ENEE 627
SPRING 2011
INFORMATION THEORY

MARKOV CHAINS

Throughout, let $\mathcal{X}$ denote a finite set, and refer to its elements as states, hence the terminology state space used sometimes to denote $\mathcal{X}$. A (square) matrix $\boldsymbol{P}$ on $\mathcal{X}$ is simply an $|\mathcal{X}| \times |\mathcal{X}|$ array of scalars, one for each ordered pair of states, namely

$$(p_{xy},\ x, y \in \mathcal{X}) .$$

We shall write $\boldsymbol{P} = (p_{xy})$ when no confusion arises.

Stochastic matrices

Consider a matrix $\boldsymbol{P} = (p_{xy})$ on $\mathcal{X}$. It is said to be a *stochastic* matrix if

$$0 \leq p_{xy} \leq 1, \quad x, y \in \mathcal{X}$$

and

$$\sum_{y \in \mathcal{X}} p_{xy} = 1, \quad x \in \mathcal{X}.$$

Thus, for each x in $\mathcal{X}$, the row

$$(p_{xy},\ y \in \mathcal{X})$$

can be interpreted as a pmf $\boldsymbol{p}_x$ on $\mathcal{X}$.

Furthermore, the matrix $\boldsymbol{P}$ is said to *doubly stochastic* if it is a stochastic matrix such that

$$\sum_{x \in \mathcal{X}} p_{xy} = 1, \quad y \in \mathcal{X}.$$

Powers of $\boldsymbol{P}$

The powers $\boldsymbol{P}$ are defined by

$$\boldsymbol{P}^0 = \boldsymbol{I}, \quad \boldsymbol{P}^{n+1} = \boldsymbol{P}\boldsymbol{P}^n = \boldsymbol{P}^n\boldsymbol{P}, \quad n = 0, 1, \ldots$$

with the identity matrix $\boldsymbol{I}$ on $\mathcal{X}$ naturally defined by

$$\boldsymbol{I} = (\delta_{xy}) .$$

We shall use the notation

$$\boldsymbol{P}^n = \left(p^{(n)}_{xy}\right), \ n = 0, 1, \ldots$$

These definitions are well posed as indicated by the following fact.

Fact 0.1 *We have*

$$\boldsymbol{P}\boldsymbol{P}^n = \boldsymbol{P}^n\boldsymbol{P}, \quad n = 0, 1, \ldots$$

Proof. Easy by induction. ■

Fact 0.2 *For every non-negative integers* $r, s, t = 0, 1, \ldots$, *it is always the case that*

$$\boldsymbol{P}^{r+s+t} = \boldsymbol{P}^r\boldsymbol{P}^s\boldsymbol{P}^t.$$

Proof. Elementary by associativity of the matrix product. ■

Fact 0.3 *If* $\boldsymbol{P}$ *is a stochastic matrix, then each of the matrices* $\{\boldsymbol{P}^n, \ n = 0, 1, \ldots\}$ *of* $\boldsymbol{P}$ *is also a stochastic matrix.*

Proof. Easy by induction. ■

Irreducibility

The stochastic matrix $\boldsymbol{P}$ is said to be *irreducible* if for every pair of distinct states x and y in $\mathcal{X}$ there exist positive integers $n(x, y)$ and $n(y, x)$ such that

$$p^{(n(x,y))}_{xy} > 0 \quad \text{and} \quad p^{(n(y,x))}_{yx} > 0.$$

Period

For any non-empty subset $\{n_\alpha,\ \alpha \in A\}$ of $\mathbb{N}$, we denote its *greatest common denominator* by

$$\text{g.c.d.}\,(n_\alpha,\ \alpha \in A)\,.$$

For each state x in $\mathcal{X}$ we define its *period* $d(x)$ as the integer

$$d(x) = \text{g.c.d.}\left(n = 1, 2, \ldots : p_{xx}^{(n)} > 0\right) \tag{1}$$

with the convention $d(x) = \infty$ if the set $(n = 1, 2, \ldots : p_{xx}^{(n)} > 0)$ is empty. The state x is said to be *periodic* if $d(x) \geq 2$ and *aperiodic* if $d(x) = 1$.

Theorem 0.1 *An irreducible Markov chain $\boldsymbol{P}$ on $\mathcal{X}$ has the property that either all its states are aperiodic or they are all periodic with the same period.*

Proof. Pick two states x and y in $\mathcal{X}$. The chain $\boldsymbol{P}$ being irreducible, there exist positive integers $n(x,y)$ and $n(y,x)$ such that

$$p_{xy}^{(n(x,y))} > 0 \quad \text{and} \quad p_{yx}^{(n(y,x))} > 0.$$

Therefore,

$$\begin{aligned} p_{yy}^{(n(y,x)+n(x,y))} &= \sum_z p_{yz}^{(n(y,x))} p_{zy}^{(n(x,y))} \\ &\geq p_{yx}^{(n(y,x))} p_{xy}^{(n(x,y))} > 0. \end{aligned} \tag{2}$$

On the other hand, whenever

$$p_{xx}^{(t)} > 0$$

for some $t = 1, 2, \ldots$, then

$$\begin{aligned} p_{yy}^{(n(y,x)+t+n(x,y))} &= \sum_z \sum_v p_{yz}^{(n(y,x))} p_{zv}^{(t)} p_{vy}^{(n(x,y))} \\ &\geq p_{yx}^{(n(y,x))} p_{xx}^{(t)} p_{xy}^{(n(x,y))} > 0. \end{aligned} \tag{3}$$

Therefore, $d(y)$ divides both $n(y,x)+n(x,y)$ and $n(y,x)+t+n(x,y)$, hence $d(y)$ divides t since $n(y,x)+t+n(x,y)-(n(y,x)+n(x,y)) = t$. Thus, $d(y)$ divides all the elements of the set $\left(t = 1, 2, \ldots :\ p_{xx}^{(t)} > 0\right)$, so that, $d(y)$ divides $d(x)$ (which is defined as the g.c.d of this set). A similar argument shows that $d(x)$ divides $d(y)$, whence $d(x) = d(y)$. ■

Markov chains

Consider a stochastic matrix $\boldsymbol{P}$ on $\mathcal{X}$. A collection of $\mathcal{X}$-valued rvs $\{X_n,\ n = 0, 1, \ldots\}$ (defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$) is said to be a (time-homogeneous) Markov chain with one-step transition probabilities $\boldsymbol{P}$ if

$$\mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right] = \mathbb{P}\left[X_0 = x_0\right] \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \tag{4}$$

for each $n = 1, 2, \ldots$ and all $x_0, x_1, \ldots, x_n$ in $\mathcal{X}$. The following fact is key to many of the arguments involving Markov chains.

Theorem 0.2 *Fix $k = 0, 1, \ldots$. Then for each $n = 1, 2, \ldots$, we have*

$$\mathbb{P}\left[X_k = x_0, X_{k+1} = x_1, \ldots, X_{k+n} = x_n\right] = \mathbb{P}\left[X_k = x_0\right] \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \tag{5}$$

with arbitrary $x_0, x_1, \ldots, x_n$ in $\mathcal{X}$.

Proof. Fix $k = 1, 2, \ldots$, $n = 1, 2, \ldots$ and states $x_0, x_1, \ldots, x_n$ in $\mathcal{X}$. For any collection of states $y_0, \ldots, y_{k-1}$ in $\mathcal{X}$, we have from (4) that

$$\begin{aligned} & \mathbb{P}\left[X_0 = y_0, \ldots, X_{k-1} = y_{k-1}, X_k = x_0, X_{k+1} = x_1, \ldots, X_{k+n} = x_n\right] \\ = \ & \mathbb{P}\left[X_0 = y_0\right] \cdot \prod_{j=0}^{k-2} p_{y_j y_{j+1}} \cdot p_{y_{k-1} x_0} \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}}. \end{aligned} \tag{6}$$

Therefore,

$$\begin{aligned} & \mathbb{P}\left[X_k = x_0, X_{k+1} = x_1, \ldots, X_{k+n} = x_n\right] \\ = \ & \sum\nolimits_{y_0, \ldots, y_{k-1}} \mathbb{P}\left[X_0 = y_0, \ldots, X_{k-1} = y_{k-1}, X_k = x_0, X_{k+1} = x_1, \ldots, X_{k+n} = x_n\right] \\ = \ & \sum\nolimits_{y_0, \ldots, y_{k-1}} \mathbb{P}\left[X_0 = y_0\right] \cdot \prod_{j=0}^{k-2} p_{y_j y_{j+1}} \cdot p_{y_{k-1} x_0} \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \\ = \ & \left(\sum\nolimits_{y_0, \ldots, y_{k-1}} \mathbb{P}\left[X_0 = y_0\right] \cdot \prod_{j=0}^{k-2} p_{y_j y_{j+1}} \cdot p_{y_{k-1} x_0} \right) \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \\ = \ & \left(\sum\nolimits_{y_0, \ldots, y_{k-1}} \mathbb{P}\left[X_0 = y_0, \ldots, X_{k-1} = y_{k-1}, X_k = x_0\right] \right) \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \\ = \ & \mathbb{P}\left[X_k = x_0\right] \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \end{aligned}$$

as desired. ■

From (4), for all $x_0, x_1, \ldots, x_n, x_{n+1}$ in $\mathcal{X}$, we get both

$$\mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right] = \mathbb{P}\left[X_0 = x_0\right] \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \tag{7}$$

and

$$\begin{aligned} & \mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = x_{n+1}\right] \\ &= \mathbb{P}\left[X_0 = x_0\right] \cdot \prod_{\ell=0}^{n} p_{x_\ell x_{\ell+1}}, \end{aligned} \tag{8}$$

whence

$$\begin{aligned} & \mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = x_{n+1}\right] \\ &= \mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right] \cdot p_{x_n x_{n+1}} \end{aligned} \tag{9}$$

upon direct comparison of (7) and (8).

Building upon these observations, if

$$\mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right] > 0, \tag{10}$$

it follows that

$$\mathbb{P}\left[X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right] = p_{x_n x_{n+1}}, \tag{11}$$

suggesting the validity of the relation[1]

$$\mathbb{P}\left[X_{n+1} = x_{n+1} | X_n = x_n\right] = p_{x_n x_{n+1}}. \tag{12}$$

To see that this is indeed the case, we argue as follows: By Theorem 0.2 we get

$$\mathbb{P}\left[X_n = x_n, X_{n+1} = x_{n+1}\right] = \mathbb{P}\left[X_n = x_n\right] p_{x_n x_{n+1}}. \tag{13}$$

Under (10) we necessarily have

$$\mathbb{P}\left[X_n = x_n\right] > 0, \tag{14}$$

[1] See discussion below.

and the standard definition

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] = \frac{\mathbb{P}[X_n = x_n, X_{n+1} = x_{n+1}]}{\mathbb{P}[X_n = x_n]} \tag{15}$$

applies. The desired conclusion (12) now follows from (13).

Alternate definition of Markov chains

In most textbooks Markov chains are given a different definition which we now present: A collection of $\mathcal{X}$-valued rvs $\{X_n, \ n = 0, 1, \ldots\}$ (defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$) is said to be a (time-homogeneous) Markov chain with one-step transition probabilities $\boldsymbol{P}$ if

$$\begin{aligned} & \mathbb{P}[X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n] \\ = \; & \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] \end{aligned} \tag{16}$$

for all $x_0, x_1, \ldots, x_n, x_{n+1}$ in $\mathcal{X}$, with

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] = p_{x_n x_{n+1}}. \tag{17}$$

The difficulty with this definition is that the conditional probabilities involved in (16) are well defined only when

$$\mathbb{P}[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n] > 0 \tag{18}$$

and

$$\mathbb{P}[X_n = x_n] > 0 \tag{19}$$

Obviously, (18) implies (19) but the converse is not true, possibly creating ambiguities with the definitions being inconsistent with each other.[2]

A possible solution to this difficulty is to read (16)-(17) as stating instead that

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n] = p_{x_n x_{n+1}} \tag{20}$$

with the understanding that if

$$\mathbb{P}[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n] = 0,$$

then the right handside of (20) is taken to be the definition of the conditional probability that $X_{n+1} = x_{n+1}$ given that $X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n$. With this definition it is easy to check that both (4) and (12) hold.

[2]Recall that the conditional probability $\mathbb{P}[A|B]$ is not uniquely defined when $\mathbb{P}[B] = 0$ with each other.

Stationary Markov chains

Consider the (time-homogeneous) Markov chain $\{X_n,\ n = 0, 1, \ldots\}$ with one-step transition probabilities $\boldsymbol{P}$. We write

$$\pi_n(x) = \mathbb{P}\left[X_n = x\right], \qquad \begin{array}{c} x \in \mathcal{X} \\ n = 0, 1, \ldots \end{array}$$

and organize these probabilities into a row vector

$$\boldsymbol{\pi}_n = (\pi_n(x),\ x \in \mathcal{X}) .$$

Using the law of total probabilities we get

$$\pi_{n+1}(x) = \sum\nolimits_y \pi_n(y) p_{yx}, \qquad \begin{array}{c} x \in \mathcal{X} \\ n = 0, 1, \ldots \end{array}$$

or in vector notation

(21) $$\boldsymbol{\pi}_{n+1} = \boldsymbol{\pi}_n \boldsymbol{P}, \quad n = 0, 1, \ldots$$

Theorem 0.3 *Let $\boldsymbol{\mu}$ denote the pmf of the initial condition X_0. Then, the (time-homogeneous) Markov chain $\{X_n,\ n = 0, 1, \ldots\}$ with one-step transition probabilities $\boldsymbol{P}$ is stationary if and only if*

(22) $$\boldsymbol{\mu} \boldsymbol{P} = \boldsymbol{\mu}.$$

Any pmf on $\mathcal{X}$ which satisfies (22) is called a *stationary* pmf for $\boldsymbol{P}$.

Proof. First, assume that the Markov chain $\{X_n,\ n = 0, 1, \ldots\}$ is stationary. This implies that for each $n = 0, 1, \ldots$, the rv X_n has the same distribution as X_0, i.e., $\boldsymbol{\pi}_n = \boldsymbol{\mu}$. Substituting this information into (21) yields (22).

Conversely, assume that the initial state X_0 is distributed according to a pmf $\boldsymbol{\mu}$ which satisfies the fixed -point equation (22). Using this fact in conjunction with (21) we get that

$$\boldsymbol{\pi}_1 = \boldsymbol{\pi}_0 \boldsymbol{P} = \boldsymbol{\mu} \boldsymbol{P} = \boldsymbol{\mu}$$

so that $\boldsymbol{\pi}_0 = \boldsymbol{\mu}$. Iterating we conclude that

$$\boldsymbol{\pi}_n = \boldsymbol{\mu}, \quad n = 0, 1, \ldots$$

Fix $k = 0, 1, \ldots$ and $n = 1, 2, \ldots$. With arbitrary $x_0, x_1, \ldots, x_n$ in $\mathcal{X}$, Theorem 0.2 states that

$$\begin{aligned} & \mathbb{P}\left[X_k = x_0, X_{k+1} = x_1, \ldots, X_{k+n} = x_n\right] \\ = & \ \mathbb{P}\left[X_k = x_0\right] \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \\ = & \ \mathbb{P}\left[X_0 = x_0\right] \cdot \prod_{\ell=0}^{n-1} p_{x_\ell x_{\ell+1}} \\ = & \ \mathbb{P}\left[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right]. \end{aligned} \tag{23}$$

This establishes the stationarity of the Markov chain. ■

Existence and uniqueness of stationary pmfs

The stationary pmf is *not* unqiue if $\boldsymbol{P}$ is *not* irreducible: For instance, with $\mathcal{X} = \{0, 1\}$ and $\boldsymbol{P} = \boldsymbol{I}$, *every* pmf on $\mathcal{X}$ is a stationary pmf.

More generally, partition $\mathcal{X}$ into two non-empty subsets $\mathcal{X}_1$ and $\mathcal{X}_2$ so that $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. Assume the stochastic matrix $\boldsymbol{P}$ on $\mathcal{X}$ to be of the form

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_1 & \boldsymbol{O}_{12} \\ \boldsymbol{O}_{21} & \boldsymbol{P}_2 \end{pmatrix} \tag{24}$$

with $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ stochastic matrices on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Here $\boldsymbol{O}_{11}$ and $\boldsymbol{O}_{21}$ are matrices with all zero entries of the appropriate dimensions. Assume now that $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are stationary pmfs for $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, respectively. For each λ in $(0, 1)$, the pmf $\boldsymbol{\mu}_\lambda$ on $\mathcal{X}$ defined by

$$\boldsymbol{\mu}_\lambda = (\lambda \boldsymbol{\mu}_1, (1 - \lambda)\boldsymbol{\mu}_2)$$

is stationary pmf for $\boldsymbol{P}$.

Limit theorems for Markov chains

Several limit results are available under certain conditions. The strongest such results guarantee the convergence

$$\lim_{n \to \infty} \pi_n(x) = \pi(x), \quad x \in \mathcal{X} \tag{25}$$

for some pmf $\boldsymbol{\pi}$ on $\mathcal{X}$, or in vector notation

$$\lim_{n\to\infty} \boldsymbol{\pi_n} = \boldsymbol{\pi}. \tag{26}$$

Sometimes it is only possible to show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \pi_k(x) = \pi(x), \quad x \in \mathcal{X} \tag{27}$$

for some pmf $\boldsymbol{\pi}$ on $\mathcal{X}$, or in vector notation

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \boldsymbol{\pi}_k = \boldsymbol{\pi}. \tag{28}$$

Obviously (25)-(26) implies (27)-(28) since usual convergence implies Cesaro convergence.

Before giving conditions for either (25)-(26) or (27)-(28) to hold, we make a couple of comments as to the identify of the limit pmf π appearing there.

If (25)-(26) takes place, then letting n go to infinity in (21) we conclude that

$$\begin{aligned} \lim_{n\to\infty} \boldsymbol{\pi}_{n+1} &= \lim_{n\to\infty} (\boldsymbol{\pi}_n \boldsymbol{P}) \\ &= \left(\lim_{n\to\infty} \boldsymbol{\pi}_n \right) \boldsymbol{P} \end{aligned} \tag{29}$$

since finite summation permute with limits. Thus, in the limit

$$\boldsymbol{\pi} = \boldsymbol{\pi P} \tag{30}$$

and $\boldsymbol{\pi}$ is necessarily a stationary pmf for $\boldsymbol{P}$.

In a similar vein, for each $n = 1, 2, \ldots$, we find

$$\begin{aligned} & \frac{1}{n+1} \sum_{k=0}^{n} \boldsymbol{\pi}_k \\ & \frac{1}{n+1} \left(\boldsymbol{\pi}_0 + \sum_{k=1}^{n} \boldsymbol{\pi}_{k-1} \boldsymbol{P} \right) \\ = \quad & \frac{1}{n+1} \boldsymbol{\pi}_0 + \frac{n}{n+1} \cdot \left(\frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\pi}_{k-1} \right) \boldsymbol{P}. \end{aligned} \tag{31}$$

Letting n go to infinity and assuming that (27)-(28) holds, we readily conclude that the limit $\boldsymbol{\pi}$ in (27)-(28) again satisfies (30), and $\boldsymbol{\pi}$ is necessarily a stationary pmf for $\boldsymbol{P}$.

The case

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

with $\mathcal{X} = \{0, 1\}$ is quite instructive. Obviously $\boldsymbol{P}$ is irreducible and periodic with all states having period two. It is also easy to see that for any pmf $\boldsymbol{\pi}$ on the initial state X_0, we have

$$\mathbb{P}[X_n = 1] = \begin{cases} \mathbb{P}[X_0 = 1] = \pi(1) & \text{if } n \text{ odd} \\ \\ \mathbb{P}[X_0 = 0] = 1 - \pi(1) & \text{if } n \text{ even} \end{cases}$$

It is now plain that (25)-(26) does not hold unless $\pi(1) = \pi(0) = \frac{1}{2}$, i.e., the uniform pmf on $\mathcal{X}$. Observe also that (27)-(28) always holds in this case with $\boldsymbol{\pi}$ uniform on $\mathcal{X}$. Thus, irreducibility is not sufficient by itself to ensure (25)-(26). Failure to have convegence can be traced to peridicity.

Theorem 0.4 *If the Markov chain is irreducible and aperiodic, then there exists a uniqiue stationary pmf $\boldsymbol{\mu}$ for $\boldsymbol{P}$ and (25)-(26) always holds with limit $\boldsymbol{\mu}$.*

Theorem 0.5 *If the Markov chain is irreducible (and possibly periodic), then there exists a uniqiue stationary pmf $\boldsymbol{\mu}$ for $\boldsymbol{P}$ and (27)-(28) always holds with limit $\boldsymbol{\mu}$.*

Consider the case

$$\boldsymbol{P} = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \quad \text{with} \quad 0 \leq a, b \leq 1 \tag{32}$$

The cases $a = b = 1$ and $a = b = 0$ have already been discussed. It is straightforward to check that (22) takes the form

$$\begin{aligned} \mu(0) &= a\mu(0) + (1-b)\mu(1) \\ \mu(1) &= (1-a)\mu(0) + b\mu(1) \end{aligned}$$

(33)

This reduces to

$$(1-a)\mu(0) = (1-b)\mu(1)$$

and the constraint $\mu(0) + \mu(1) = 1$ yields

$$\mu(0) = \frac{1-a}{2-(a+b)} \quad \text{and} \quad \mu(1) = \frac{1-b}{2-(a+b)}$$

provided $a + b < 2$, in which case (22) has a unique solution! The case $a + b = 2$ is equivalent to $a = b = 1$, for which there are infinitely solutions as we have seen earlier.
