UPPSALA UNIVERSITY
Department of Mathematics
© Docent ÖS

Markov Processes, 1MS012
Spring semester 2024

Lecture 5 Markov Processes, 1MS012

1 Markov chains in the long run

Example: Let

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

From computer computations we see that

$$\mathbf{P}^n = \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{pmatrix} \approx \begin{pmatrix} 0.33 & 0.67 \\ 0.33 & 0.67 \end{pmatrix}$$

if n is large. It seems to be a limiting matrix with identical rows.

Thus

$$\underbrace{(P(X_n = 0), P(X_n = 1))}_{(\pi_0^n, \pi_1^n)} = (\pi_0^0, \pi_1^0)\mathbf{P}^n \approx (0.33, 0.67),$$

if n is large, and the initial distribution seems to be "forgotten".

What can be said about the limit vector (distribution?) (π_0, π_1) if it exists?

1.1 Markov chains (finite state-space)

Suppose (X_n) is a Markov chain with finite state-space $S = (1, 2, ..., n_0)$, $n_0 < \infty$, and

$$\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k,$$

for all j and k in S.

Then

(a) $\pi = (\pi_1, \ldots, \pi_{n_0})$ is a probability distribution since

$$1 = \sum_k p_{jk}^{(n)} \Rightarrow 1 = \lim_{n\to\infty} \sum_k p_{jk}^{(n)} \underbrace{=}_{\text{since } S \text{ is finite}} \sum_k \lim_{n\to\infty} p_{jk}^{(n)} = \sum_k \pi_k$$

(b) π is a solution to the equation $\pi = \pi\mathbf{P}$, i.e. $\pi_k = \sum_j \pi_j p_{jk}$ for all $k \in S$, since

$$\begin{aligned} \pi_k &= \lim_{n\to\infty} p_{ik}^{(n)} \underbrace{=}_{\text{Chapm.Kolm.}} \lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n-1)} p_{jk} \\ &\underbrace{=}_{\text{since } S \text{ is finite}} \sum_{j\in S} \underbrace{\lim_{n\to\infty} p_{ij}^{(n-1)}}_{\pi_j} p_{jk} \end{aligned} \tag{1}$$

A probability distribution satisfying $\pi = \pi\mathbf{P}$ is called a **stationary distribution** for the Markov chain since if the Markov chain starts according to π, i.e. $P(X_0 = k) = \pi_k$ for all k, then X_n is π-distributed for each n since

$$\underbrace{\boldsymbol{\mu}_n}_{\text{dist. of } X_n} = \underbrace{\boldsymbol{\mu}_0}_{\text{dist. of } X_0} \mathbf{P}^n = \pi\mathbf{P}^n = \underbrace{\pi\mathbf{P}}_{\pi}\mathbf{P}^{n-1} = \ldots = \pi.$$

More generally it follows that $(X_n)_{n=0}^{\infty}$ is a stationary sequence i.e. $(X_0, X_1, ..., X_k)$ has the same distribution as $(X_n, X_{n+1}, ..., X_{n+k})$, for any fixed n and k.

If X_n is π-distributed where π is stationary, then we sometimes say that the chain is in equilibrium.

(c) In the long run the proportion of time spent in state j is $\pi_j = \frac{1}{\mathrm{E}(T_j)}$. This can be seen by first observing that it is true if j is transient. If j is recurrent and $T_{jj}^{(n)}$ denotes the time for the nth revisit to j starting at j then by the law of large numbers

$$\frac{T_{jj}^{(n)}}{n} = \frac{T_{jj}^{(1)} + (T_{jj}^{(2)} - T_{jj}^{(1)}) + ... + (T_{jj}^{(n)} - T_{jj}^{(n-1)})}{n} \to \mathrm{E}(T_j),$$

as $n \to \infty$ (with probability one). Therefore it follows that

$$\frac{\text{number of revisits in } j \text{ up to time } T_{jj}^{(n)}}{T_{jj}^{(n)}} = \frac{n}{T_{jj}^{(n)}} \to \frac{1}{\mathrm{E}(T_j)}, \tag{2}$$

as $n \to \infty$ (with probability one).

Let $n(N)$ denote the number of revisits to state j made at time N. If we define $I_n = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{otherwise} \end{cases}$, then,

$$\frac{n(N)}{T_{jj}^{(n(N)+1)}} \leq \frac{\sum_{n=1}^{N} I_n}{N} \leq \frac{n(N)}{T_{jj}^{(n(N))}},$$

and therefore from (2) and a "sandwich"-argument, it follows that

$$\frac{\sum_{n=1}^{N} I_n}{N} = \frac{|\{n \leq N : X_n = j\}|}{N} \to \frac{1}{\mathrm{E}(T_j)},$$

as $N \to \infty$ (with probability one).

Now

$$\mathrm{E}(\frac{\sum_{n=1}^{N} I_n}{N}) = \frac{1}{N} \sum_{n=1}^{N} \underbrace{\mathrm{E}(I_n)}_{P(X_n=j)} \to \pi_j,$$

as $N \to \infty$. Therefore $\pi_j = \frac{1}{\mathrm{E}(T_j)}$.

Example: (continued) Let (X_n) be a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}.$$

Note that (X_n) is irreducible. We want to find all stationary distributions, i.e. solutions to $\pi\mathbf{P} = \pi$, where $\pi = (\pi_1, \pi_2)$, $\pi_1 + \pi_2 = 1$, $\pi_1 \geq 0$, $\pi_2 \geq 0$:

$$\pi\mathbf{P} = \pi \Leftrightarrow (\pi_1, \pi_2) = (0.6\pi_1 + 0.2\pi_2, 0.4\pi_1 + 0.8\pi_2).$$

We thus need to solve

$$\begin{cases} 0.6\pi_1 + 0.2\pi_2 & = \pi_1 \\ 0.4\pi_1 + 0.8\pi_2 & = \pi_2 \\ \pi_1 + \pi_2 & = 1 \\ \pi_1 \geq 0, \ \pi_2 \geq 0. & \end{cases}$$

This system has unique solution $\pi_1 = 1/3$, $\pi_2 = 2/3$, and thus $\pi = (1/3, 2/3)$ is the unique stationary distribution for the Markov chain.

We have earlier seen that $\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k$, for all j and k in S, in this example, and therefore, by property c) above, it follows that $\mathrm{E}(T_1) = 3$ and $\mathrm{E}(T_2) = 3/2$.

Example: Let

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.2 & 0.8 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix $\mathbf{P}$ is a transition matrix for a reducible (i.e. not irreducible) Markov chain.

If $\pi\mathbf{P} = \pi$ for a distribution π i.e. a vector $\pi = (\pi_1, \pi_2, \pi_3)$ with $\sum_{i=1}^{3} \pi_i = 1$, and $\pi_i \geq 0$, $i = 1, 2, 3$, then

$$\begin{cases} 0.6\pi_1 + 0.2\pi_2 & = \pi_1 \\ 0.4\pi_1 + 0.8\pi_2 & = \pi_2 \\ \pi_3 & = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 & = 1 \\ \pi_1 \geq 0,\ \pi_2 \geq 0,\ \pi_3 \geq 0 & \end{cases}$$

This system has many solutions. Each solution is of the form

$$(\pi_1, \pi_2, \pi_3) = t(1/3, 2/3, 0) + (1-t)(0, 0, 1),$$

for some fixed $0 \leq t \leq 1$.

Thus there are many stationary distributions in this case. We cannot hope that $\lim_{n\to\infty} p_{jk}^{(n)}$ exists and is independent of j in this case since if we start in $\{1, 2\}$, then we will stay there, and similarly if we start in $\{3\}$.

The limiting behavior of $p_{jk}^{(n)}$ here thus depends on *both* j and k.

If $X_0 \in \{1, 2\}$, then we may regard this as the state space and thus $p_{j1}^{(n)} \to 1/3$, $p_{j2}^{(n)} \to 2/3$ as $n \to \infty$, for any $j \in \{1, 2\}$. Trivially, $p_{33}^{(n)} \to 1$ and $p_{3j}^{(n)} \to 0$ as $n \to \infty$, for any $j \in \{1, 2\}$, and therefore

$$\mathbf{P}^n \to \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

as $n \to \infty$.

Recall: Decomposition theorem:

$$S = T \cup C_1 \cup C_2 \cup ...,$$

where T are the set of transient states and C_i, $i \geq 1$ are closed irreducible sets of recurrent states.

If S is finite and (X_n) starts in a transient state then it will end up in some set of recurrent states in the long run, but it is random what set it will be.

1.2 Markov chains (countably infinite state space)

If S is infinite and $\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k$ for all j and k then it can happen that

$$1 = \sum_{k\in S} p_{jk}^{(n)} = \lim_{n\to\infty} \sum_{k\in S} p_{jk}^{(n)} > \sum_{k\in S} \lim_{n\to\infty} p_{jk}^{(n)} = \sum_{k\in S} \pi_k.$$

Example: Simple random walk: $\lim_{n\to\infty} p_{jk}^{(n)} = 0$ for all j and k. No stationary distribution can exist.
Recall: All states are null-recurrent for a simple random walk (with $p = 1/2$), i.e. $\mathrm{E}(T_k) = \infty$, for all k.

Theorem: If

$$\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k,$$

and $\sum_k \pi_k = 1$, then π is a stationary distribution.

Proof: In general, taking limits as in (1) gives $\pi_k \geq \sum_{j\in S} \pi_j p_{jk}$, for any $k \in S$, but if the inequality is strict for some k, then

$$1 = \sum_{k\in S} \pi_k > \sum_{k\in S} \sum_{j\in S} \pi_j p_{jk} = \sum_{j\in S} \pi_j = 1,$$

which is a contradiction. Thus $\pi_k = \sum_{j\in S} \pi_j p_{jk}$, for any $k \in S$, i.e. π is a stationary distribution.

Theorem: If (X_n) is an irreducible Markov chain then

All states are positive-recurrent $\Leftrightarrow$ A stationary distribution exists

If a stationary distribution π exists then it is unique and given by $\pi_k = \frac{1}{\mathrm{E}(T_k)}$.

Proof: ($\Rightarrow$)
Suppose k is a positive recurrent state. Let $\nu_i(k)$ be the expected number of visits of the Markov chain to state i between two successive visits to state k. That is, if $X_0 = k$, we define

$$\nu_i(k) = \mathrm{E}(\sum_{n=1}^{\infty} I(X_n = i, T_k \geq n)) = \sum_{n=1}^{\infty} P(X_n = i, T_k \geq n | X_0 = k),$$

where I denotes the indicator function. It is sufficient to check that $\nu = (\nu_i(k) : i \in S)$ satisfies $\nu P = \nu$, since by definition $\sum_i \nu_i(k) = \mathrm{E}(T_k) < \infty$, and we can therefore normalize ν and obtain a stationary probability vector, by dividing each component with $\mathrm{E}(T_k)$.

We have

$$\begin{aligned}
\nu_j(k) &= \sum_{n=1}^{\infty} P(X_n = j, T_k \geq n | X_0 = k) \\
&= \sum_{n=1}^{\infty} \sum_{i \in S} P(X_n = j, X_{n-1} = i, T_k \geq n | X_0 = k) \\
&= \sum_{n=1}^{\infty} \sum_{i \in S} \underbrace{P(X_n = j | X_{n-1} = i, T_k \geq n, X_0 = k)}_{p_{ij}} P(X_{n-1} = i, T_k \geq n | X_0 = k) \\
&= \sum_{i \in S} \sum_{n=1}^{\infty} p_{ij} P(X_{n-1} = i, T_k \geq n | X_0 = k) \\
&= \sum_{i \in S} p_{ij} \underbrace{\sum_{m=0}^{\infty} P(X_m = i, T_k > m | X_0 = k)}_{\nu_i(k)} = \sum_{i \in S} \nu_i(k) p_{ij}.
\end{aligned}$$

Thus if $\pi_i = \frac{\nu_i(k)}{\mathrm{E}(T_k)}$, then $\pi = (\pi_i : i \in S)$ is a stationary distribution.

Proof: ($\Leftarrow$)
Suppose $\pi = (\pi_k : k \in S)$ is a stationary distribution.

We first observe that all states then have to be recurrent, since if all states were transient, then $p_{jk}^{(n)} \to 0$, as $n \to \infty$, for all j and k, and then $\pi_k = \sum_{j \in S} \pi_j p_{jk}^{(n)} \to 0$, as $n \to \infty$, for all j and k which contradicts the assumption that π is a probability distribution.

Next observe that $\pi_k > 0$ for all k. We can see this by noting that if $\pi_k = 0$, for some $k \in S$, then by stationarity $0 = \pi_k = \sum_{j \in S} \pi_j p_{jk}^{(n)} \geq \pi_j p_{jk}^{(n)}$, for all j and n. Since we may choose n such that $p_{jk}^{(n)} > 0$ (irreducibility), it follows that $\pi_j = 0$ for all j which contradicts the assumption that π is a probability distribution.

It remains to prove that $\mathrm{E}(T_k)\pi_k = 1$, for all $k \in S$.

Let (X_n) be a stationary Markov chain starting according to π, i.e. $P(X_0 = k) = \pi_k$, for all $k \in S$. Then

$$
\begin{aligned}
\mathrm{E}(T_k)\pi_k &= \sum_{n=1}^{\infty} P(T_k \geq n | X_0 = k) P(X_0 = k) \\
&= \sum_{n=1}^{\infty} P(T_k \geq n, X_0 = k) \\
&= P(X_0 = k) + \sum_{n=2}^{\infty} P(T_k \geq n, X_0 = k) \\
&= P(X_0 = k) + \sum_{n=2}^{\infty} P(X_0 = k, X_m \neq k, 1 \leq m \leq n-1) \\
&= P(X_0 = k) \\
&\quad + \sum_{n=2}^{\infty} (P(X_m \neq k, 1 \leq m \leq n-1) - P(X_m \neq k, 0 \leq m \leq n-1)) \\
&= P(X_0 = k) \\
&\quad + \sum_{n=2}^{\infty} (P(X_m \neq k, 0 \leq m \leq n-2) - P(X_m \neq k, 0 \leq m \leq n-1)) \\
&\qquad\qquad \text{(by stationarity)} \\
&= P(X_0 = k) \\
&\quad + P(X_0 \neq k) + \lim_{n \to \infty} P(X_m \neq k, 0 \leq m \leq n-1) \\
&\qquad\qquad \text{(sum of telescoping serie)} \\
&= 1 \\
&\qquad\qquad \text{(using recurrence of } k\text{)}
\end{aligned}
$$

Thus $\mathrm{E}(T_k) = \frac{1}{\pi_k} < \infty$, for any $k \in S$, and the stationary distribution π is therefore unique. □

It follows from the proof above that positive recurrence is a class property i.e. if $\mathrm{E}(T_i) < \infty$ and $i \leftrightarrow j$ then $\mathrm{E}(T_j) < \infty$. Therefore in particular if S is

finite then irreducible Markov chains have only positively recurrent states.

Example: Let (X_n) be a Markov chain on $S = (0, 1, 2, \ldots)$ with $p_{i,i+1} = 1/2 = p_{i,0}$, $i \geq 0$. Then $\pi = (\pi_0, \pi_1, \pi_2, \ldots) = (1/2, 1/4, 1/8, \ldots)$ is the unique stationary distribution for (X_n). Thus if T_k denotes the first return time to state k then $E(T_k) = 1/\pi_k = 2^{k+1}$. □

The following theorem generalizes the law of large numbers in the discrete setting:

Theorem: (Law of large numbers for discrete Markov chains)
Let (X_n) be an irreducible Markov chain with state space $S = \{0, 1, 2 \ldots\}$ and suppose a stationary distribution π exists. Then π is unique and for any bounded function $f : S \to \mathbb{R}$,

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{n} \to \sum_k f(k)\pi_k$$

with probability one as $n \to \infty$.

Remark: The law of large numbers for discrete Markov chains is an example of an ergodic theorem where the "time average" $\frac{\sum_{k=0}^{n-1} f(X_k)}{n}$ of f converges to the "space average" $\sum_k f(k)\pi_k$ of f. □

Proof: The uniqueness of π, where $\pi_k = \frac{1}{\mathrm{E}(T_k)}$, $k \in S$, follows from the theorem above. By the same argument as in proving property (c) above we see that the long run the proportion of time spent in state k is π_k, for any $k \in S$, i.e. if $n_k(n)$ denotes the number of visits to state k before time n, then $\frac{n_k(n)}{n} \to \pi_k$ as $n \to \infty$, with probability one for any $k \in S$. Note that if F is a finite set then $\sum_{k \notin F} \frac{n_k(n)}{n} = 1 - \sum_{k \in F} \frac{n_k(n)}{n} = \sum_{k \notin F} \pi_k - \sum_{k \in F} (\frac{n_k(n)}{n} - \pi_k)$,

so

$$\begin{aligned}
\left|\frac{\sum_{k=0}^{n-1} f(X_k)}{n} - \sum_k f(k)\pi_k\right| &= \left|\sum_{k\in S}(\frac{n_k(n)}{n} - \pi_k)f(k)\right| \\
&\leq \sup_{k\in S}|f(k)|\sum_{k\in S}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&= \sup_{k\in S}|f(k)|\sum_{k\in F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\quad + \sup_{k\in S}|f(k)|\sum_{k\notin F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\leq \sup_{k\in S}|f(k)|\sum_{k\in F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\quad + \sup_{k\in S}|f(k)|\sum_{k\notin F}\left(\frac{n_k(n)}{n} + \pi_k\right) \\
&\leq 2\sup_{k\in S}|f(k)|\sum_{k\in F}\left|\frac{n_k(n)}{n} - \pi_k\right| \\
&\quad + 2\sup_{k\in S}|f(k)|\sum_{k\notin F}\pi_k,
\end{aligned}$$

and by chosing F large enough we can make the right hand side as small as we want by chosing n sufficiently large.

□

2 Matlab

2.1 Stationary distribution

```
P=[0.6 0.4; 0.2 0.8]
pi=null(transpose(P-eye(2)))/sum(null(transpose(P-eye(2))))
% πP = π ⇔ π(P − I) = 0 ⇔ (P − I)^t π^t = 0^t.
% Thus π^t need to be in the null-space of (P − I)^t in order for π to be
% a solution.
% A stationary distribution is a normalized such solution.

P=[0.6 0.4 0; 0.2 0.8 0; 0 0 1]
v=null(transpose(P-eye(3)))                % Gives a matrix with 2 column-
                                           % vectors spanning the null-space
```

```
pi1=v(:,1)/sum(v(:,1))
pi2=v(:,2)/sum(v(:,2))          % normalized-null-space-spanning-vectors

% Both examples above can be treated with:
v=null(transpose(P-eye(length(P))))
pi=v.*(ones(length(P),1)*(1./sum(v)))
% (In the second example the matrix pi will have column-vectors
% pi1= (0.3333 0.6667 0) and pi2= (0 0 1).)
```

3 Suggested exercises

Basic exercises:
8–10, 12–16

Extra problems:
a1

Exercises Lawler:
1.9 de, 1.18

Lecture 6 Markov Processes, 1MS012

4 Markov chain convergence theorem

Convergence theorem: Let (X_n) be an aperiodic irreducible Markov chain with transition probability matrix $\mathbf{P}$ and stationary distribution π. Then

$$p_{jk}^{(n)} \to \pi_k = \frac{1}{\mathrm{E}(T_k)} \quad \text{as } n \to \infty \tag{3}$$

for all j and k.

Thus here: $S = T \cup C_1 \cup C_2 \cup \ldots = C_1$ (i.e. no transient states and only one closed irreducible set of recurrent states)

Note that (3) does not mean that any particular trajectory converge.

Proof: We will prove the convergence theorem using a probabilistic technique known as "coupling".

Suppose (X_n) starts at $X_0 = j$.

Let (Y_n) be another Markov chain with transition probability matrix $\mathbf{P}$ starting according to the stationary distribution (i.e. with $P(Y_0 = k) = \pi_k$ for any k) and suppose (Y_n) is independent of (X_n). (Note that Y_n has distribution π for all n.)

Let $T = \min(n \geq 1 : X_n = Y_n)$.

Let

$$Z_n = \begin{cases} X_n & \text{if } T > n, \\ Y_n & \text{if } T \leq n. \end{cases}$$

Clearly X_n and Z_n have the same distribution for any fixed n.

Thus

$$P(X_n = k) = P(Z_n = k) = \underbrace{P(Z_n = k, T \leq n)}_{P(Y_n = k, T \leq n)} + \underbrace{P(Z_n = k, T > n)}_{P(X_n = k, T > n)}$$

and

$$P(Y_n = k) = \underbrace{P(Y_n = k, T \leq n)}_{P(Y_n = k, T \leq n)} + \underbrace{P(Y_n = k, T > n)}_{P(Y_n = k, T > n)}.$$

This implies that

$$\underbrace{|P(X_n = k) - P(Y_n = k)|}_{|p_{jk}^{(n)} - \pi_k|} \leq P(T > n) \to 0, \tag{4}$$

as $n \to \infty$, since $((X_n, Y_n))$ is an irreducible[1] Markov chain with stationary distribution $\pi(j,k) = \pi_j \pi_k$ and therefore positively recurrent and in particular, all states of the form (i,i), $i \in S$ (i.e. the "diagonal") will eventually be reached. We can interpret T as the time to hit the "diagonal".

Remark: If $(X_n)_{n=0}^{\infty}$ is irreducible and *periodic* with period d then we can use the convergence theorem for the aperiodic Markov chains $(X_{nd+k})_{n=0}^{\infty}$, for any fixed $k = 0, 1, ..., d-1$. Note however, that irreducibility may be lost:

Example: The Markov chain (X_n) on $S = (1,2)$ with transition matrix

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

is irreducible and periodic with period 2, and has unique stationary distribution $\pi = (1/2, 1/2)$. The Markov chain (X_{2n}) has transition matrix $\mathbf{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and is not irreducible. Any probability distribution on S is stationary for (X_{2n}).

5 Branching processes

Branching processes are processes used in modeling how sizes of a population grows in generations. Let (X_n) be a stochastic sequence, where we regard X_n as the size of a population at generation n. Consider the following simple model:

Assumption: Suppose each individual in generation n, gives birth to a family of offspring of random size in generation $n+1$, where the offspring sizes

[1]Irreducibility here means that $(i_0, i_1) \to (j_0, j_1)$ for any states i_0, i_1, j_0, and j_1. We know by irreducibility of the individual chains that there exist numbers k and l such that $p_{i_0 j_0}^{(k)} > 0$, and $p_{i_1 j_1}^{(l)} > 0$. By aperiodicity we may choose $l = k$ (See Lecture 3) which means that $(i_0, i_1) \to (j_0, j_1)$. Note that we have exponentially quick convergence rate in (4) in case S is finite, since by irreducibility of $((X_n, Y_n))$ we see that there exists a $p > 0$ and a positive integer n_0 such that $P(T > nn_0) \leq (1-p)^n$.

are independent and identically distributed non-negative integer valued random variables. (We also tacitly assume that an individual only lives for one generation so an individual dies immediately after its family of offspring has been created.)

Let p_m denote the probability that an individual has m offspring.

Then (X_n) is a Markov chain and

$$P(X_{n+1} = j | X_n = k) = P(Y_{n,1} + \ldots + Y_{n,k} = j),$$

where $Y_{n,1}, \ldots, Y_{n,k}$ are independent and identically distributed random variables with $P(Y_{n,j} = m) = p_m$, for all n, j and $m \geq 0$.

Let $D_n = P(X_n = 0 | X_0 = 1)$ denote the extinction probability by the n:th generation for a line starting with one individual, and let

$$D = \lim_{n \to \infty} D_n$$

denote the ultimate extinction probability of the line. (The limit exists since (D_n) is a non-decreasing sequence bounded from above.)

If there are j offspring in the first generation, then to die out by the nth generation, each of these lines must die out in $n-1$ generations. Since they proceed independently, this probability is D_{n-1}^j.

Thus

$$D_n = p_0 + p_1 D_{n-1} + p_2 D_{n-1}^2 + p_3 D_{n-1}^3 + \ldots = G(D_{n-1}), \tag{5}$$

and thus, by taking limits, $D = G(D)$, where

$$G(s) = \sum_{k=0}^{\infty} p_k s^k = p_0 + p_1 s + p_2 s^2 + \ldots$$

denotes the probability generating function of the offspring distribution.

Note that if $p_0 > 0$ and $p_0 + p_1 < 1$, then $G(s) > 0$ for $s \geq 0$ and

$$G'(s) = p_1 + 2p_2 s + 3p_3 s^2 + \ldots \geq 0,$$

and

$$G''(s) = 2p_2 + 6p_3 s + \ldots > 0,$$

so G is increasing and convex. By definition $G(1) = 1$ and $G'(1) = \mu$ where μ denotes the expected number of offspring of an individual.

Thus all positive solutions to $G(x) = x$ must be larger than or equal to one if and only if $\mu \leq 1$, and consequently $D = 1$ if $\mu \leq 1$.

If $\mu > 1$ then there exist a solution to $G(x_0) = x_0$ with $x_0 < 1$. We claim that D is the smallest solution to $G(x) = x$, i.e. $x_0 = D$. In order to prove the claim note first that $D_0 = 0$, and thus we see from (5), and the non-decreasing property of G on $[0, 1]$, that $D_n = G^n(0) \leq G^n(x_0) = x_0$. The claim now follows by taking limits.

Let us call the offspring distribution non-degenerate if $p_1 < 1$. We have proved the following theorem:

Theorem: Suppose the offspring distribution is non-degenerate with probability generating function G. If $D = \lim_{n\to\infty} P(X_n = 0 \mid X_0 = 1)$ denotes the ultimate extinction probability, then

$$D = 1 \quad \Leftrightarrow \quad \mu \leq 1,$$

and if $\mu > 1$, then D is the unique solution to the equation $D = G(D)$, with $0 < D < 1$.

Thus the population will eventually die out if the expected number of offspring of an individual is less than or equal to one, and there is a positive probability that the population does not die out otherwise.

Remark: In the degenerated case when $p_1 = 1$, then obviously $D = 0$. More generally $D > 0$ if and only if $p_0 > 0$.

Remark: Equation (5) can be generalized, and we can express the distribution of $(X_n|X_0 = 1)$ in terms of G; Let $G_{X_n}(s) = \mathrm{E}(s^{X_n}|X_0 = 1)$ denote the probability generating function of X_n, given $X_0 = 1$. Arguing in a similar way as in (5) we get

$$\begin{aligned} G_{X_n}(s) &= \mathrm{E}(s^{X_n}|X_0 = 1) = \sum_{k=0}^{\infty} p_k \mathrm{E}(s^{X_n}|X_1 = k) = \sum_{k=0}^{\infty} p_k (G_{X_{n-1}}(s))^k \\ &= G(G_{X_{n-1}}(s)) = \ldots = G^n(s), \end{aligned}$$

where we, in the last step, used that $G_{X_1}(s) = G(s)$. Thus in particular we get

$$\mathrm{E}(X_n) = G'_{X_n}(1) = (G^n)'(1) = G'(\underbrace{G^{n-1}(1)}_{1}) \underbrace{(G^{n-1})'(1)}_{\mathrm{E}(X_{n-1})} = \mu \mathrm{E}(X_{n-1}) = \ldots = \mu^n.$$

6 Appendix

The Markov chain convergence theorem can also be proved using non-probabilistic techniques and in particular, if the state space is finite, as a corollary of the Perron-Frobenius theorem of linear algebra;

6.1 The Perron–Frobenius theorem

Theorem: Let A be a matrix with non-negative entries such that A is primitive, i.e. A^n has only positive entries for some n. Then

(a) There is an eigenvalue $r > 0$ of A with algebraic multiplicity 1 such that $|\lambda| < r$ for any other eigenvalue λ of A.

(b) There exists a unique positive *probability vector* π satisfying $\pi A = r\pi$, and a unique positive vector v such that $Av^T = rv^T$ with $\pi v^T = 1$.

(c) $\lim_{k\to\infty} A^k/r^k = v^T\pi$.

Remark: If we extend the definition of irreducibility and period for transition matrices to general non-negative matrices in the natural way, then it can be proved that a non-negative matrix A is irreducible and aperiodic if and only if A is primitive.

If $\mathbf{P}$ is a transition matrix then the row sums are equal to one and thus $v^T = \mathbf{1}^T$ (the column vector containing only ones) is an eigenvector corresponding to the eigenvalue $r = 1$. As a corollary of the Perron-Frobenius theorem we thus get:

Corollary: Let $\mathbf{P}$ be the transition matrix of an irreducible aperiodic Markov chain with finite state space. Then

(a) $r = 1$ is an eigenvalue of $\mathbf{P}$ with algebraic multiplicity 1, and $|\lambda| < 1$ for any other eigenvalue λ of $\mathbf{P}$.

(b) There exists a unique positive *probability vector* π satisfying $\pi\mathbf{P} = \pi$.

(c) $\lim_{k\to\infty} \mathbf{P}^k = \mathbf{1}^T\pi$.

We will not prove the core part ((a), and (b)) of the Perron-Frobenius theorem here, but an outline of proof is given in Lawler, exercise 1.20.

All real matrices cannot be diagonalized, but if A is an irreducible and aperiodic non-negative matrix, then there exist an invertible matrix $\mathcal{P}$ such that $\mathcal{P}^{-1}A\mathcal{P} = J$, where the first row vector of $\mathcal{P}^{-1}$ is π, the first column vector of $\mathcal{P}$ is v^T, and where J is a block diagonal matrix with

$$\lim_{k\to\infty} J^k/r^k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Thus

$$\lim_{k\to\infty} A^k/r^k = \lim_{k\to\infty} (\mathcal{P}J^k\mathcal{P}^{-1})/r^k = \mathcal{P}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}\mathcal{P}^{-1} = v^T\pi,$$

explaining the main idea why (a) and (b) implies (c).

7 Suggested exercises

Extra problems:
a3

Exercises Lawler:
1.12, 2.2, 2.3, 2.8

Lecture 7 Markov Processes, 1MS012

8 Markov chains generated by random iterations of functions

Example: Simple random walk on $S = \mathbb{Z}$ can be generated by random iterations of functions where the function $w_1(x) = x + 1$ is chosen with probability p and the function $w_2(x) = x - 1$ is chosen with probability $1-p$. More precisely, if (I_n) is a sequence of independent random variables with $P(I_n = 1) = p$ and $P(I_n = 2) = 1-p$ and we define $X_{n+1} = f_{I_{n+1}}(X_n)$, i.e.

$$X_n = (f_{I_n} \circ f_{I_{n-1}} \circ \cdots \circ f_{I_1})(X_0),$$

then (X_n) is a simple random walk.

In general, random independent iterations of functions from a set S to S generates a Markov chain with state space S.

Example: $S = (0, 1, 2)$. Random iterations with $w_1(x) = 0$ chosen with probability p_1, $w_2(x) = x$ chosen with probability p_2, and $w_3(x) = 2$ chosen with probability p_3, where $p_1 + p_2 + p_3 = 1$, $p_i \geq 0$, generates a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_1 + p_2 & 0 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & 0 & p_2 + p_3 \end{pmatrix}.$$

Conversely, we can also generate any (time-homogeneous) Markov chain by i.i.d. iterations of functions. We use this property when we want to simulate a trajectory of a Markov chain.

Example: Suppose $S = (0, 1, 2)$, and

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

Most computers can generate sequences of numbers $(U_n)_{n\geq 1}$ that more or less well behaves like a sequence of independent $U(0,1)$-distributed random variables.

At time-step n we use U_{n+1} to perform the transitions.

If $0 < U_{n+1} < 1/2$ and $X_n = 0$ we let $X_{n+1} = f_{U_{n+1}}(0) = 0$,
If $1/2 \leq U_{n+1} < 3/4$ and $X_n = 0$ we let $X_{n+1} = f_{U_{n+1}}(0) = 1$,
If $3/4 \leq U_{n+1} < 1$ and $X_n = 0$ we let $X_{n+1} = f_{U_{n+1}}(0) = 2$,

If $0 < U_{n+1} < 1/4$ and $X_n = 1$ we let $X_{n+1} = f_{U_{n+1}}(1) = 0$,
If $1/4 \leq U_{n+1} < 3/4$ and $X_n = 1$ we let $X_{n+1} = f_{U_{n+1}}(1) = 1$,
If $3/4 \leq U_{n+1} < 1$ and $X_n = 1$ we let $X_{n+1} = f_{U_{n+1}}(1) = 2$,

If $0 < U_{n+1} < 1/4$ and $X_n = 2$ we let $X_{n+1} = f_{U_{n+1}}(2) = 0$,
If $1/4 \leq U_{n+1} < 1/2$ and $X_n = 2$ we let $X_{n+1} = f_{U_{n+1}}(2) = 1$,
If $1/2 \leq U_{n+1} < 1$ and $X_n = 2$ we let $X_{n+1} = f_{U_{n+1}}(2) = 2$.

By construction $X_{n+1} = f_{U_{n+1}}(X_n)$, i.e.

$$X_n = (f_{U_n} \circ f_{U_{n-1}} \circ \cdots \circ f_{U_1})(X_0)$$

where (U_n) is a sequence of independent random variables uniformly distributed on $(0,1)$, and $p_{jk} = P(f_{U_n}(j) = k)$, for all $j, k \in S$.

Note that the functions $f_u, u \in (0,1)$ are only of 4 different types in this case. Let

$$g = \begin{pmatrix} 0 & 1 & 2 \\ i & j & k \end{pmatrix},$$

denote a function $g : \{0,1,2\} \to \{0,1,2\}$ with $g(0) = i$, $g(1) = j$ and $g(2) = k$.

If

$$w_1 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

i.e. $w_1(x) = 0$,

$$w_2 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix},$$

$$w_3 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 2 \end{pmatrix},$$

$$w_4 = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 2 & 2 \end{pmatrix},$$

i.e. $w_4(x) = 2$, then random (i.i.d.) iterations with these functions (chosen with equal probabilities) generates a Markov chain with transition matrix $\mathbf{P}$.

The above examples illustrates and can in a straightforward manner be generalized to a proof of the following theorem:

Theorem: (X_n) is a (time-homogeneous) Markov chain $\Leftrightarrow$ (X_n) can be generated by random (i.i.d.) iterations of functions

Remark: If the state space of (X_n) is uncountable, then we may regard the statement of this theorem as the definition of a time-homogeneous Markov chain.

8.1 Fractals

Fractals are sets with self-similarity properties. Many fractals can be obtained by a recursive construction or by random iterations with affine maps.

Example: Sierpinski triangle: $S =$ a triangle in $\mathbb{R}^2$
Pick three points with coordinates A, B, C in the plane building the vertices of a triangle. Consider a Markov chain starting at $X_0 = A$ where, in each step, if $X_n = x$ we let

$$\begin{aligned} X_{n+1} &= f_A(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } A \\ &= x + \frac{1}{2}(A - x) = \frac{x}{2} + \frac{A}{2}, \end{aligned}$$

with probability $1/3$,

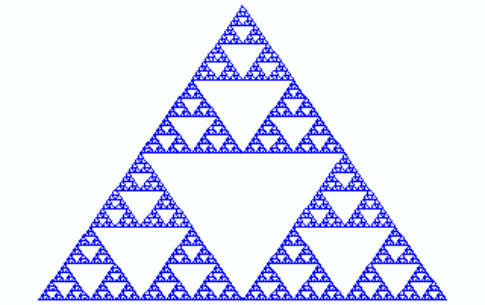
$$\begin{aligned} X_{n+1} &= f_B(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } B \\ &= x + \frac{1}{2}(B - x) = \frac{x}{2} + \frac{B}{2}, \end{aligned}$$

with probability $1/3$,

$$\begin{aligned} X_{n+1} &= f_C(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } C \\ &= x + \frac{1}{2}(C - x) = \frac{x}{2} + \frac{C}{2}, \end{aligned}$$

with probability $1/3$.

If (I_n) is a sequence of independent random variables uniformly distributed on (A, B, C), then $X_{n+1} = f_{I_{n+1}}(X_n)$. The only reachable points are points on the *Sierpinski triangle*. A random trajectory will, with probability one, "draw" the Sierpinski triangle.



The Sierpinski triangle

Example: Cantor set: $S = [0, 1]$
The Markov chain (X_n) generated by independent random iterations with the functions $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$ chosen with equal probabilities (starting at e.g. $X_0 = 0$) generates points on the middle-third *Cantor set*. A random trajectory will, with probability one, "draw" the Cantor set.

The first 5 sets approximating the limiting Cantor set. If (X_n) is a Markov chain on $S = [0, 1]$ generated by i.i.d. iterations with the maps $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$, then the value of X_n can be found in the $n : th$ approximating set of the Cantor set.

Example: Barnsley's fern: $S = \mathbb{R}^2$
The Markov chain (X_n) generated by independent random iterations with the functions

$$f_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0.16 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$f_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1.6 \end{pmatrix},$$

$$f_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.2 & -0.26 \\ 0.23 & 0.22 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1.6 \end{pmatrix},$$

$$f_4\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0.44 \end{pmatrix},$$

(starting at e.g. the origin) generates points on *Barnsley's fern*. A random trajectory will, with probability one, "draw" Barnsley's fern. Barnsley suggested to choose f_1 with probability 0.01, f_2 with probability 0.85, f_3 with probability 0.07, and f_4 with probability 0.07, in order to make this algorithm for drawing Barnsley's fern quick.

Barnsley's fern

9 Matlab

9.1 Simulating a Markov chain with a given transition matrix

```
n=79                                    % Number of steps to be simulated
P=[1/2 1/4 1/4; 1/4 1/2 1/4; 1/4 1/4 1/2];
startstate=1;                           % Choose initial state
MCsim(n,P,startstate)


% In the textfile MCsim.m ...
function MCsim(n,P,startstate)
i=0;
j=startstate;
x=[j];
```

```
while i< n
   j=f(j,rand,P);
   x=[x j];
   i=i+1;
end;
x
end


function y=f(x,u,P)
r=[0 cumsum(P(x,:))];
for j=1:length(P)
   if ((u>r(j)) & (u<=r(j+1)));
      y=j;
   end
end
end
```

9.2 Simulating a trajectory of a Markov chain drawing the Sierpinski triangle

```
n=50000;              % Number of steps to be simulated
A1=[1/2 0; 0 1/2];
A2=[1/2 0; 0 1/2];
A3=[1/2 0; 0 1/2];
b1=[0;0];
b2=[1/2;0];
b3=[1/4;sqrt(3)/4];
p1=1/3;
p2=1/3;
p3=1/3;
startstate=[0;0];
x=startstate;
trajectory=[x];       % We store the trajectory in the
                      % matrix ``trajectory''
for i = 1:n
  u=rand;
  if u<p1
     x=A1*x+b1;
  elseif u<p1+p2
     x=A2*x+b2;
  else
     x=A3*x+b3;
  end
trajectory=[trajectory x];
end
plot(trajectory(1,:),trajectory(2,:),'.')
                      % Draws the Sierpinski triangle
title('The Sierpinski triangle')
axis off;   %  To avoid printing the coordinate axes
```

9.3 Simulating a trajectory of a Markov chain drawing Barnsley's fern

```
n=50000;   % Number of iterations
A1=[0 0;0 0.16];
A2=[0.85 0.04;-0.04 0.85];
A3=[0.2 -0.26;0.23 0.22];
A4=[-0.15 0.28;0.26 0.24];
b1=[0;0];
b2=[0;1.6];
b3=[0;1.6];
b4=[0;0.44];
p1=0.01;
p2=0.85;
p3=0.07;
p4=0.07;
startstate=[0;0];
x=startstate;
trajectory=[x];
for i = 1:n
  u=rand;
  if u<p1
     x=A1*x+b1;
  elseif u<p1+p2
     x=A2*x+b2;
  elseif u<p1+p2+p3
     x=A3*x+b3;
  else
     x=A4*x+b4;
  end
trajectory=[trajectory x];
end
plot(trajectory(1,:),trajectory(2,:),'g.')
                    % Draws a green fern
title('Barnsleys fern')
axis off;
```

10 Suggested exercises

Basic exercises:
17