E-mail: yms@statslab.cam.ac.uk

MARKOV CHAINS

1. Definition and basic facts

Definitions and basic properties, the transition matrix. Calculation of n-step transition probabilities.

This course is a logical continuation of IA Probability. We will study a class of *random processes* describing a wide variety of systems of theoretical and practical interest (and sometimes simply amusing). This class is called Markov chains, after the Russian mathematician A. Markov (1856–1922) who introduced and developed this elegant concept in the 1910's, 25 years before the notion of probability was shaped (in the 1930's) in the manner we use it today. This shows that a deep insight into the subject is possible without using sophisticated mathematical tools. It may also be an explanation why Markov chains are popular in so many different disciplines that are seemingly far from pure mathematics.

The basic model for the course will be a system that changes its state in *discrete* time, according to some random mechanism. The collection of states is called a state-space and throughout the whole course will be assumed finite or countable; we will denote it by I. Each $i \in I$ is called a state; our system will be always in one of these states. Sometimes we will know what state the system occupies and sometimes we will only know that the system is in state i with some probability. Therefore it makes sense to introduce a *probability measure* or *probability distribution* (shortly, a distribution) on I. A probability measure λ on I is simply a collection $(\lambda_i, i \in I)$ of non-negative numbers of total sum one:

$$\lambda_i \geq 0, \ \sum_{i \in I} \lambda_i = 1. \tag{1.1}$$

We can think of a unit 'mass' spread over set I where point i has mass λ_i. For that reason it is sometimes convenient to speak of a probability mass

function $i \in I \mapsto \lambda_i$. Then the probability of a set $J \subseteq I$ is $\lambda(J) = \sum_{i \in J} \lambda_j$.

If $\lambda_i = 1$ for some $i \in I$ and $\lambda_j = 0$ when $j \neq i$, the distribution is 'concentrated' at point i. Then the state of our system becomes 'deterministic'. We will denote such a distribution by δ_i.

Sometimes the condition $\sum_{i \in I} \lambda_i = 1$ is not fulfilled; then we simply say that λ is a *measure* on I. If the total mass $\sum_{i \in I} \lambda_i < \infty$, the measure is called finite and can be transformed into a probability distribution by the normalisation: $\widetilde{\lambda}_i = \lambda_i \Big/ \sum_{j \in I} \lambda_j$ will define a probability measure on I, as $\sum_{i \in I} \widetilde{\lambda}_i = \sum_{i \in I} \lambda_i \Big/ \sum_{j \in I} \lambda_i = 1$. But even if $\sum_{i \in I} \lambda_i = \infty$ (i.e., the total mass is infinite), we still can assign a finite value $\lambda(J) = \sum_{i \in J} \lambda_i$ to finite subsets $J \subset I$.

The random mechanism that causes the change of the state is described by a *transition matrix* P, with entries p_{ij}, $i, j \in I$. Entry p_{ij} gives the probability that the system will change state i to j in a unit of time. That is, p_{ij} is the conditional probability that the system will occupy state j at the next time given that it is currently in state i. Hence, we have that each entry of P is non-negative but not greater than 1, and the sum of entries along every row equals 1:

$$0 \leq p_{ij} \leq 1 \ \forall \, i, j \in I \text{ and } \sum_{j \in I} p_{ij} = 1 \ \forall \, i \in I. \tag{1.2}$$

Matrix P with these properties is called *stochastic*. By analogy, a probability distribution (λ_i) on I is often called a stochastic vector. Then a stochastic matrix is the one where every row is a stochastic vector.

Examples. 1.1. The simplest case is 2×2 (a two-state space). Without loss of generality, we may think that the states are 0 and 1, then the entries will be p_{ij}, $i, j = 0, 1$. Here, the stochastic matrix has the form

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

where $0 \leq \alpha, \beta \leq 1$. In particular, $\alpha = \beta = 0$ gives a unit matrix and $\alpha = \beta = 1$ the anti-diagonal matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The system with the unit matrix stays in the initial state forever; in the anti-diagonal case it changes its state every time, from 0 to 1 and vice versa.

On the other hand, $\alpha = \beta = 1/2$ gives the matrix

$$\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

In this case the system may keep its state or change it with probability $1/2$.

It is convenient to represent the transition matrix with a diagram where arrows show possible transitions and are labelled with the corresponding transition probabilities (arrows leading back to their own origin are often omitted as well as labels for deterministic transitions). See Diagram 1.1.

1.2. The 4×4 matrix

$$\begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is represented on Diagram 1.2.

The time will take values $n = 0, 1, 2, \ldots$. To complete the picture, we have to specify in what state our system is at the initial time $n = 0$. Typically, we will assume that the system at time $n = 0$ is in state i with probability λ_i for some given 'initial' distribution λ on I.

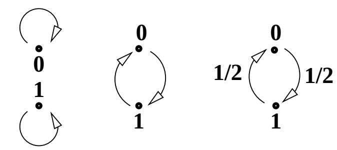

Diagram 1.1.

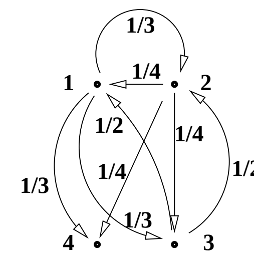

Diagram 1.2.

Denote by X_n the state of our system at time n. The rules specifying a Markov chain with initial distribution λ and transition matrix P are that (i) X_0 has distribution λ:

$$\mathbb{P}(X_0 = i) = \lambda_i, \ \forall i \in I,$$

(ii) more generally, $\forall\ n$ and $i_0, \ldots, i_n \in I$, the probabilities $\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)$ that the system occupies states i_0, i_1, ..., i_n at times 0, 1, ..., n is written as a product

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \tag{1.3}$$

Of course, (i) is a particular case of (ii), with $n = 0$.

An important consequence of (1.3) is the equation for the conditional probability $\mathbb{P}(X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i)$ that the state at time

$n+1$ is j, given states i_0, ..., i_{n-1} and $i_n = i$ at times 0, ..., $n-1$, n:

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i)$$
$$= \frac{\mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i, X_{n+1} = j)}{\mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i)} \tag{1.4}$$
$$= \frac{\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij}}{\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i}} = p_{ij}.$$

That is, conditional on $X_0 = i_0$, ..., $X_{n-1} = i_{n-1}$ and $X_n = i$, X_{n+1} has the distribution $(p_{ij}, j \in I)$. In particular, the conditional distribution of X_{n+1} does not depend on i_0, ..., i_{n-1}, i.e. depends only on the state i at the last preceding time n.

Formula (1.4) illustrates the memoryless property of a Markov chain.

Another consequence of (1.3) is an elegant formula involving matrix multiplication, for the marginal probability distribution of X_n. Here we ask the question: what is the probability $\mathbb{P}(X_n = j)$ that at time n our system is in state i. For example, for $n = 1$ we can write:

$$\mathbb{P}(X_1 = j) = \sum_{i \in I} \mathbb{P}(X_0 = i, X_1 = i),$$

by considering all possible initial states i. In fact, the events

$$\{\text{state } i \text{ at time } 0, \text{ state } j \text{ at time } 1\}$$

do not intersect for different $i \in I$ and their union gives the event

$$\{\text{state } j \text{ at time } 1\}.$$

Now use (1.3) and recall the rules of matrix algebra:

$$\sum_{i \in I} \mathbb{P}(X_0 = i, X_1 = j) = \sum_{i \in I} \lambda_{i_0} p_{ij} = (\lambda P)_j.$$

By a direct calculation, this formula is extended to a general n:

$$\begin{aligned} &\mathbb{P}(X_n = j) \\ &= \sum_{i_0, \ldots, i_{n-1}} \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0, \ldots, i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} = (\lambda P^n)_j, \end{aligned} \tag{1.5}$$

where P^n is the nth power of matrix P. That is, the stochastic vector describing the distribution of X_n is obtained by applying matrix P^n to the initial stochastic vector λ.

Then, similarly,

$$\begin{aligned} &\mathbb{P}(X_n = i, X_{n+1} = j) \\ &= \sum_{i_0, \ldots, i_{n-1}} \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i, X_{n+1} = j) \\ &= \sum_{i_0, \ldots, i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij} = (\lambda P^n)_i \, p_{ij}, \end{aligned}$$

and hence

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \frac{\mathbb{P}(X_n = i, X_{n+1} = j)}{\mathbb{P}(X_n = i)} = \frac{(\lambda P^n)_i \, p_{ij}}{(\lambda P^n)_i} = p_{ij}. \tag{1.6}$$

In other words, entry p_{ij} is the conditional probability that the state at the next time is j given that at the preceding time it is i.

Moreover,

$$\begin{aligned} &\mathbb{P}(X_0 = i, X_n = j) \\ &= \sum_{i_1, \ldots, i_{n-1}} \mathbb{P}(X_0 = i, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_1, \ldots, i_{n-1}} \lambda_i p_{i i_1} \cdots p_{i_{n-1} j} = \lambda_i (P^n)_{ij}, \end{aligned}$$

and

$$\mathbb{P}(X_n = j | X_0 = i) = \frac{\mathbb{P}(X_0 = i, X_n = j)}{\mathbb{P}(X_0 = i)} = \frac{\lambda_i (P^n)_{ij}}{\lambda_i} = (P^n)_{ij}. \tag{1.7}$$

That is, the entry $(P^n)_{ij}$ of matrix P^n gives the n-step transition probability from state i to j. We will also denote it sometimes by $p_{ij}^{(n)}$.

More generally,

$$\mathbb{P}(X_k = i, X_{n+k} = j) = (\lambda P^k)_i (P^n)_{ij}$$

and

$$\mathbb{P}(X_{k+n} = j | X_k = i) = \frac{\mathbb{P}(X_k = i, X_{k+n} = j)}{\mathbb{P}(X_k = i)} = \frac{(\lambda P^k)_i (P^n)_{ij}}{(\lambda P^k)_i} = (P^n)_{ij}. \tag{1.8}$$

A consequence of this observation is that the power P^n of a stochastic matrix is again stochastic, viz. $\sum_{j\in I} p_{ij}^{(n)} = 1 \ \forall\ i \in I$. Of course, this fact can be verified directly:

$$\sum_{j\in I} p_{ij}^{(n)} = \sum_{i_1,\dots,i_{n-1},j} p_{ii_1}\cdots p_{i_{n-1}j} = \sum_{i_1} p_{ii_1}\cdots\sum_j p_{i_{n-1}j} = 1$$

as at each step (beginning with $\sum_j$) we get the sum one, owing to (1.2).

Another consequence is that if we apply to a stochastic vector a stochastic matrix (P or more generally P^n), we again obtain a stochastic vector. Again, the direct calculation confirms, viz., by (1.1):

$$\sum_j (\lambda P^n)_j = \sum_{i,j} \lambda_i (P^n)_{ij} = \sum_i \lambda_i \sum_j (P^n)_{ij} = \sum_i \lambda_i = 1.$$

An ultimate generalisation of (1.3) is the formula

$$\begin{aligned}&\mathbb{P}\big(X_{k_1} = i_1, X_{k_2} = i_2, \dots, X_{k_n} = i_n\big)\\ &= (\lambda P^{k_1})_{i_1}\big(P^{k_2-k_1}\big)_{i_1 i_2}\cdots\big(P^{k_n-k_{n-1}}\big)_{i_{n-1}i_n}\end{aligned} \tag{1.9}$$

valid $\forall$ times $0 \le k_1 < k_2 < \ldots < k_n$ and states i_1, ..., $i_n \in I$.

It is now the time to summarise our findings. Suppose that $\lambda = (\lambda_i)$ is a stochastic vector and $P = (p_{ij})$ a transition matrix on I. The random state X_n at time n is considered as a random variable with values in I.

Definition 1.1. *We say that a sequence of random variables X_n with values in a finite or countable set I is a Markov chain with the initial distribution λ and transition matrix P if $\forall$ i_0, ..., $i_n \in I$, the joint probability $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n)$ is given by formula* (1.3). In this case we say that (X_n) is Markov (λ, P).

Theorem 1.1. *If (X_n) is Markov (λ, P) then:*

(i) *The conditional probability*

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i)$$

is equal to the conditional probability $\mathbb{P}(X_{n+1} = j | X_n = i)$ and coincides with p_{ij}. In particular, the conditional distribution of X_{n+1} given that $X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i$ does not depend on $i_0, \dots, i_{n-1}$ and coincides with $(p_{ij}, j \in I)$, i.e., with the row i of P.

(ii) *The probability $\mathbb{P}(X_n = i)$ that the state at time n is i equals $(\lambda P^n)_i$.*

(iii) *The entry $p_{ij}^{(n)}$ of matrix P^n coincides with the conditional probability $\mathbb{P}(X_{k+n} = j | X_k = i)$, i.e. gives the n-step transition probability from i to j.*

(iv) *The general probability*

$$\mathbb{P}\big(X_{k_1} = i_1, X_{k_2} = i_2, \dots, X_{k_n} = i_n\big)$$

is given by (1.9).

Examples and remarks. 1.3. Suppose that all rows of P are the same, i.e. $p_{ij} = p_j$ does not depend on i. In addition, suppose that $\lambda_j = p_j$, i.e., λ coincides with the row of P. Then, by (1.3)

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p_{i_0} p_{i_1} \cdots p_{i_n}.$$

Also, in this example $P^n = P$, as

$$p_{ij}^{(n)} = \sum_{i_1,\dots,i_{n-1}} p_{i_1}\cdots p_{i_{n-1}} p_j = \sum_{i_1} p_{i_1} \sum_{i_2} p_{i_2} \cdots \sum_{i_{n-1}} p_{i_{n-1}} p_j = p_j,$$

owing to the fact that $\sum_{l\in I} p_l = 1$. Hence, $\mathbb{P}(X_n = j) = (\lambda P^n)_j = \sum_{i\in I} p_i p_{ij}^{(n)} = \sum_{i\in I} p_i p_j = p_j$. We see that

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1)\cdots\mathbb{P}(X_n = i_n).$$

That is, (X_n) is a sequence of independent, identically distributed random variables.

1.4. If P is diagonal then it must coincide with the unit matrix where row i is given by the stochastic vector δ_i:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0\\ 0 & 1 & 0 & \dots & 0\\ 0 & 0 & 1 & \dots & 0\\ 0 & 0 & 0 & \dots & 1\end{pmatrix}.$$

In this case, every power P^n again equals the unit matrix. Hence, by (1.4): $\mathbb{P}(X_n = i) = \lambda_i$. That is, the distribution of X_n is the same as X_0. In other words, the initial distribution is preserved in time.

1.5. For a two-state Markov chain, $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$. Entries of P^n can be found by a straightforward calculation. In fact, $P^n = P^{n-1}P$, which for entry $p_{00}^{(n)}$ yields

$$p_{00}^{(n)} = p_{00}^{(n-1)}(1-\alpha) + p_{01}^{(n-1)}\beta$$
$$= p_{00}^{(n-1)}(1-\alpha) + \left(1 - p_{00}^{(n-1)}\right)\beta = \beta + (1-\alpha-\beta)p_{00}^{(n-1)}.$$

This is a recursion in n, with $p_{00}^{(0)} = 1$ and $p_{00}^{(1)} = 1-\alpha$. Hence,

$$p_{00}^{(n)} = A + B(1-\alpha-\beta)^n,$$

with

$$A + B = 1, \ A + B(1-\alpha-\beta) = 1-\alpha,$$

and, clearly,

$$p_{00}^{(n)} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n, & \text{if } \alpha+\beta > 0, \\ 1, & \text{if } \alpha = \beta = 0. \end{cases}$$

Entry $p_{11}^{(n)}$ is obtained by swapping α and β and entries $p_{01}^{(n)}$ and $p_{10}^{(n)}$ as complements to 1.

1.6. In the general case, we can use the eigen-values and eigen-vectors of P to find elements of P^n. Consider a 3×3 example:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 \end{pmatrix}.$$

The eigen-values are solutions to the characteristic equation:

$$\det \begin{pmatrix} -\mu & 1 & 0 \\ 0 & 2/3-\mu & 1/3 \\ 1/3 & 0 & 2/3-\mu \end{pmatrix}$$
$$= -\mu^3 + \frac{4}{3}\mu^2 - \frac{4}{9}\mu + \frac{1}{9} = -(\mu-1)\left(\mu^2 - \frac{1}{3}\mu + \frac{1}{9}\right) = 0,$$

whence

$$\mu_0 = 1, \quad \mu_\pm = \frac{1 \pm i\sqrt{3}}{6}.$$

As the eigen-values are distinct, matrix P is diagonalisable: there exists an invertible matrix C such that

$$C^{-1}PC = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1+i\sqrt{3}}{6} & 0 \\ 0 & 0 & \frac{1-i\sqrt{3}}{6} \end{pmatrix}, \quad \text{i.e., } P = C \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1+i\sqrt{3}}{6} & 0 \\ 0 & 0 & \frac{1-i\sqrt{3}}{6} \end{pmatrix} C^{-1}$$

Then

$$P^n = C \begin{pmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1+i\sqrt{3}}{6}\right)^n & 0 \\ 0 & 0 & \left(\frac{1-i\sqrt{3}}{6}\right)^n \end{pmatrix} C^{-1},$$

and each entry of P^n is a sum of the form

$$A + B\left(\frac{1+i\sqrt{3}}{6}\right)^n + C\left(\frac{1-i\sqrt{3}}{6}\right)^n.$$

Coefficients A, B and C may be complex; they vary from entry to entry and are found from the initial values $n = 0, 1, 3$. For $n = 0$, P^0 is the unit matrix (just as in the scalar case $p^0 = 1$ for any p ($p = 0$ included!)); for $n = 1$, we use matrix P and for $n = 2$ we have to square it, to obtain P^2. For instance, suppose that the states are 1, 2 and 3; then the entries are $p_{ij}^{(n)}$, $i, j = 1, 2, 3$. Then, for $p_{12}^{(n)}$:

$$p_{12}^{(0)} = A + B + C = 0, \ p_{12}^{(1)} = A + B\frac{1+i\sqrt{3}}{6} + C\frac{1-i\sqrt{3}}{6} = 1,$$

and

$$p_{12}^{(2)} = A + B\left(\frac{1+i\sqrt{3}}{6}\right)^2 + C\left(\frac{1-i\sqrt{3}}{6}\right)^2 = \frac{2}{3}.$$

The calculations may be simplified if we get rid of imaginary parts (as all entries $p_{ij}^{(n)}$ of P^n are real non-negative). To this end, observe that $\mu_\pm$ are complex conjugate roots and write

$$\frac{1\pm i\sqrt{3}}{6}=\frac{1}{3}\,\frac{1\pm i\sqrt{3}}{2}=\frac{1}{3}e^{\pm i\pi/3}=\frac{1}{3}\left(\cos\,\frac{\pi}{3}+i\sin\,\frac{\pi}{3}\right).$$

Then

$$\left(\frac{1\pm i\sqrt{3}}{6}\right)^n=\left(\frac{1}{3}\right)^n e^{\pm in\pi/3}\left(\frac{1}{3}\right)^n\left(\cos\,\frac{\pi n}{3}+i\sin\,\frac{\pi n}{3}\right),$$

and

$$p_{ij}^{(n)}=\alpha+\left(\frac{1}{3}\right)^n\left(\beta\cos\,\frac{\pi n}{3}+\gamma\sin\,\frac{\pi n}{3}\right),$$

where $\alpha=A$, $\beta=B+C$ and $\gamma=i(B-C)$ must be real. Again, we have the equations for $n=0,1,2$; for $p_{12}^{(n)}$ they are

$$\alpha+\beta=0,\ \alpha+\frac{1}{3}\left(\frac{1}{2}\beta+\frac{\sqrt{3}}{2}\gamma\right)=1,\ \alpha+\frac{1}{9}\left(-\frac{1}{2}\beta+\frac{\sqrt{3}}{2}\gamma\right)=\frac{2}{3},$$

whence

$$\alpha=\frac{3}{7},\ \ \beta=\frac{-3}{7},\ \ \gamma=\frac{9}{7}\sqrt{3}.$$

In particular, $\lim\limits_{n\to\infty}p_{12}^{(n)}=3/7$.

1.7. Consider another three-state example:

$$P=\begin{pmatrix}1/3 & 0 & 2/3\\ 1/3 & 2/3 & 0\\ 1/3 & 1/3 & 1/3\end{pmatrix}.$$

Here the characteristic equation is:

$$-\mu^3+\frac{4}{3}\mu^2-\frac{1}{3}\mu=-(\mu-1)\left(\mu-\frac{1}{3}\right)\mu=0,$$

and the eigen-values are:

$$\mu_0=1,\ \mu_1=\frac{1}{3},\ \mu_2=0.$$

Hence, the entries $p_{ij}^{(n)}$ have a simple form:

$$p_{ij}^{(n)}=A+B\left(\frac{1}{3}\right)^n+C\cdot 0^n.$$

Again we use three initial conditions, with P^0, P and P^2. For instance, for $p_{11}^{(n)}$:

$$A+B+C=1,\ A+\frac{1}{3}B=\frac{1}{3},\ A+\left(\frac{1}{3}\right)^2B=\frac{1}{3},$$

whence $A=1/3$, $B=0$, $C=2/3$ and $p_{11}^{(n)}\equiv 1/3$. Similarly, $p_{12}^{(n)}=1/3-(1/3)^n$ and $p_{13}^{(n)}=1/3+(1/3)^n$. As $n\to\infty$, all entries of the first row of P^n approach $1/3$ (in fact, the same is true for all 9 entries of P^n).

1.8. We can make a number of observations. First, 1 is always an eigen-value of any stochastic matrix P. This is because (i) $\det\,(\mu I-P)=\det(\mu I-P)^{\mathrm{T}}=\det\,(\mu I-P^{\mathrm{T}})$, i.e., the eigen-values of P and the transposed matrix P^{T} coincide, and (ii) 1 is always an eigen-value of P^{T}: the corresponding eigen-vector is the row $\mathbf{1}=(1,\dots,1)$ of 1's. Formally: $\mathbf{1}P^{\mathrm{T}}=\mathbf{1}$, or equivalently, $P\mathbf{1}^{\mathrm{T}}=\mathbf{1}^{\mathrm{T}}$ for the column $\mathbf{1}^{\mathrm{T}}$. To check the last equation, observe that every entry of the column $P\mathbf{1}^{\mathrm{T}}$ is 1:

$$\left(P\mathbf{1}^{\mathrm{T}}\right)_i=\sum_{j\in I}p_{ij}=1,$$

because P is stochastic.

Therefore, the characteristic polynomial of a stochastic matrix is divisible by $(\mu-1)$; in the 3×3 case this leads to a quadratic quotient polynomial, and all eigen-values can be found.

Second, if there is a complex eigen-value μ_+ of P then the complex conjugate $\mu_-=\overline{\mu}_+$ is also an eigen-value, as this is the only possibility to produce a real characteristic polynomial from the product of linear monomials. I mean the product $(\mu-\mu_+)(\mu-\mu_-)=\mu^2-(\mu_+\mu_-)\mu+\mu_+\mu_-$, with real coefficients $\mu_++\mu_-$ and $\mu_+\mu_-=|\mu_\pm|^2$. Then, writing

$$\mu_\pm=|\mu_\pm|e^{\pm i\phi}=|\mu_\pm|(\cos\,\phi\pm i\sin\,\phi),$$

we can work with real summands only, of the form $\beta\cos(n\phi)$ and $\gamma\sin(n\phi)$.

Third, the coefficient A (in front of 1^n) in the equation for $p_{ij}^{(n)}$ typically identifies the limit $\lim_{n\to\infty} p_{ij}^{(n)}$. This is because the modulus $|\mu|$ of any eigen-value μ of P is ≤ 1, and 'generically' (although not always), any eigen-value $\mu \neq 1$ has $|\mu| < 1$. This fact is more delicate and will be commented on in subsequent lectures. Then in the decomposition

$$p_{ij}^{(n)} = A + \sum_{\text{eigenvalues } \mu_s \neq 1} B_s \mu_s^n$$

all terms except for A are suppressed as $n \to \infty$. [In the case of $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ this is not true: the eigen-values are 1 and -1 and there is no limit $\lim_{n\to\infty} p_{ij}^{(n)}$ as P^n vascillates betveen I for n even and P for n odd.]

It has to be said that many (even very simple) examples may lead to rather cumbersome formulas for entries $p_{ij}^{(n)}$. For example, the matrix

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 1/3 & 2/3 & 0 \end{pmatrix}$$

has the characteristic equation

$$-\mu^3 + \frac{5}{6}\mu^2 + \frac{4}{9}\mu - \frac{5}{18} = -(\mu - 1)\left(\mu^2 + \frac{1}{6}\mu - \frac{5}{18}\right) = 0,$$

with eigen-values

$$\mu_0 = 1,\ \mu_\pm = \frac{-1 \pm \sqrt{41}}{12}.$$

This leads to the equation

$$p_{ij}^{(n)} = A + B\left(\frac{-1+\sqrt{41}}{12}\right)^n + C\left(\frac{-1-\sqrt{41}}{12}\right)^n,$$

with

$$A + B + C = \delta_{ij},\ A + B\frac{-1+\sqrt{41}}{12} + \frac{-1+\sqrt{41}}{12} = p_{ij},$$

and

$$A + B\left(\frac{-1+\sqrt{41}}{12}\right)^2 + C\left(\frac{-1+\sqrt{41}}{12}\right)^2 = p_{ij}^{(2)}.$$

For instance, for $p_{21}^{(n)}$ the final expression is

$$\frac{3}{4}\,\frac{13\sqrt{41}-164}{41\sqrt{41}+533} + \frac{3}{4\sqrt{41}}\left(\frac{-1+\sqrt{41}}{12}\right)^n - \frac{3}{4}\,\frac{13-\sqrt{41}}{41+13\sqrt{41}}\left(\frac{-1-\sqrt{41}}{12}\right)^n.$$

1.9. A helpful property is the presence of symmetries in P: it may reduce the number of states in the Markov chain. For example, the $N \times N$ matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha/(N-1) & \ldots & \alpha/(N-1) \\ \alpha/(N-1) & 1-\alpha & \ldots & \alpha/(N-1) \\ \vdots & \vdots & \ldots & \vdots \\ \alpha/(N-1) & \alpha/(N-1) & \ldots & 1-\alpha \end{pmatrix}$$

describes a model of a virus mutation where a virus retains its type or changes to one of different types with equal probability (the types are 1, ..., N).

To calculate $p_{11}^{(n)}$, we reduce the number of states to two (say, 1 and 0 (another)), by considering original transitions from a state 1 to itself or to another state, and backwards, without further specification (as for our problem all other states are indistinguishable). The reduced two-state chain has the 2×2 transition matrix

$$\begin{pmatrix} 1-\alpha & \alpha \\ \alpha/(N-1) & 1-\alpha/(N-1) \end{pmatrix}.$$

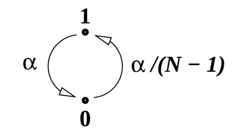

Diagram 1.3.

We can apply formulas of Example 1.5 (with $\beta = \alpha/(N-1)$):

$$p_{11}^{(n)} = \frac{\alpha/(N-1)}{\alpha+\alpha/(N-1)} + \frac{\alpha}{\alpha+\alpha/(N-1)}\left(1-\alpha-\frac{\alpha}{N-1}\right)^n$$
$$= \frac{1}{N} + \frac{N-1}{N}\left(1-\frac{\alpha N}{N-1}\right).$$

We are now in position to establish the famous Markov property of a Markov chain. It asserts that the Markov chain begins afresh after any given time n.

Theorem 1.2. *Let (X_n) be Markov (λ, P). Then, $\forall\ m \geq 1$ and $i \in I$, conditional on $X_m = i$, $(X_{m+n}, n \geq 0)$ is Markov (δ_i, P). In particular, conditional on $X_m = i$, random variables X_{m+1}, X_{m+2}, ... are independent of variables X_0, ..., X_{m-1}.*

In other words, in a Markov chain, the past (X_0, ..., X_{m-1}) and the future $(X_{m+1}, X_{m+2}, \ldots)$ are conditionally independent, given the present $(X_m = i)$.

Proof. Recall, the stochastic vector δ_i has entries δ_{ij}, $j \in I$. We want to check that for any event A determined by X_0, ..., X_{m-1} and B determined by X_{m+1}, ..., X_{m+1+n} for some n, (i) the conditional probability $\mathbb{P}(A \cap B | X_m = i)$ decouples:

$$\mathbb{P}(A\cap B|X_m = i) = \mathbb{P}(B|X_m = i)\mathbb{P}(A|X_m = i), \tag{1.10}$$

and (ii) the conditional probability $\mathbb{P}(B|X_m = i)$ is calculated as in the Markov chain (δ_i, P):

$$\begin{aligned}\mathbb{P}(B|X_m = i) &= \sum_{j_0\in I}\delta_{j_0 i}\sum_{(j_1,\ldots,j_n)\in B} p_{j_0 j_1}\cdots p_{j_{n-1}j_n}\\ &= \sum_{(j_1,\ldots,j_n)\in B} p_{ij_1}\cdots p_{j_{n-1}j_n}.\end{aligned} \tag{1.11}$$

First, let A and B be of the form

$$A = \{X_0 = i_0, \ldots, X_{m-1} = i_{m-1}\},\ B = \{X_{m+1} = j_1, \ldots, X_{m+1+n} = j_n\}$$

for some sequence of states i_0, ..., i_{m-1}, j_1, ..., $j_n \in I$. General A and B are disjoint unions of such 'elementary' events.

For A and B as above,

$$\mathbb{P}\big(A\cap B\cap\{X_m = i\}\big)$$
$$= \mathbb{P}\big(X_0 = i_0, \ldots, X_{m-1} = i_{m-1}, X_m = i, X_{m+1} = j_1, \ldots, X_{m+1+n} = j_n\big)$$
$$= \lambda_{i_0}p_{i_0 i_1}\cdots p_{i_{m-1}i}p_{ij_1}\cdots p_{j_{n-1}j_n}.$$

For a general B we have to sum over $(j_1, \ldots, j_n) \in B$:

$$\lambda_{i_0}p_{i_0 i_1}\cdots p_{i_{m-1}i}\sum_{(j_1,\ldots,j_n)\in B} p_{ij_1}\cdots p_{j_{n-1}j_n}.$$

The sum $\sum\limits_{(j_1,\ldots,j_n)\in B}$ gives the conditional probability $\mathbb{P}(B|X_m = i)$, and it is calculated as in the Markov chain (δ_i, P).

Next, for a general A we sum over $(i_0, \ldots, i_{m-1}) \in A$:

$$\mathbb{P}\big(A\cap B\cap\{X_m = i\}\big) = \sum_{(i_0,\ldots,i_{m-1})\in A}\lambda_{i_0}p_{i_0 i_1}\cdots p_{i_{m-1}i}\mathbb{P}(B|X_m = i)$$
$$= \mathbb{P}(A\cap\{X_m = i\})\mathbb{P}(B|X_m = i).$$

Finally, to produce the conditional probability $\mathbb{P}\big(A\cap B|X_m = i\big)$, we divide by $\mathbb{P}(X_m = i)$:

$$\mathbb{P}\big(A\cap B|X_m = i\big) = \frac{\mathbb{P}\big(A\cap B\cap\{X_m = i\}\big)}{\mathbb{P}(X_m = i)}$$
$$= \frac{\mathbb{P}\big(A\cap\{X_m = i\}\big)}{\mathbb{P}(X_m = i)}\mathbb{P}(B|X_m = i) = \mathbb{P}(A|X_m = i)\mathbb{P}(B|X_m = i),$$

as required. □

In future we will write $\mathbb{P}_i$ for the conditional probabilities $\mathbb{P}(\ \cdot\ |X_0 = i)$ given that the state at time 0 is i.

Example 1.10. (Math Tripos, Part IB, 1991, Question 307D; parts *(a)* and *(b)*) Three girls A, B and C are playing table tennis. In each game, two of the girls play against each other and the third girl does not play. The winner of any given game n plays again in game $n+1$. The probability that girl x will beat girl y in any game that they play against each other is $s_x/(s_x+s_y)$ for $x, y \in \{A, B, C\}$, $x \neq y$, where s_A, s_B, s_C represent the playing strengths of the three girls.

(a) Represent this process as a Markov chain by defining the possible states and constructing the transition matrix.

(b) Determine the probability that the two girls who play each other again in the first game will play each other in the fourth game. Show that this probability does not depend on which two girls play in the first game.

Solution. *(a)* Label states by A, B, C indicating which player is *not* playing in a given game. Then the transition matrix is $\{A, B, C\} \times \{A, B, C\}$:

$$\begin{pmatrix} 0 & \dfrac{s_C}{s_B+s_C} & \dfrac{s_B}{s_B+s_C} \\ \dfrac{s_C}{s_A+s_C} & 0 & \dfrac{s_A}{s_A+s_C} \\ \dfrac{s_B}{s_A+s_B} & \dfrac{s_A}{s_A+s_B} & 0 \end{pmatrix}.$$

The process is a Markov chain because the results of the subsequent games are independent.

(b) Here, we look for the probability that after three steps the chain returns to a given initial state. See Diagram 1.4.

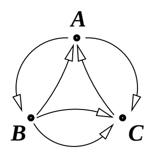

Diagram 1.4.

From the symmetry, this probability is the same for any choice of the initial state and is equal to

$$p_{AB}p_{BC}p_{CA} + p_{AC}p_{CB}p_{BA} = \frac{2s_As_Bs_C}{(s_A+s_B)(s_B+s_C)(s_C+s_A)}.$$