0$. A problem which is analogous to a classical boundary value problem and is of intrinsic interest for example in Markov decision problems is to solve the equation

$$(P - I)u = \phi \text{ on } U, \quad \text{subject to } u = \psi \text{ on } \partial U. \quad (1.1.7)$$

The solution u can in fact be defined on the entire state space S . The domain U can have empty boundary for example if $U = S$. As before we let $\mathbb{E}_j[Y]$, where Y is a random variable depending on the Markov chain, denote the expectation of Y conditioned on $X_0 = j$.

Proposition 1.1.3 *Let T be the first hitting time of the boundary ∂U and assume for every $j \in U$, $P[T < \infty] = 1$. Then the function*

$$u(j) = \mathbb{E}_j[\psi(X_T) - \sum_{k=0}^{T-1} \phi(X_k)],$$

defined for $j \in U \cup \partial U$, is a solution to (1.1.7). If ϕ and ψ are non-negative then u is the unique bounded non-negative solution.

Proof - The fact that U solves the problem (1.1.7) follows easily by conditioning to X_1 and is left to the reader. To prove uniqueness let u be a solution to (1.1.7) and $Y_n^u = Y_n$ be the Lévy martingale associated to u :

$$Y_n = u(X_n) - u(X_0) - \sum_{k=0}^{n-1} (P - I)u(X_k).$$

For every integer m , $T \wedge m$ is also a stopping time and it follows that

$$\mathbb{E}[Y_{T \wedge m}] = \mathbb{E}_j[Y_0] = 0.$$

Therefore the fact that $(P - I)u = \phi$ on U implies

$$u(j) = \mathbb{E}_j[u(X_{T \wedge m}) - \sum_{k=0}^{(T \wedge m)-1} \phi(X_k)]$$

Now taking $\lim_{m \rightarrow \infty}$ we see that

$$u(j) = \mathbb{E}_j[\psi(X_T) - \sum_{k=0}^{T-1} \phi(X_k)]$$

which is the desired uniqueness. Notice that in the application of $\lim_{m \rightarrow \infty}$ we made use of the assumption of non-negativity since otherwise the monotone convergence theorem may not have been applicable. ♣

1.2 Some Proofs

There are results similar in spirit to the Optional Stopping Time theorem which are useful in martingale theory and in the proof of theorem 1.1.1. One such result is lemma 1.2.1 below which we first describe as a strategy in a game of chance. Consider an individual playing a game of chance against a casino. The game is played at regular intervals and every time the player has the option of playing or not playing. The player decides to bet or not bet at $(k + 1)^{\text{st}}$ trial (codified by $\epsilon_{k+1} = 1$ or 0 in lemma 1.2.1) according to the information he/she has from the first k trials. The point of lemma 1.2.1 is that no matter what strategy is used (i.e., choice of ϵ_k 's) the expected value will not improve. More precisely, we have

Lemma 1.2.1 *Let $(X_n, \mathcal{A}_n)$ be a submartingale, $\epsilon_k, k = 1, 2, \dots, \mathcal{A}_k$ -admissible random variables taking values 1 and 0 only. Set $Y_1 = X_1$ and*

$$Y_{k+1} = Y_k + \epsilon_k(X_{k+1} - X_k).$$

Then $E[Y_n] \leq E[X_n]$. If $(X_n, \mathcal{A}_n)$ is a martingale, then $E[Y_n] = E[X_n]$.

Proof - We have

$$\begin{aligned} E[Y_{n+1} \mid \mathcal{A}_n] &= E[Y_n + \epsilon_n(X_{n+1} - X_n) \mid \mathcal{A}_n] \\ &= Y_n + \epsilon_n E[X_{n+1} - X_n \mid \mathcal{A}_n] \\ &\geq Y_n + \epsilon_n(X_n - X_n) \\ &= Y_n. \end{aligned}$$

Therefore $(Y_n, \mathcal{A}_n)$ is a submartingale (or martingale if $(X_n, \mathcal{A}_n)$ is so). We show by induction on n that $E[Y_n] \leq E[X_n]$. For $n = 1$ the result is obvious. Since

$$X_{n+1} - Y_{n+1} = (1 - \epsilon_n)(X_{n+1} - X_n) + X_n - Y_n,$$

the submartingale property and the induction hypothesis imply

$$E[X_{n+1} - Y_{n+1} \mid \mathcal{A}_n] \geq E[X_n - Y_n \mid \mathcal{A}_n] \geq 0. \quad (1.2.1)$$

Notice that if $(X_n, \mathcal{A}_n)$ were a martingale, then the inequalities in (1.2.1) can be replaced by $=$. The required result follows by taking another E and using $E[E[Z \mid \mathcal{A}_n]] = E[Z]$. ♣

For a random variable X we let

$$X_+ = \max(X, 0), \quad X_- = \min(X, 0).$$

It is immediate that if $(X_n, \mathcal{A}_n)$ is a submartingale, then so is the sequence $(X_{n+}, \mathcal{A}_n)$.

The key point in the proof of (sub)martingale convergence theorem is the upcrossing lemma 1.2.2 below. We need some notation. Let $X_1, \dots, X_N$ be real-valued random variables (defined as usual on a probability space Ω), $a < b$ be real numbers. For $\omega \in \Omega$ let $T_1(\omega)$ be the smallest integer $i_1 \leq N$ such that $X_{T_1}(\omega) \leq a$. If no such integer exists, we set $T_1(\omega) = \infty$. Let $T_2(\omega)$ be the smallest integer $i_2, i_1 < i_2 \leq N$, such that $X_{T_2}(\omega) \geq b$ with the same proviso $X_{T_2}(\omega) = \infty$ if no such i_2 exists. Similarly, let $T_3(\omega)$ be the smallest

integer $i_3, i_2 < i_3 \leq N$, such that $X_{T_3}(\omega) \leq a$ etc.etc. For each ω let l be the number of finite T_i 's and define the *upcrossing* random variable as

$$U_{ab}(\omega) = \begin{cases} \frac{l}{2} & \text{if } l \text{ is even;} \\ \frac{l-1}{2} & \text{if } l \text{ is odd.} \end{cases}$$

Now we can prove

Lemma 1.2.2 *With the above notation if $(X_k, \mathcal{A}_k)$, $k = 1, 2, \dots, N$, is a submartingale, then*

$$\mathbb{E}[U_{ab}] \leq \frac{1}{b-a} \mathbb{E}[(X_N - a)_+].$$

Proof - It suffices to prove the lemma for the case $a = 0$ and X_j take only non-negative values. In fact, we consider $((X_k - a)_+, \mathcal{A}_k)$ which is also a submartingale and whose validity implies that of the general case. For the proof of the assertion for $a = 0$ and $X_j \geq 0$, define

$$\epsilon_j(\omega) = \begin{cases} 1 & \text{if } T_1 \leq j < T_2, \text{ or } T_3 \leq j < T_4, \text{ or } \dots \\ 0 & \text{if } T_2 \leq j < T_3, \text{ or } T_4 \leq j < T_5, \text{ or } \dots \end{cases}$$

Define random variables Y_k according to lemma 1.2.1. It is clear that $Y_n(\omega) \geq bU_{0b}(\omega)$ and therefore lemma 1.2.1 implies

$$\mathbb{E}[U_{0b}] \leq \frac{1}{b} \mathbb{E}[Y_N] \leq \frac{1}{b} \mathbb{E}[X_N]$$

which is the required result. ♣

Proof of Submartingale Convergence Theorem - The set

$$\{\omega \mid X_k(\omega) \text{ does not converge to a finite or infinite limit}\}$$

coincides with the set

$$\cup_{a,b \in \mathbf{Q}} \{\omega \mid \liminf_{k \rightarrow \infty} X_k(\omega) \leq a < b \leq \limsup_{k \rightarrow \infty} X_k(\omega)\}.$$

($\mathbf{Q}$ is the set of rational numbers.) Therefore if X_k does not converge on a subset of Ω of positive probability, then

$$P[\omega \mid \liminf_{k \rightarrow \infty} X_k(\omega) \leq a < b \leq \limsup_{k \rightarrow \infty} X_k(\omega)] > 0$$

for some rational numbers $a < b$. It follows that X_k has an infinite number of upcrossings on a set of positive probability and consequently $\mathbb{E}[U_{ab}] = \infty$. Clearly U_{ab} is the monotone limit of $U_{ab,N}$ as $N \rightarrow \infty$ where $U_{ab,N}$ means we are computing upcrossing for random variables $X_1, \dots, X_N$ only. Therefore

$$\mathbb{E}[U_{ab,N}] \longrightarrow \mathbb{E}[U_{ab}].$$

Now

$$\mathbb{E}[U_{ab,N}] \leq \frac{1}{b-a} \mathbb{E}[(X_N - a)_+] \leq \frac{1}{b-a} \left(\sup_N \mathbb{E}[X_{N+}] + a_- \right) < \infty$$

contradicting $\mathbb{E}[U_{ab}] = \infty$. Therefore $\mathbb{E}[U_{ab}] < \infty$ and $X_N \rightarrow X_\infty$ almost everywhere. It remains to show X_∞ is integrable. We have $|X_N| = 2X_{N+} - X_N$ and submartingale property implies $\mathbb{E}[X_N] \geq \mathbb{E}[X_1]$. Therefore

$$\mathbb{E}[|X_N|] \leq 2 \sup_N \mathbb{E}[X_{N+}] - \mathbb{E}[X_1] < \infty.$$

By Fatou's lemma $\mathbb{E}[X_\infty] \leq \liminf \mathbb{E}[|X_N|] < \infty$ and therefore X_∞ is integrable. Convergence in L^1 of uniformly integrable submartingales follows from general measure theory. ♣

Propositions 1.1.1 and 1.1.2 are all in the same spirit, and their proofs have much in common as demonstrated below.

Lemma 1.2.3 *With the above notation let $T \wedge n = \min(T, n)$, then $\mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_1]$.*

Proof - We have

$$\begin{aligned} \mathbb{E}[X_n] &= \sum_{i=1}^n \mathbb{E}[X_n \mid T = i]P[T = i] + \mathbb{E}[X_n \mid T > n]P[T > n] \\ &= \sum_{i=1}^n \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{A}_1, \dots, \mathcal{A}_i] \mid T = i]P[T = i] \\ &\quad + \mathbb{E}[X_n \mid T > n]P[T > n] \\ &= \sum_{i=1}^n \mathbb{E}[X_i \mid T = i]P[T = i] + \mathbb{E}[X_n \mid T > n]P[T > n] \\ &= \mathbb{E}[X_T \mid T \leq n]P[T \leq n] + \mathbb{E}[X_n \mid T > n]P[T > n] \\ &= \mathbb{E}[X_{T \wedge n}]. \end{aligned}$$

Since $(X_n, \mathcal{A}_n)$ is a martingale, we have $\mathbb{E}[X_n] = \mathbb{E}[X_1]$, and the requires result follows. ♣

Proof of proposition 1.1.1 - We have

$$\mathbb{E}[X_T] = \mathbb{E}[X_T \mid T \leq n]P[T \leq n] + \mathbb{E}[X_T \mid T > n]P[T > n].$$

From the last equality in the proof of lemma 1.2.3 it follows that

$$\mathbb{E}[X_T] = \mathbb{E}[X_{T \wedge n}] - \mathbb{E}[X_n | T > n]P[T > n] + \mathbb{E}[X_T | T > n]P[T > n]. \quad (1.2.2)$$

Hypothesis (2) implies $\mathbb{E}[X_T | T > n]P[T > n]$ tends to 0 as $n \rightarrow \infty$. From hypothesis (1) we have

$$\mathbb{E}[|X_T|] = \sum_{k=1}^{\infty} \mathbb{E}[|X_T| | T = k]P[T = k] < \infty.$$

Therefore the tail of the convergent series on the right hand side tends to 0. It follows that

$$\mathbb{E}[X_T | T > n]P[T > n] \leq \sum_{k=n+1}^{\infty} \mathbb{E}[|X_T| | T = k]P[T = k] \rightarrow 0$$

as $n \rightarrow \infty$. Taking limit of $n \rightarrow \infty$ on right hand side of (1.2.2) we obtain the desired result for martingales. Minor modification of the same argument implies the assertion for submartingales. ♣

For the proof of proposition 1.1.2 we need

Lemma 1.2.4 *Let $(X_n, \mathcal{A}_n)$ be a martingale and T a stopping time such that with probability 1 $T \leq N < \infty$, then*

$$\mathbb{E}[X_N | \mathcal{A}_T] = X_T.$$

Proof - For $A \in \mathcal{A}_T$ of positive probability we have

$$\mathbb{E}[X_N I_A] = \sum_{n=1}^N \mathbb{E}[X_N I_{A \cap [T=n]}].$$

In view of the martingale property this becomes

$$\mathbb{E}[X_N I_A] = \sum_{n=1}^N \mathbb{E}[X_n I_{A \cap [T=n]}] = \mathbb{E}[X_T I_A],$$

proving the lemma. ♣

Proof of proposition 1.1.2 - To prove the first assertion it suffices to show for $A \in \mathcal{A}_{T_1}$ we have

$$\mathbb{E}[X_{\infty} I_A] = \mathbb{E}[X_{T_1} I_A]. \quad (1.2.3)$$

In view of corollary 1.1.2 and lemma 1.2.4 we have

$$\mathbb{E}[X_\infty|\mathcal{A}_m] = X_m, \quad \mathbb{E}[X_m|\mathcal{A}_{T_1 \wedge m}] = X_{T_1 \wedge m}.$$

It is clear that $\mathbb{E}[X_\infty|\mathcal{A}_{T_1 \wedge m}] = X_{T_1 \wedge m}$ and consequently

$$\mathbb{E}[X_\infty I_{A \cap [T_1 \leq m]}] = \mathbb{E}[X_{T_1 \wedge m} I_{A \cap [T_1 \leq m]}] = \mathbb{E}[X_{T_1} I_{A \cap [T_1 \leq m]}].$$

Taking $\lim_{m \rightarrow \infty}$ and applying the monotone convergence theorem to positive and negative parts of X_∞ we obtain (1.2.3). The assertion $\mathbb{E}[X_{T_2}|\mathcal{A}_{T_1}] = X_{T_1}$ follows from the first assertion, corollary ?? and the general property $\mathbb{E}[\mathbb{E}[X_\infty|\mathcal{A}_{T_2}]|\mathcal{A}_{T_1}] = \mathbb{E}[X_\infty|\mathcal{A}_{T_1}]$. ♣

1.3 An Idea from Analysis

Spectral theory is one of most fruitful ideas of analysis. In this and the following subsections we show how this idea can be successfully applied to solve some problems in probability theory. We do not intend to give a sophisticated treatment of the spectral theorem, rather we loosely explain the general idea and apply it to some random walk problems. Even understanding the broad outlines of the theory is a useful mathematical tool.

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ counted with multiplicity, and assume A is diagonalizable. The diagonalization of the matrix A can be accurately interpreted as taking the underlying vector space to be the space $L(S)$ of functions (generally complex-valued) on a set $S = \{s_1, \dots, s_n\}$ of cardinality n and the matrix A to be the operator of multiplication of an element $\psi \in L(S)$ by the function φ_A which has value λ_j at s_j . In this manner we have obtained the simplest form that a matrix can reasonably be expected to have. Of course not every matrix is diagonalizable but large classes of matrices including an open dense subset of them are. The idea of spectral theory is to try to do the same, to the extent possible, for infinite matrices or linear operators on infinite dimensional spaces. Even when diagonalization is possible for infinite matrices, several distinct possibilities present themselves which we now describe:

1. (*Pure Point Spectrum*) There is a countably infinite set $S = \{s_1, s_2, \dots\}$ and a complex valued function ϕ_A defined on S such that the action of the matrix A is given by multiplication by φ_A . The underlying vector space is the vector space of square summable functions on S , i.e., if we set $\psi_k = \psi(s_k)$, then $\sum_k |\psi_k|^2 < \infty$. It often becomes necessary to take a weighted sum in the sense that there is a positive weight function $\frac{1}{c_k}$ and the underlying vector space is the space of sequences $\{\psi_k\}$ such that

$$\sum_k \frac{1}{c_k} |\psi_k|^2 < \infty.$$

The weight function $\frac{1}{c_k}$ is sometimes called *Plancherel measure*.

2. (*Absolutely Continuous Spectrum*) There is an interval (a, b) (closed or open, a and/or b possibly ∞), a function φ_A such that A is the operator of multiplication of functions on (a, b) by φ_A . There is a positive or

non-negative function $\frac{1}{c(\lambda)}$ (called the *Plancherel measure*) such that the underlying vector space is the space of functions ψ on (a, b) with the property

$$\int_a^b |\psi(\lambda)|^2 \frac{d\lambda}{c(\lambda)} < \infty.$$

3. (*Singular Continuous Spectrum*) There is an uncountable set S of Lebesgue measure zero (such as the Cantor set) such that A can be realized as multiplication by a function φ_A on S . The underlying vector space is again the space of square integrable functions on S relative to some measure on S . One often hopes that the problem does not lead to this case.

4. Any combination of the above.

The first two cases and their combination is of greatest interest to us. In order to demonstrate the idea we look at some familiar examples and demonstrate the diagonalization process.

Example 1.3.1 Let A be the differentiation operator $\frac{d}{dx}$ on the space of periodic functions with period 2π . Since

$$\frac{d}{dx} e^{inx} = in e^{inx},$$

in is an eigenvalue of $\frac{d}{dx}$. Writing a periodic function as a Fourier series

$$\psi(x) = \sum_{n \in \mathbf{Z}} a_n e^{inx},$$

we see that the appropriate set S is $S = \mathbf{Z}$ and

$$\varphi_{\frac{d}{dx}}(n) = in.$$

The Plancherel measure is $\frac{1}{c_k} = \frac{1}{2\pi}$ and from the basic theory of Fourier series (Parseval's theorem) we know that

$$\int_{-\pi}^{\pi} \psi \frac{d}{dx} \bar{\phi} dx = -\frac{i}{2\pi} \sum_{n \in \mathbf{Z}} n a_n \bar{b}_n,$$

where $\psi(x) = \sum_{n \in \mathbf{Z}} a_n e^{inx}$ and $\phi(x) = \sum_{n \in \mathbf{Z}} b_n e^{inx}$. ♠

Example 1.3.2 Let f be a periodic function of period 2π and A_f be the operator of convolution with f , i.e.,

$$A_f : \psi \longrightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y)\psi(y)dy$$

Assume $f(x) = \sum_n f_n e^{inx}$ and $\psi(x) = \sum_n a_n e^{inx}$. Substituting the Fourier series for f and ψ in the definition of A_f we get

$$A_f(\psi) = \frac{1}{2\pi} \sum_{n,m} e^{imx} \int_{-\pi}^{\pi} f_n a_m e^{i(n-m)y} dy = \sum_m f_m a_m e^{imx}.$$

This means that in the diagonalization of the operator A_f , the set S is $\mathbf{Z}$, and the function φ_{A_f} is

$$\varphi_{A_f}(n) = f_n.$$

The underlying vector space and Plancherel measure is the same as in example 1.3.1. Thus Fourier series transforms convolutions into multiplication of Fourier coefficients. The fact that Fourier series simultaneously diagonalizes convolutions and differentiation reflects the fact that convolutions and differentiation commute and commuting diagonalizable matrices can be simultaneously diagonalized. Convolution operators occur frequently in many areas of mathematics and engineering. ♠

Example 1.3.3 Examples 1.3.1 and 1.3.2 for functions on $\mathbf{R}$ or $\mathbf{R}^n$ when the periodicity assumption is removed. For a function ψ of *compact support* (i.e., ψ vanishes outside a closed interval $[a, b]$) Fourier transform is defined by

$$\tilde{\psi}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} \psi(x) dx.$$

Integration by parts shows that under Fourier transform the operator of differentiation $\frac{d}{dx}$ becomes multiplication by $-i\lambda$:

$$\left(\widetilde{\frac{d\psi}{dx}} \right)(\lambda) = (-i\lambda) \tilde{\psi}(\lambda).$$

Similarly let A_f denotes the operator of convolution by an integrable function f :

$$A_f(\psi) = \int_{-\infty}^{\infty} f(x-y)\psi(y)dy.$$

Then a change of variable shows

$$\int_{-\infty}^{\infty} e^{-i\lambda x} \int_{-\infty}^{\infty} f(x-y)\psi(y)dydx = \tilde{f}(\lambda)\tilde{\psi}(\lambda).$$

Therefore the diagonalization process of the differentiation and convolution operators on functions on $\mathbf{R}$ leads to case (2) with $S = \mathbf{R}$, $\frac{d}{dx} \leftrightarrow -i\lambda$, and $A_f \leftrightarrow \tilde{f}$. For the underlying vector space it is convenient to start with the space of compactly supported functions on $\mathbf{R}$ and then try to extend the operators to $L^2(S)$. There are technical points which need clarification, but for the time being we are going to ignore them. ♠

Example 1.3.4 Let us apply the above considerations to the simple random walk on $\mathbf{Z}$. The random walk is described by convolution with the function f on $\mathbf{Z}$ defined by

$$f(n) = \begin{cases} p, & \text{if } n = 1; \\ q, & \text{if } n = -1; \\ 0, & \text{otherwise.} \end{cases}$$

Convolution on $\mathbf{Z}$ is defined similar to the cases on $\mathbf{R}$ except that the integral is replaced by a sum:

$$A_f(\psi) = f \star \psi(n) = \sum_{k \in \mathbf{Z}} f(n-k)\psi(k).$$

Let e_j be the function on $\mathbf{Z}$ defined by $e_j(n) = \delta_{jn}$ where δ_{jn} is 1 if $j = n$ and 0 otherwise. It is straightforward to see that the matrix of the operator A_f relative to the basis e_j for the vector space of functions on $\mathbf{Z}$ is the matrix of transition probabilities for the simple random walk on $\mathbf{Z}$. Example 1.3.2 suggests that this situation is dual to one described in that example. For S we take the interval $[-\pi, \pi]$, to state j corresponds the periodic function e^{ijx} and the action of the transition matrix P is given by multiplication by the function

$$pe^{ix} + qe^{-ix} = (p+q)\cos x + i(p-q)\sin x.$$

The probability of being in state 0 at time $2l$ is the therefore the constant term in the Fourier expansion of $(pe^{ix} + qe^{-ix})^{2l}$, viz.,

$$\binom{2l}{l} p^l q^l,$$

which we had easily established before. ♠

Example 1.3.5 A more interesting example is the application of the idea of the spectral theorem to the reflecting random walk on $\mathbf{Z}_+$ where the point 0 is a reflecting barrier. The matrix of transition probabilities is given by

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Assume the diagonalization can be implemented so that P becomes multiplication by the function $\varphi_P(x) = x$ on the space of functions on an interval which we take to be $[-1, 1]$. If the state n corresponds to the function ϕ_n then we must require

$$x\phi_0 = \phi_1, \quad x\phi_n(x) = \frac{1}{2}\phi_{n-1}(x) + \frac{1}{2}\phi_{n+1}(x).$$

Using the elementary trigonometric identity

$$\cos \alpha \cos \beta = \frac{1}{2} \cos(\alpha - \beta) + \frac{1}{2} \cos(\alpha + \beta)$$

we obtain the following functions ϕ_n :

$$\phi_0(x) = 1, \quad \phi_n(x) = \cos n\theta, \quad \text{where } \theta = \cos^{-1} x. \quad (1.3.1)$$

The polynomials $\phi_n(x) = \cos(n \cos^{-1} x)$ are generally called *Chebyshev polynomials*. The above discussion should serve as a good motivation for the introduction of these polynomials which found a number of applications. For the Plancherel measure we seek a function $\frac{1}{c(x)}$ such that

$$\int_{-1}^1 \phi_n(x) \phi_m(x) \frac{dx}{c(x)} = \delta_{mn}.$$

It is clear that we can take

$$\frac{1}{c(x)} = \frac{1}{\pi\sqrt{1-x^2}}.$$

The coefficients of the matrix P^n can be computed in terms of the Chebycheff polynomials. In fact it is straightforward to see that

$$P_{jk}^{(n)} = \int_{-1}^1 x^n \phi_j(x) \phi_k(x) \frac{dx}{c(x)}.$$

The idea of this simple example can be extended to considerably more complex Markov chains and the machinery of orthogonal polynomials can be used to this effect. ♠

In the next subsection we apply some ideas from Fourier analysis for the analysis of Markov chains.