

Stat 8931 Fall 2005 Class Notes

©2005 Charles J. Geyer

Some More Markov Chain Theory

Version of November 16, 2005

## Contents

<b>1</b>	<b>Markov Transition Matrices</b>	<b>2</b>
1.1	Eigenvectors Associated with Eigenvalue One . . . . .	2
1.1.1	Right Eigenvectors . . . . .	2
1.1.2	Left Eigenvectors . . . . .	3
1.1.3	Invariant Probability Vectors . . . . .	3
1.2	Communication Classes . . . . .	4
1.3	Irreducibility . . . . .	5
1.4	Transient States . . . . .	6
1.5	Recurrent States . . . . .	8
1.6	Closed Sets . . . . .	8
1.7	Indecomposability . . . . .	8
<b>2</b>	<b>Markov Transition Kernels</b>	<b>9</b>
2.1	General State Space . . . . .	10
2.2	Eigenvectors Associated with Eigenvalue One . . . . .	10
2.2.1	Right Multiplication . . . . .	10
2.2.2	Right Eigenvectors . . . . .	11
2.2.3	Signed Measures . . . . .	12
2.2.4	Left Multiplication . . . . .	13
2.2.5	Left Eigenvectors . . . . .	14
2.3	$\varphi$ -Irreducibility . . . . .	15
2.3.1	Occupation Times . . . . .	16
2.3.2	Return Times . . . . .	16
2.3.3	Randomly Subsampled Markov Chains . . . . .	17
2.3.4	Maximal Irreducibility Measures . . . . .	17
<b>3</b>	<b>Minorization, Splitting, and Regeneration</b>	<b>18</b>
3.1	Small Sets and Minorization . . . . .	19

# 1 Markov Transition Matrices

A *Markov transition matrix*, also called a *stochastic matrix* is a square matrix whose rows are *probability vectors*, meaning their entries are nonnegative and sum to one. If  $P$  is a Markov transition matrix, we denote its entries  $p(x, y)$ , rather than  $p_{ij}$ , where  $x$  and  $y$  range over a finite set  $S$  called the *state space*.

The Markov chain interpretation of  $P$  is that  $p(x, y)$  is the probability that the Markov chain is in state  $y$  at time  $t + 1$  given that it was in state  $x$  at time  $t$ . If  $X_1, X_2, \dots, X_\nu$  is a Markov chain with transition matrix  $P$ , then this means

$$p(x, y) = \Pr\{X_{t+1} = y | X_t = x\}, \quad t = 1, \dots, \nu - 1, \quad x, y \in S.$$

Thus  $P$  determines everything about the law of the Markov chain except the marginal distribution of  $X_1$ , which is called the *initial distribution* of the Markov chain.

## 1.1 Eigenvectors Associated with Eigenvalue One

### 1.1.1 Right Eigenvectors

Let  $e_1$  denote the column vector having all elements one. Then the requirement that each row of  $P$  sum to one can be written

$$Pe_1 = e_1 \tag{1}$$

which says that  $e_1$  is a right eigenvector of  $P$  associated with the eigenvalue one. There may be more right eigenvectors associated with the eigenvalue one, but there is always at least one.

Another way to write (1) is

$$(I - P)e_1 = 0 \tag{2}$$

where  $I$  denotes the identity matrix of the same dimension as  $P$ . This says that the set of right eigenvectors of  $P$  with eigenvalue one is the null space of the linear operator

$$f \mapsto (I - P)f \tag{3}$$

that is represented by the matrix  $I - P$ .

In general, a right eigenvector of  $P$  associated with the eigenvalue one, that is, an  $h$  such that  $Ph = h$  is called a *harmonic function*. In the finite state space case “harmonic vector” might be more appropriate, but this

terminology was developed in the theory of general Markov chains where  $h$  is a function on the state space, as it is, of course, in the finite dimensional case too, as we acknowledge when we write its coordinates  $h(x)$  and the equation  $Ph = h$  as

$$\sum_{y \in S} p(x, y)h(y) = h(x), \quad \text{for all } x \in S.$$

The linear operator (3) represented by the matrix  $I - P$  is called the *Laplacian operator*. It arises again in the general theory.

### 1.1.2 Left Eigenvectors

Since  $I - P$  does not have full rank, the null space of the linear operator

$$\nu \mapsto \nu(I - P)$$

that is represented by the matrix  $(I - P)^T$  is also not the zero subspace, since both left and right null spaces have the same dimension by a basic theorem of linear algebra Halmos (1958, Theorem 1 of Section 50).

Thus there also exists a vector  $\nu$  satisfying

$$\nu P = \nu$$

which is a left eigenvector of  $P$  associated with the eigenvalue one. It is a consequence of the Perron-Frobenius theorem that  $\nu$  can always be taken to be a probability vector.

### 1.1.3 Invariant Probability Vectors

A probability vector  $\pi$  is said to be invariant for a transition matrix  $P$  if

$$\pi P = \pi \tag{4}$$

In this terminology, the discussion in the preceding section establishes the following theorem (an immediate consequence of the Perron-Frobenius theorem).

**Theorem 1.1.** *Every Markov transition matrix has at least one invariant probability vector.*

We have seen that existence of invariant probability vectors is connected with the rank of  $I - P$ . If  $P$  is a Markov transition matrix, then we say that  $I - P$  has *maximal rank* if it has the maximum possible rank, which is one less than the dimension of the state space, because its null space must have dimension at least one. Thus  $I - P$  is never *full rank*, but can have *maximal rank*.

Thus  $I - P$  having maximal rank is a *sufficient* condition for uniqueness of the invariant probability vector. A question we still need to address is: is this condition also *necessary*?

## 1.2 Communication Classes

Define a relation

$$x \longrightarrow y \quad \text{if and only if} \quad p^n(x, y) > 0 \text{ for some } n$$

( $x \longrightarrow y$  is read “can get from  $x$  to  $y$ ”). This is a transitive relation, meaning

$$x \longrightarrow y \quad \text{and} \quad y \longrightarrow z \quad \text{implies} \quad x \longrightarrow z$$

because

$$p^m(x, y) > 0 \quad \text{and} \quad p^n(y, z) > 0 \quad \text{implies} \quad p^{m+n}(x, z) > 0.$$

The reflexive closure of this relation defined by

$$x \succeq y \quad \text{if and only if} \quad x = y \text{ or } x \longrightarrow y.$$

(Note that  $x \succeq x$  is true by definition and does not imply  $x \longrightarrow x$ .)

For each  $x \in S$  define

$$C_x = \{ y \in S : x \succeq y \text{ and } y \succeq x \}.$$

We call  $C_x$  the *communication class* of  $x$ . Observe that if  $z \in C_x$  then either  $x = z$  or we can get from  $x$  to  $z$  and vice versa and also from  $x$  to any other element of  $C_x$  (if any) and vice versa, hence  $C_z = C_x$ . Thus the communication class are either equal and disjoint and hence form a partition of the state space.

Now define

$$C_x \succeq C_y \quad \text{if and only if} \quad x \succeq y.$$

Now as a relation among the communication classes this is a reflexive, anti-symmetric, and transitive relation, reflexivity and transitivity being inherited from the relation among states and antisymmetry, that is

$$C_x \succeq C_y \quad \text{and} \quad C_y \succeq C_x \quad \text{implies} \quad C_x = C_y$$

coming from the definition of the communication classes. A reflexive, anti-symmetric, and transitive relation is also known as a *partial order* relation.

The irreflexive version of this relation is defined by

$$C_x \succ C_y \quad \text{if and only if} \quad C_x \succeq C_y \quad \text{and} \quad C_x \neq C_y.$$

Call  $C_x$  a *minimal element* of the family of communication classes if there does not exist a  $y$  such that  $C_x \succ C_y$ . Pick any  $C_{x_1}$ . Then either  $C_{x_1}$  is minimal or there exists  $x_2$  such that  $C_{x_1} \succ C_{x_2}$ . Then either  $C_{x_2}$  is minimal or we can repeat the process, obtaining a sequence

$$C_{x_1} \succ C_{x_2} \succ \cdots \succ C_{x_k}$$

By transitivity, no communication class can appear twice in the sequence. Since there are only a finite number of communication classes, the sequence must end somewhere, and the last element must be minimal.

A minimal communication class  $C_x$  has the property that the Markov chain never moves from inside to outside the class. Hence if we restrict  $P$  to this class, it is again a Markov transition matrix. Moreover, multistep transitions are possible from any point of this class to any other. In general, we do not know  $x \longrightarrow x$  but if  $C_x$  is a minimal class, we do, because either  $C_x$  contains some other point  $y$  and  $x \longrightarrow y \longrightarrow x$ , or  $C_x = \{x\}$  and since no move outside the class is possible, we then must have  $p(x, x) = 1$ .

Since movement outside a minimal communication class is impossible,  $P$  restricted to such a class is still a Markov transition matrix. If  $A$  is the minimal communication class, then

$$\sum_{y \in A} p(x, y) = 1, \quad x \in A$$

so  $P$  is a Markov transition matrix on  $A \times A$ . The property such a restricted  $P$  has is formalized in the following section.

### 1.3 Irreducibility

If  $P$  is a Markov transition matrix such that we have  $x \longrightarrow y$  and  $y \longrightarrow x$  for all  $x$  and  $y$  in the state space, then we say  $P$  is *irreducible*. Recall that this means for every  $x$  and  $y$  in the state space there exists an  $n$ , which may depend on  $x$  and  $y$ , such that  $p^n(x, y) > 0$ .

**Theorem 1.2.** *If a  $P$  is an irreducible Markov transition matrix having invariant probability vector  $\pi$ , then  $\pi(x) > 0$  for all  $x$ .*

We abbreviate the conclusion of the theorem as  $\pi > 0$ .

*Proof.* Suppose  $\pi(y) = 0$ . Note that  $\pi P = \pi$  implies  $\pi P^n = \pi$  for all  $n$ . This implies

$$\sum_{x \in S} \pi(x) p^n(x, y) = 0, \quad \text{for all } n.$$

Now we have  $\pi(x) > 0$  for some  $x$ . Hence for this  $x$  we have

$$p^n(x, y) = 0, \quad \text{for all } n.$$

But this implies  $x \not\leftrightarrow y$ , which contradicts irreducibility. Hence  $\pi(y) = 0$  is impossible.  $\square$

**Theorem 1.3.** *If a  $P$  is an irreducible Markov transition matrix, then  $I - P$  has maximal rank, and  $P$  has exactly one invariant probability vector.*

*Proof.* One invariant probability vector  $\pi$  is guaranteed by the Perron-Frobenius theorem. If  $I - P$  does not have maximal rank, then there is a nonzero vector  $\nu$  satisfying  $\nu P = \nu$ , which does not satisfy  $\nu = t\pi$  for some real  $t$ . Now  $\nu$  need not have all components the same sign, but  $\pi + \epsilon\nu$  is a probability vector for sufficiently all small positive or negative  $\epsilon$ . Define

$$\begin{aligned} \epsilon^+ &= \sup\{\epsilon > 0 : \pi + \epsilon\nu > 0\} \\ \epsilon^- &= \inf\{\epsilon < 0 : \pi + \epsilon\nu > 0\} \end{aligned}$$

At least one of these is finite because  $\nu \neq 0$ . But then either  $\pi + \epsilon^+\nu$  or  $\pi + \epsilon^-\nu$  has some component zero, and that contradicts irreducibility. Hence there exists no such  $\nu$  and  $I - P$  has maximal rank, which implies  $\pi$  is the unique invariant probability vector.  $\square$

## 1.4 Transient States

Let us now consider a maximal communicating class, a  $C_x$  such that there does not exist a  $C_y$  such that  $C_y \succ C_x$ . Again, such a class exists by finiteness of the state space. Now we have no probability of entering the class from outside. Calling the maximal class  $A$ , we have

$$p(x, y) = 0, \quad x \notin A, y \in A.$$

If  $\pi$  is an invariant probability vector, this implies

$$\sum_{x \in A} \pi(x) p(x, y) = \pi(y) \tag{5}$$

and summing over  $y$  gives

$$\sum_{x \in A} \pi(x) \sum_{y \in A} p(x, y) = \sum_{y \in A} \pi(y)$$

and this implies

$$\sum_{y \in A} p(x, y) = 1, \quad x \in A, y \in A \quad (6a)$$

or

$$\sum_{y \in A} \pi(y) = 0. \quad (6b)$$

If (6a) holds, then  $A$  is not only but minimal and the Markov chain restricted to  $A$  is irreducible. Otherwise, (6b) holds, and we must have  $\pi(x) = 0$  for all  $x \in A$ .

Now we proceed by induction on the strict partial order  $\succ$ . Suppose (the induction hypothesis) we have

$$C_x \succ C_y \quad \text{implies} \quad \pi(z) = 0, \quad z \in C_x$$

Fix  $y$ . Write

$$U = \bigcup \{ C_x : C_x \succ C_y \}$$

and  $A = C_y$ . Then

$$p(x, y) = 0, \quad x \notin U \cup A, y \in A.$$

If  $\pi$  is an invariant probability vector, this implies

$$\sum_{x \in U \cup A} \pi(x) p(x, y) = \pi(y).$$

Now by the induction hypothesis we have  $\pi(z) = 0$  for  $z \in U$  so this reduces to (5). So we get the same conclusion about  $C_y$ : either it is minimal or we also have  $\pi(z) = 0$  for  $z \in C_y$ .

We say that any communicating class that is not minimal is *transient* and also that all of the states it contains are *transient*. The calculations above prove the following theorem.

**Theorem 1.4.** *An invariant probability vector gives probability zero to any transient state.*

## 1.5 Recurrent States

We have arrived at an interesting and important dichotomy. A state in a minimal communication class is said to be *recurrent*. We have just defined a state in a non-minimal communication class to be *transient*. Thus every state is either transient or recurrent.

Theorem 1.4 says  $\pi(x) = 0$  for every transient  $x$  and every invariant probability vector  $\pi$ . Theorem 1.2 says  $\pi(x) > 0$  for every recurrent  $x$  and every invariant probability vector  $\pi$ .

## 1.6 Closed Sets

A set out of which no transitions are possible is called *closed*. Clearly a communication class is closed if and only if it is minimal (“closed set” is standard Markov chain terminology but “minimal communication class” is not).

If  $A$  is a closed set and  $x \in A$  then we must have  $C_y \subset A$  whenever  $C_x \succeq C_y$ . So every closed set is a union of communication classes and contains at least one minimal communication class.

## 1.7 Indecomposability

Let us rearrange the states of our Markov chain putting all of the communication classes together and treating them as blocks but lumping all of the transient states in one block. Then the transition matrix has the form

$$P = \begin{pmatrix} P_{11} & 0 & \cdots & 0 & 0 \\ 0 & P_{22} & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & P_{k-1,k-1} & 0 \\ P_{k1} & P_{k2} & \cdots & P_{k,k-1} & P_{kk} \end{pmatrix}$$

All of the diagonal blocks except the last, being irreducible, have unique invariant probability vectors; call them  $\pi_1, \dots, \pi_{k-1}$ . It is clear that the full dimensional vector  $(\pi_1 \ 0 \ \cdots \ 0 \ 0)$  is an invariant probability vector of  $P$ . So is  $(0 \ \pi_2 \ \cdots \ 0 \ 0)$ , and so forth.

Now consider an arbitrary left eigenvector of  $P$  associated with eigenvalue one

$$(\nu_1 \ \nu_2 \ \cdots \ \nu_{k-1} \ \nu_k). \tag{7}$$

We have

$$\nu_j P_{jj} + \nu_k P_{kj} = \nu_j, \quad j = 1, \dots, k-1 \tag{8a}$$



and

$$\nu_k P_{kk} = \nu_k. \tag{8b}$$

We already know from Theorem 1.4 that if (7) is required to be a probability vector, then the only solution of (8b) is  $\nu_k = 0$  because all of the states in the  $k$ -th block are transient. Plugging that back in to (8a) we see that  $\nu_j$  must be proportional to the unique invariant probability vector  $\pi_j$  for  $P_{jj}$ .

Thus we see that every invariant probability vector for  $P$  has the form

$$(t_1\pi_1 \quad t_2\pi_2 \quad \cdots \quad t_{k-1}\pi_{k-1} \quad 0),$$

where  $t_1, \dots, t_{k-1}$  are nonnegative real numbers that sum to one, and if the  $k$ -th block is empty (there are no transient states), then every left eigenvector associated with eigenvalue one has the form

$$(t_1\pi_1 \quad t_2\pi_2 \quad \cdots \quad t_{k-1}\pi_{k-1}),$$

where  $t_1, \dots, t_{k-1}$  are arbitrary real numbers. (A transient block can have multiple left eigenvectors that are not proportional to probability vectors. Theorem 1.4 only says it has no left eigenvectors that are proportional to probability vectors.)

A transition probability matrix with only one minimal communication class, which is the same thing as saying there do not exist two disjoint closed sets, is said to be *indecomposable*.

**Theorem 1.5.** *A Markov transition matrix is indecomposable if and only if it has a unique invariant probability vector.*

## 2 Markov Transition Kernels

A *Markov transition kernel*, on a measurable space  $(S, \mathcal{S})$  is a function from  $P : (S, \mathcal{S}) \rightarrow \mathbb{R}$  such that

- $x \mapsto P(x, A)$  is a measurable function for each fixed  $A \in \mathcal{S}$ .
- $A \mapsto P(x, A)$  is a probability measure for each fixed  $x \in S$ .

Here we are using real measure theory with  $\mathcal{S}$  a sigma-algebra of subsets of  $S$ .

The Markov chain interpretation of  $P$  is that  $p(x, A)$  is the probability that the Markov chain is in  $A$  at time  $t + 1$  given that it was in state  $x$  at

time  $t$ . If  $X_1, X_2, \dots, X_\nu$  is a Markov chain with transition matrix  $P$ , then this means

$$P(x, A) = \Pr\{X_{t+1} \in A | X_t = x\}, \quad t = 1, \dots, \nu - 1, \quad x, y \in S.$$

Thus  $P$  determines everything about the law of the Markov chain except the marginal distribution of  $X_1$ , which is called the *initial distribution* of the Markov chain.

## 2.1 General State Space

General state space Markov chains were until fairly recently (about 1980) considered so difficult that there was no satisfactory general theory. In fact, they were so difficult that most books with “Markov chain” in the title considered state spaces that were at most countable.

By “general” we do not mean completely general (as with many uses of “general” it only means more general than the previously widely presented theory). However the only restriction is that the sigma-algebra, the  $\mathcal{S}$  in  $(S, \mathcal{S})$  be countably generated, meaning there is a countable subset  $\mathcal{B}$  such that  $\mathcal{S}$  is the smallest sigma-algebra containing  $\mathcal{B}$ .

Recall that any separable metric space has a countable basis of the topology. For example,  $\mathbb{R}^d$  has a basis consisting of all balls whose centers have rational coordinates and whose radii are also rational (because the rational numbers are countable and because the cartesian product of countable sets is countable). This family also generates the Borel sigma-algebra for  $\mathbb{R}^d$ .

It is fair to say that any probability space for which one might find a real application will be countably generated. So this “general” despite not being completely general, is very “general” indeed.

## 2.2 Eigenvectors Associated with Eigenvalue One

### 2.2.1 Right Multiplication

If  $f$  is a measurable function on  $(S, \mathcal{S})$ , then we interpret the “right multiplication”  $g = Pf$  to mean

$$g(x) = \int P(x, dy)f(y) \tag{9}$$

assuming the integral makes sense (for example, it always does when  $f$  is bounded). The Markov chain interpretation of this right multiplication is

$$g(X_n) = E\{f(X_{n+1})|X_n\}.$$

Using the conditional Jensen inequality we have for any  $p \geq 1$ .

$$E\{|g(X_n)|^p\} \leq E[E\{|f(X_{n+1})|^p | X_n\}] = E\{|f(X_{n+1})|^p\} \quad (10)$$

Introducing the  $L^p(\pi)$  seminorms

$$\|f\|_{L^p(\pi)}^p = \int |f(x)|^p \pi(dx) \quad (11)$$

we see when the chain is stationary with invariant measure  $\pi$  that (10) has the interpretation

$$\|Pf\|_{L^p(\pi)} \leq \|f\|_{L^p(\pi)}, \quad f \in L^p(\pi) \quad (12)$$

where  $L^p(\pi)$  denotes the set of measurable functions  $f$  such that  $\|f\|_{L^p(\pi)}$  is finite.

Thus, assuming  $P$  has an invariant probability measure  $\pi$  (it need not have one), we can consider  $P$  as the map  $f \mapsto Pf$  which maps  $L^p(\pi) \rightarrow L^p(\pi)$ . This makes  $P$  a linear operator on an infinite-dimensional vector space (in general).

For many applications, this is good enough. If one really needs to use functional analysis (the fancy name for infinite-dimensional linear algebra), then one redefines  $L^p(\pi)$ , changing the definition only slightly. The problem is that (11) does not define a norm, because  $\|f\|_{L^p(\pi)}^p = 0$  does not imply  $f = 0$  but only  $f = 0$  almost everywhere  $[\pi]$ . Thus we do the usual thing of “modding out the equivalence classes” and consider  $L^p(\pi)$  the set of all equivalence classes of functions  $f$  having finite  $L^p(\pi)$  norm, where the equivalence is that of being equal almost everywhere  $[\pi]$ . This makes  $L^p(\pi)$  a Banach space, (a complete normed vector space), a concept about which functional analysis books have a lot to say (see, for example, Rudin, 1987, pp. 66-67 and 95).

Another Banach space (also covered in Rudin, 1987, pp. 66-67) is  $L^\infty(\pi)$  which is the space whose points are equivalence class of functions that are equal almost everywhere  $[\pi]$  to a bounded function, the least bound being the  $L^\infty(\pi)$  norm. One easily checks that (11) then also holds for  $p = \infty$ . Thus  $P$  can be considered a “right multiplication” operator  $L^p(\pi) \rightarrow L^p(\pi)$  for any  $1 \leq p \leq \infty$ .

### 2.2.2 Right Eigenvectors

Let  $e_1$  denote the constant function  $S \rightarrow \mathbb{R}$  whose constant value is one. Then  $e_1 \in L^p(\pi)$  for any  $1 \leq p \leq \infty$ , or more precisely (a pedantry that

most writers do not repeat after the first occurrence or two) the equivalence class of functions that are equal to one almost everywhere  $[\pi]$  is an element of  $L^p(\pi)$ .

Then the requirement that each  $A \mapsto P(x, A)$  be a probability measure can be written

$$Pe_1 = e_1 \tag{13}$$

which says that  $e_1$  is a right eigenvector of  $P$  associated with the eigenvalue one. There may be more right eigenvectors associated with the eigenvalue one, but there is always at least one.

Another way to write (13) is

$$(I - P)e_1 = 0 \tag{14}$$

where  $I$  denotes the identity operator on the same space, some  $L^p(\pi)$ , that we are considering  $P$  an operator on.

This says that the set of right eigenvectors of  $P$  with eigenvalue one is the null space of the linear operator  $I - P$ , considered as a “right multiplication”

$$f \mapsto (I - P)f. \tag{15}$$

As we noted in the finite state space case, a right eigenvector of  $P$  associated with the eigenvalue one, that is, an  $h$  such that  $Ph = h$  is called a *harmonic function*. The linear operator (15) represented by the kernel  $I - P$  is called the *Laplacian operator*.

### 2.2.3 Signed Measures

A *signed measure* on  $(S, \mathcal{S})$  is a map  $\mu : \mathcal{S} \rightarrow \mathbb{R}$  that is countably additive. Unlike measures that are studied elsewhere in measure theory, note that the values  $+\infty$  and  $-\infty$  are not allowed. This limitation has strong consequences.

For each  $A \in \mathcal{S}$  define

$$|\mu|(A) = \sup \sum_{i=1}^{\infty} |\mu(A_i)|$$

the supremum being taken over all countable, measurable partitions  $A_1, A_2, \dots$  of  $A$ . It turns out that  $A \mapsto |\mu|(A)$  is countably additive (Rudin (1987, Theorem 6.2) hence a *positive measure* which is denoted  $|\mu|$  and that  $|\mu|(S)$  is finite. Hence  $|\mu|$  is also real-valued (cannot be infinite) and is a signed measure as defined above.

Let  $\mathfrak{M}(S)$  denote the space of all signed measures on  $(S, \mathcal{S})$ . This is a Banach space when total variation measure is used at the norm, that is

$$\|\mu\|_{\text{TV}} = |\mu|(S).$$

If  $\mu$  is a signed measure, we write

$$\begin{aligned}\mu^+(A) &= \frac{|\mu|(A) + \mu(A)}{2} \\ \mu^-(A) &= \frac{|\mu|(A) - \mu(A)}{2}\end{aligned}$$

also define positive measures that are elements of  $\mathfrak{M}(S)$ . They are called the positive and negative parts of  $\mu$  and we can write

$$\begin{aligned}|\mu| &= \mu^+ + \mu^- \\ \mu &= \mu^+ - \mu^-\end{aligned}$$

Since both  $\mu^+$  and  $\mu^-$  are finite, they are renormalizable to probability measures if nonzero. Call these measures  $\pi^+$  and  $\pi^-$ . Then we can write

$$\mu = a^+ \pi^+ - a^- \pi^- \tag{16}$$

where  $a^+$  and  $a^-$  are nonnegative real numbers. Thus we see that every signed measure is a linear combination of probability measures. The main reason for introducing signed measures in probability theory is to deal with linear combinations of probability measures in a satisfactory way.

#### 2.2.4 Left Multiplication

If  $\mu$  is a signed measure on  $(S, \mathcal{S})$ , then we interpret the “left multiplication”  $\nu = \mu P$  to mean

$$\nu(A) = \int \mu(dx) P(x, A). \tag{17}$$

If  $\mu$  has the form (16), then (17) becomes

$$\nu(A) = a^+ \int \pi^+(dx) P(x, A) - a^- \int \pi^-(dx) P(x, A). \tag{18}$$

(for those uncomfortable with integrating w. r. t. a general signed measure).

The Markov chain interpretation of the right multiplication  $\pi_2 = \pi_1 P$  is that if  $\pi_1$  is the (marginal) distribution of  $X_1$ , then  $\pi_2$  is the (marginal) distribution of  $X_2$ .

### 2.2.5 Left Eigenvectors

Infinite-dimensional vector spaces are just like finite-dimensional vector spaces except for the ways they are different (how tautological!) One of the ways in which infinite-dimensional spaces are different is the following.

A linear operator  $T$  on a finite-dimensional vector space  $V$  is invertible if and only if any one of the following conditions holds.

- (a) It is injective (its null space is the zero subspace).
- (b) It is surjective (its range is  $V$ ).
- (c) The matrix representing  $T$  has full rank, which is the dimension of  $V$ .

(this is equivalent to the right and left null spaces having the same dimension used in constructing an invariant probability vector).

In infinite-dimensional spaces, we have only the general condition. Any function (linear or not) is invertible if and only if it is injective and surjective (one-to-one and onto). Thus we can say the following.

A linear operator  $T$  on an infinite-dimensional vector space  $V$  is invertible if and only if both of the following conditions hold.

- (a) It is injective (its null space is the zero subspace).
- (b) It is surjective (its range is  $V$ ).

In summary, in the finite-dimensional case, it is enough that either (a) or (b) hold, because each implies the other, but in the infinite-dimensional case, both (a) and (b) must hold, because neither implies the other. Moreover, the concept of rank makes no sense (in general) for operators on an infinite-dimensional space.

We are left with no tools that (in general) guarantee the existence of an invariant probability measure. Here is a counterexample, that shows there can be no proof like that in the finite-dimensional case.

Consider a Markov chain with state space the nonnegative integers that moves deterministically

$$X_{n+1} = X_n + 1$$

If  $\nu$  is an invariant measure (never mind probability measure), then we must have

$$\nu(\{x\}) = \nu(\{x-1\}), \quad x > 0$$

and

$$\nu(\{0\}) = 0.$$

But together these imply  $\nu = 0$ , the zero measure that attributes measure zero to all sets, and zero never counts as an eigenvector.

### 2.3 $\varphi$ -Irreducibility

The notions of indecomposability and irreducibility that we used to analyze finite state space Markov chains are of little use for general state space Markov chains.

Indecomposability still makes sense. A subset  $A$  of the state space is *closed* if

$$P(x, A) = 1, \quad x \in A,$$

and  $P$  is *indecomposable* if there do not exist two disjoint closed sets.

Irreducibility makes no sense, in general. Consider a Markov chain with each random variable continuous, for example an AR(1) time series. The probability of going from any point to any point is zero, because all points in the state space have probability zero at all times. As we shall see, an AR(1) time series is about as well behaved as a Markov chain on a general state space can be. But the irreducibility concept from finite state space theory is useless for telling us anything about it.

Clearly, we need to consider probabilities of hitting *sets* rather than points. Hence the following definition. Let  $\varphi$  be a positive measure on  $(S, \mathcal{S})$  that is not the zero measure ( $\varphi$  is not necessarily real-valued, the value  $+\infty$  is allowed). The Markov kernel  $P$  is  $\varphi$ -irreducible if for every  $x$  and every  $A$  such that  $\varphi(A) > 0$  there exists an  $n$ , which may depend on  $x$  and  $A$ , such that  $P^n(x, A) > 0$ .

The only thing that matters for the definition is whether  $\varphi(A)$  is zero or nonzero. When  $\varphi(A) > 0$  we have no interest in the actual value of  $\varphi(A)$ . Thus  $\varphi$  is only being used to specify a class of non-null sets that must be checked.

When specialized to finite state spaces,  $\varphi$ -irreducibility corresponds to *indecomposability* not irreducibility. Consider  $\varphi$  concentrated at the point  $y$ . Then  $\varphi$ -irreducibility implies  $x \rightarrow y$  for all  $x \in S$ , but it does not imply irreducibility. Consider for example

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

This  $P$  is  $\varphi$ -irreducible, because it can get from any state to state 2. But it is not irreducible, because it cannot get from state 2 to the other state.

**Theorem 2.1.**  *$\varphi$ -irreducibility implies indecomposability.*

*Proof.* Suppose  $P$  is decomposable, so there exist disjoint closed sets  $A$  and  $B$ , and let  $C$  be a set such that  $\varphi(C) > 0$ . Now one of  $C \cap A$  or  $C \setminus A$  has positive  $\varphi$  measure. If  $\varphi(C \cap A) > 0$ , then  $P^n(x, C \cap A) = 0$  for  $x$  in  $B$  and  $P$  is not irreducible. If  $\varphi(C \setminus A) > 0$ , then  $P^n(x, C \setminus A) = 0$  for  $x$  in  $A$  and  $P$  is again not irreducible.  $\square$

### 2.3.1 Occupation Times

Let  $X_0, X_1, \dots$ , be a Markov chain with transition kernel  $P$  that is started at  $x$ , that is,  $X_0 = x$  with probability one.

For any set  $A$ , the *occupation time*  $\eta_A$  is the number of visits the Markov chain makes to  $A$ , that is,

$$\eta_A = \sum_{i=1}^{\infty} I_A(X_i)$$

The occupation time is a random variable. Its expectation is a kernel (not a Markov kernel, but a general kernel)

$$U(x, A) = E\{\eta_A\} = \sum_{i=1}^{\infty} P^i(x, A)$$

Observe that  $P$  is  $\varphi$ -irreducible if and only if for every  $x \in S$  and every  $A \in \mathcal{S}$  such that  $\varphi(A) > 0$  we have  $U(x, A) > 0$ . (There is no need to check higher powers  $U^n$ , because  $U$  already contains all powers of  $P$ .)

### 2.3.2 Return Times

For any set  $A$ , the *first return time*  $\tau_A$  is the time of the first visits the Markov chain makes to  $A$  after time zero, that is,

$$\tau_A = \min\{i \geq 1 : I_A(X_i) = 1\}$$

where the min is defined to be infinity if the set is empty. The first return time is a random variable. The probability it is finite is a kernel (not a Markov kernel, but a general kernel)

$$L(x, A) = \Pr\{\tau_A < \infty\}$$

Meyn and Tweedie (1993, p. 72) give a complicated explicit formula for  $L(x, A)$ .



Observe that  $P$  is  $\varphi$ -irreducible if and only if for every  $x \in S$  and every  $A \in \mathcal{S}$  such that  $\varphi(A) > 0$  we have  $L(x, A) > 0$ . Since we do not have an explicit formula for  $L(x, A)$ , this must be argued abstractly, but it is clear that we are not  $\varphi$ -irreducible, then there exist  $x$  and  $A$  with  $\varphi(A) > 0$  such that the chain started at  $x$  never enters  $A$  with probability one. Hence with probability one we have  $\tau_A = \infty$  and hence  $L(x, A) = 0$ . And vice versa.

### 2.3.3 Randomly Subsampled Markov Chains

For various reasons, randomly subsampled Markov chains are an important technical tool, used extensively by Meyn and Tweedie (1993). Practically, this is a “combining method” that combines mixing and composition. The compositions in question are  $P^n$ ,  $n \geq 0$ , where  $P^0 = I$ . One step of the Markov chain chooses a random  $n$  (independent of the current state) and then executes the mechanism described by  $P^n$ , which is the same thing as running  $n$  iterations of the mechanism described by  $P$ .

When  $n$  has a geometric distribution with success probability  $\epsilon$  which is neither zero or one, the kernel of the randomly subsampled chain is

$$K_\epsilon(x, A) = (1 - \epsilon) \sum_{n=0}^{\infty} \epsilon^n P^n(x, A)$$

Observe that  $P$  is  $\varphi$ -irreducible if and only if for every  $x \in S$  and every  $A \in \mathcal{S}$  such that  $\varphi(A) > 0$  we have  $K_\epsilon(x, A) > 0$ . (There is no need to check higher powers  $K_\epsilon^n$ , because  $K_\epsilon$  already contains all powers of  $P$ .)

Meyn and Tweedie (1993) call  $K_\epsilon(x, A)$  the *resolvent*, a term from functional analysis. Using the geometric series we have

$$K_\epsilon = (1 - \epsilon)(I - \epsilon P)^{-1}$$

when the geometric series converges in an operator sense, which is always does when  $0 < \epsilon < 1$  and  $P$  is considered an operator on  $L^2(\pi)$  where  $\pi$  is invariant for  $P$ . The resolvent in functional analysis is the map

$$\lambda \mapsto (\lambda I - P)^{-1}$$

defined on the set of complex numbers  $\lambda$  for which the inverse exists (which is always a nonempty open set in the complex plane).

### 2.3.4 Maximal Irreducibility Measures

The following is Proposition 4.2.2 in Meyn and Tweedie (1993). Recall that if  $\mu$  and  $\nu$  are positive measures on the same measurable space we say  $\nu$

dominates  $\mu$  or  $\mu$  is absolutely continuous with respect to  $\nu$  and write  $\mu \ll \nu$  if  $\nu(A) = 0$  implies  $\mu(A) = 0$ .

**Theorem 2.2.** *If  $P$  is  $\varphi$ -irreducible for some  $\varphi$ , then there exists a probability measure  $\psi$  such that*

- (a)  $P$  is  $\psi$ -irreducible.
- (b) If  $P$  is  $\varphi'$ -irreducible, then  $\varphi' \ll \psi$ .
- (c) If  $\psi(A) = 0$  and  $B = \{y : L(y, A) > 0\}$ , then  $\psi(B) = 0$ .
- (d) If  $P$  is  $\varphi'$ -irreducible and

$$\psi'(A) = \int \varphi'(dx) K_\epsilon(x, A),$$

then  $\psi' \ll \psi$  and  $\psi \ll \psi'$ .

Any  $\psi$  satisfying the conditions of the theorem is called a *maximal irreducibility measure* for  $P$ . (If  $P$  is  $\varphi$ -irreducible then  $\varphi$  is a, not necessarily maximal, irreducibility measure for  $P$ ).

Maximal irreducibility measures are not unique, but consider that the purpose of irreducibility measures is characterizing a certain class of non-null sets (or of null sets). From (b) of the theorem it is clear that the class of non-null sets determined by a maximal irreducibility measure is unique and is the largest class of non-null sets determined by any irreducibility measure.

In what is very bad notation Meyn and Tweedie (1993) just say  $\psi$ -irreducible to mean  $\psi$  is a *maximal* irreducibility measure for  $P$  and say  $\varphi$ -irreducible to mean  $\psi$  is a not necessarily maximal irreducibility measure for  $P$ , thus making  $\varphi$  and  $\psi$  “frozen letters” in this context. We shall try not to follow their example and always say that  $\psi$  is maximal when that is what we mean.

### 3 Minorization, Splitting, and Regeneration

This section introduces the notions at the heart of the modern theory of Markov chains on general state spaces. These notions and constructions changed everything, making general state space chains no more difficult than countable state space chains. We do not attempt proofs, which are very deep.

It is understood throughout the discussion that  $P$  is a  $\varphi$ -irreducible Markov kernel, and the state space is, as usual,  $(S, \mathcal{S})$ .

### 3.1 Small Sets and Minorization

A subset  $C$  of the state space is said to be *small* if there exists a nonzero measure  $\nu$  and a positive integer  $m$  such that

$$P^m(x, A) \geq \nu(A), \quad x \in C, A \in \mathcal{S}. \quad (19)$$

The following theorem, originally due to Jain and Jameson is Theorem 5.2.2 in Meyn and Tweedie (1993).

**Theorem 3.1.** *Suppose  $P$  is a  $\psi$ -irreducible Markov kernel. Then for every  $B$  such that  $\psi(B) > 0$ , there exists a  $C \subset B$ , such that  $\psi(C) > 0$ , an  $m \geq 1$ , and a nonzero sigma-finite measure  $\nu$  such that (19) holds and, moreover,  $\nu(C) > 0$ .*

The theorem says at least one small set exists. It is fairly easily shown that the existence of one implies many more. By (Meyn and Tweedie, 1993, Theorem 5.2.4) there exists a countable collection of small sets cover the state space.

The fact that small sets involve powers of  $P$  is a nuisance. It would be much easier if we could assume the following, which is called the *minorization condition*. For some  $\delta > 0$ , and some  $C$  such that  $\psi(C) > 0$  and some probability measure  $\nu$  on the state space such that  $\nu(C) = 1$

$$P(x, A) \geq \delta\nu(A), \quad x \in C, A \in \mathcal{S} \quad (20)$$

holds. Clearly  $C$  is a small set when (20) holds.

Although it is not true that (20) holds for a general  $\psi$ -irreducible  $P$ . The minorization condition does hold for the corresponding resolvent kernels  $K_\epsilon$ . (Meyn and Tweedie, 1993, Proposition 5.2.3). This fact accounts for the intensive use of resolvent kernels in Meyn and Tweedie (1993). The resolvent kernel satisfies the minorization condition, which can be used to prove facts about the subsampled chain. Then these facts are transferred to the original chain with kernel  $P$  by some argument or other.

## References

- Halmos, P. R. (1958). *Finite-Dimensional Vector Spaces*. Springer-Verlag. Originally published by Litton Educational Publishing.
- Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.
- Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, 3rd edition.