Stochastic processes

Károly Simon

This course is based on the book:
Essentials of Stochastic processes
by R. Durrett

Department of Stochastics
Budapest University of Technology and Economics
www.math.bme.hu/~simonk

2023 File A

1. We collect a lot of natural examples (see slide 297 for the collection of examples) which can be studied by the theory of Markov chains.
2. We introduce the most important notions and most important theorems without proofs. (Proofs come in File BB.)
3. Compute the stationary distributions.
4. Recurrence properties of Markov chains.
5. We study the death and birth processes as a special case of reversible Markov chains.
6. Exist distributions for absorbing Markov chains.
7. Branching processes.

Gambler's ruin

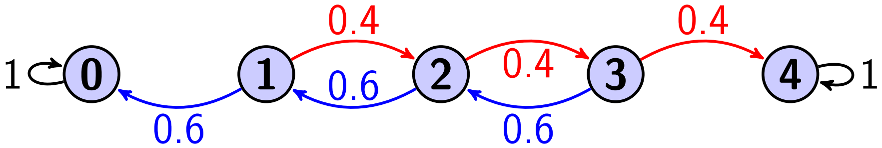
Example 1.1

We start with a gambling game, in which in every turn:

- we win \$1 with probability $p = 0.4$,
- we lose \$1 with probability $1 - p = 0.6$.

The game stops if we reach a fixed amount of $N = \$4$ or if we lose all our money.

We start at $\$X_0$, where $X_0 \in \{1, 2, 3\}$.
Let X_n be the amount of money we have after n turns.
In this case



X_n has the "Markov property". That is:
if we know X_n, any other information about the past is irrelevant for predicting the next state of X_{n+1}. Thus:

$$(1)\quad \mathbb{P}\left(X_{n+1}=j \middle| X_n=i, X_{n-1}=b_{n-1},\dots,X_0=b_0\right) = \mathbb{P}\left(X_{n+1}=j \middle| X_n=i\right),$$

which is 0.4, in the given example.

Homogeneous discrete-time Markov chain

Definition 1.2

Let S be a finite or a countably infinite (we call it countable) set. We say that X_n is a (time) homogeneous discrete-time Markov chain on state space S, with transition matrix $\mathbf{P} = p(i,j)$, if for any n, and any $i, j, b_{n-1}, \dots, b_0 \in S$:

$$\text{(2)} \quad \mathbb{P}\left(X_{n+1} = j \middle| X_n = i, X_{n-1} = b_{n-1}, \dots, X_0 = b_0\right) = p(i,j)$$

We consider only time homogeneous Markov chains and some times we abbreviate them MC.

Initial distribution

A Markov chain is determined by its initial distribution and its transition matrix. The initial distribution $\alpha = (\alpha_i)_{i \in S}$, $(\alpha_i \geq 0, \sum_{i \in S} \alpha_i = 1)$ is the distribution of the state from which a Markov chain starts. When we insist that the Markov chain starts from a given $i \in S$ (in this case $\alpha_i = 1$ and $\alpha_j = 0$ for $j \in S,\ j \neq i$) then all probabilities and expectations are denoted by

$$\mathbb{P}_i\left(\cdot\right), \mathbb{E}_i\left[\cdot\right].$$

In some cases, we write $\mathbb{P}_\alpha\left(\cdot\right), \quad \mathbb{E}_\alpha\left[\cdot\right]$ or we specify the initial distribution α in words, and then we write simply $\mathbb{P}\left(\cdot\right), \mathbb{E}\left[\cdot\right]$.

In the Gambler's ruin example, if $N = 4$ then the transition matrix $\mathbf{P}$ is a 5×5 matrix

	0	**1**	**2**	**3**	**4**
0	1	0	0	0	0
1	0.6	0	0.4	0	0
2	0	0.6	0	0.4	0
3	0	0	0.6	0	0.4
4	0	0	0	0	1

Here and many places later, the bold green numbers like $0, \ldots, 4$ are the elements of the state space. So, they are NOT part of the matrix. They are the indices. The matrix above is a 5×5 matrix. For example: $p(0,0) = 1$ and $p(3,4) = 0.4$.

A simulation with Mathematica

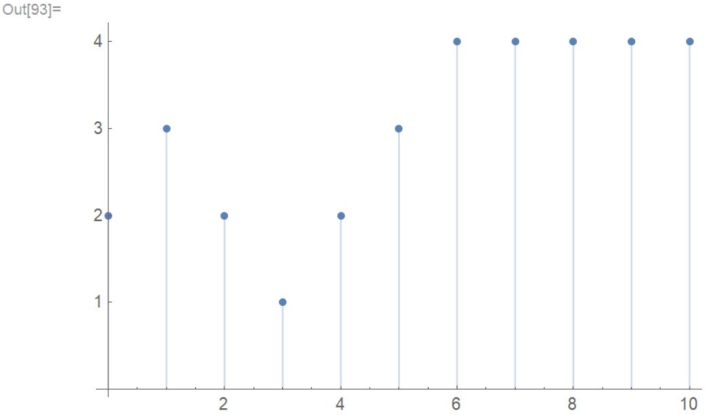


Figure: Gamblar's ruin simulation

The Mathematica code for the previous simulation

```
𝒫 = DiscreteMarkovProcess[{0, 0, 1, 0, 0},
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{6}{10} & 0 & \frac{4}{10} & 0 & 0 \\ 0 & \frac{6}{10} & 0 & \frac{4}{10} & 0 \\ 0 & 0 & \frac{6}{10} & 0 & \frac{4}{10} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

```
]
```

```
In[92]:= data = RandomFunction[𝒫, {0, 10}]

Out[92]=
TemporalData[ + Time: 0 to 10
                Data points: 11    Paths: 1 ]

In[93]:= ListPlot[data - 1, Filling → Axis, Ticks → {Automatic, {0, 1, 2, 3, 4}}]
```

Andrey Markov, 1856 – 1922

Ehrenfest chain

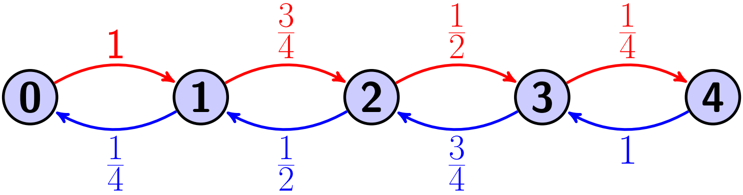
Example 1.3

We have two urns (left and right urn), in which there are a total of N balls. We pick a random ball and take it into the other urn. Let X_n be the number of balls in the left urn after the n^{th} draw. X_n has the Markov-property, because

$$p(i, i+1) = \frac{N-i}{N}, \; p(i, i-1) = \frac{i}{N} \text{ if } 0 \leq i \leq N$$

and $p(i, j) = 0$ otherwise.

$N = 4$, the corresponding graph and transition matrix:



	0	**1**	**2**	**3**	**4**
0	0	1	0	0	0
1	1/4	0	3/4	0	0
2	0	2/4	0	2/4	0
3	0	0	3/4	0	1/4
4	0	0	0	1	0

A simulation with Mathematica

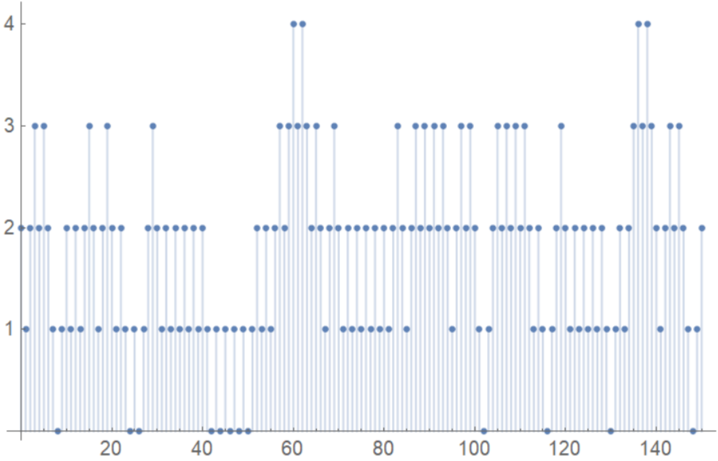


Figure: A simulation for Ehrenfest chain simulation

Another simulation with Mathematica

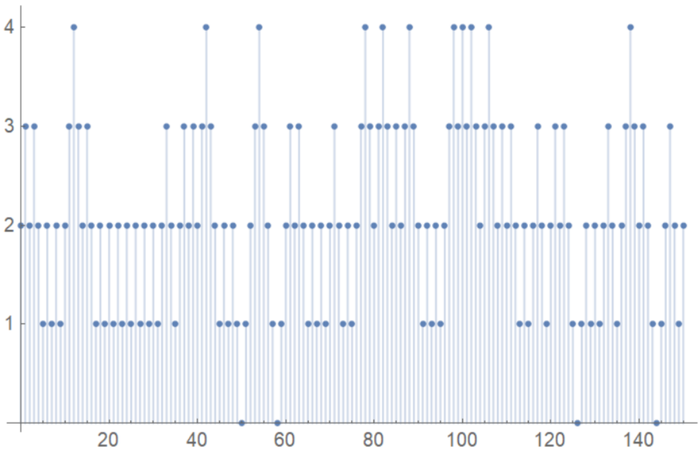


Figure: Another simulation for Ehrenfest chain simulation

The Mathematica code for the previous two simulations

$$\mathcal{P} = \text{DiscreteMarkovProcess}\left[\{0, 0, 1, 0, 0\}, \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{2}{4} & 0 & \frac{2}{4} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}\right]$$

```
data = RandomFunction[𝒫, {0, 150}]
```

```
TemporalData[ Time: 0 to 150
              Data points: 151    Paths: 1 ]
```

```
ListPlot[data - 1, Filling → Axis, Ticks → {Automatic, {0, 1, 2, 3, 4}}]
```

Tatyana Pavlovna Ehrenfest (1876–1964)

Compare the previous two chains I.

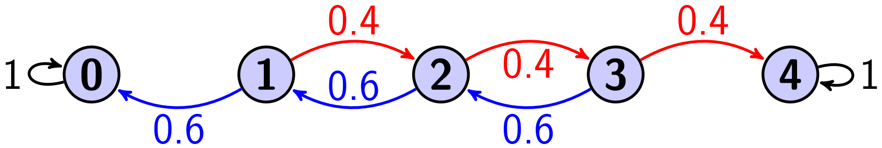


Figure: Gambler's ruin chain:

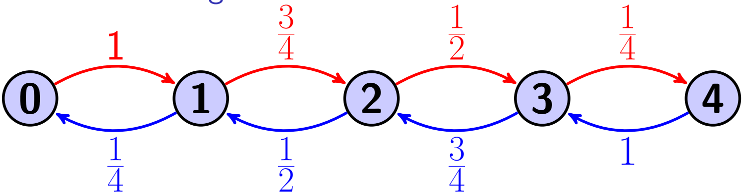


Figure: Ehrenfest chain

Compare the previous two chains II.

First, we consider the Gambler's ruin case. Let us say we start from state 2. In the gambler' ruin case with probability 0.16 we reach state 4 in two steps, and with probability 0.36 we reach state 0 and then we stay there forever. Therefore the states 0 and 4 are absorbing states. That is the probability that starting from 2 we ever return to 2 at least one more time is less than $p := 0.48 = 1 - (0.16 + 0.36)$. Then after the first return, everything starts as before independently. So, the probability that we return to 2 at least twice is less than p^2, and similarly, the probability that we return to 2 at least n times is less than p^n.

Compare the previous two chains III.

So, the probability that we return to 2 infinitely many times is $\lim_{n\to\infty} p^n = 0$. That is starting from 2, we visit 2 only finitely many times almost surely. We call those states where we return only finitely many times almost surely, transient states. Since the same reasoning applies for states $1, 3$ we can see that in the Gambler's ruin example, states $1, 2, 3$ are transient. The states where we return infinitely many times almost surely are called recurrent. Every state is either transient or recurrent.

Compare the previous two chains IV.

We spend only finite time at each transient states. So, if the state space S is finite, then we spend finite time altogether at all transient states together. This implies that

for a finite state MC we always have recurrent states.

Clearly the absorbing states $\{0, 4\}$ are always recurrent states. The following interesting questions will be answered later. To answer the first of the following two problems we need to learn about the so-called exit distributions (see Section 10.1) and to answer the second one we need to study the so-called exit times (see Section 10.2).

Compare the previous two chains V.

Problem 1.4

Starting from 2 *what the probability that the gambler eventually wins is? That is she gets to* 4*?*

We answer this on slide 202, see also slide 45.

Problem 1.5

Starting from 2*, what is the expected number of steps until the gambler gets to either* 0 *(ruin) or to* 4 *(success)?*

We answer this question on slide 238.

Compare the previous two chains VI.

Now we turn to the Ehrenfest chain:

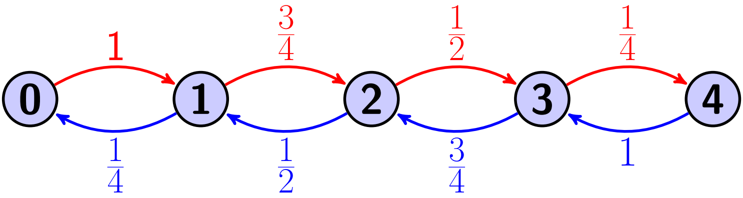


Figure: Ehrenfest chain

We consider the case again when we start from state 2. Then with 1/2-1/2 probability, we jump to either state 1 or 3.

Compare the previous two chains VII.

The probability that we do not return to 2 in any of the next $2n$ steps is $(1/4)^n$. So, the probability that we actually never return to state 2 is $\lim_{n\to\infty}\left(\frac{1}{4}\right)^n = 0$. So we return to 2 almost surely. But when we are at 2 then the whole argument repeats. So we obtain that we return to 2 infinitely many times almost surely. This means that 2 is a recurrent state. With a very similar argument, one can show that the same holds for all the other states. This means that all of the states are recurrent. In this case, we can reach from every state to every state with positive probability (after some steps). In such a situation we say that the MC is irreducible.

Compare the previous two chains VII.

Here we can ask the following question:

Problem 1.6

What is the expected number of steps so that starting from $i \in \{0, \ldots, 4\}$ we get back to i for the first time?

The answer is the reciprocal of the i-th component of the so-called stationary distribution which is a probability vector $\pi = (\pi_i)_{i \in S}$, $\pi_i \geq 0$, $\sum\limits_{i \in S} \pi_i = 1$ satisfying:

$$\pi^T \cdot P = \pi^T \tag{3}$$

This is computed in a more general case, on slide 155.

Mathematica code for the stationary distribution

In this special case we use Mathematica we get
$\pi = \left(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}\right)$.

In[119]:= $p = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{2}{4} & 0 & \frac{2}{4} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$

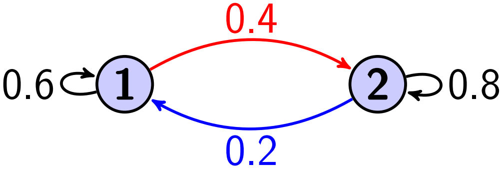
```
In[120]:= invmatrep =
  Inverse[ReplacePart[p - IdentityMatrix[Length[p[[1]]]], {i_, Length[p[[1]]]} :> 1]]

In[121]:= invmatrep[[Length[p[[1]]]]]
```

Out[121]= $\left\{\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}\right\}$

Weather chain

Let X_n be the weather on day n on a given island, with

(4) $$X_n := \begin{cases} 1, & \text{if day } n \text{ is rainy;} \\ 2, & \text{if day } n \text{ is sunny} \end{cases}$$



	1	**2**
1	0.6	0.4
2	0.2	0.8

Question: What is the long-run fraction of sunny days?

π for the Weather chain

For weather chain: $\mathbf{P} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$ We are looking for a random vector $\boldsymbol{\pi} = (\pi_1, \pi_2)$ for which:

$$(\pi_1, \pi_2) \cdot \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} = (\pi_1, \pi_2).$$

The solution is $\boldsymbol{\pi} = (\frac{1}{3}, \frac{2}{3})$. This follows from the general result about the stationary distribution of two-states MC:

Stationary state for general two states MC

Lemma 2.1

A two-state MC's transition matrix can be written in the following way:

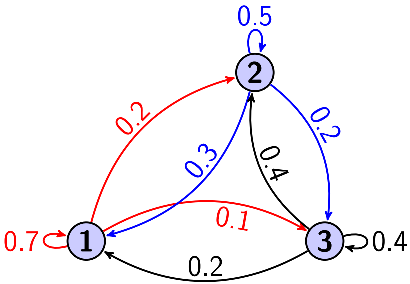
$$\mathbf{P} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

Then the stationary distribution is $\boldsymbol{\pi} = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$.

The proof is trivial.

Social mobility chain

Let X_n be a family's social class in the n^{th} generation, if
lower class:1 middle class:2 upper class:3



	1	2	3
1	0.7	0.2	0.1
2	0.3	0.5	0.2
3	0.2	0.4	0.4

Question: Do the fractions of people in the three classes stabilize after a long time?

For the social mobility chain

For the social mobility chain $\mathbf{P} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$ the equation of $\boldsymbol{\pi}^T \cdot \mathbf{P} = \boldsymbol{\pi}^T$ is

$$\begin{aligned} 0.7\pi_1 &+ 0.3\pi_2 + 0.2\pi_3 = \pi_1 \\ 0.2\pi_1 &+ 0.5\pi_2 + 0.4\pi_3 = \pi_2 \\ 0.1\pi_1 &+ 0.2\pi_2 + 0.4\pi_3 = \pi_3 \end{aligned}$$

The 3^{rd} equation gives us no more information than we have already known. So, we can throw it away, and we

For the social mobility chain (cont.)

replace it with the condition that the sum of the components of $\boldsymbol{\pi}$ equals to 1. We obtain after this replacement:

$$
\begin{array}{rcrcrcl}
0.7\pi_1 & + & 0.3\pi_2 & + & 0.2\pi_3 & = & \pi_1 \\
0.2\pi_1 & + & 0.5\pi_2 & + & 0.4\pi_3 & = & \pi_2 \\
\pi_1 & + & \pi_2 & + & \pi_3 & = & 1
\end{array} \tag{5}
$$

After straightforward algebraic manipulations we get:

$$
\begin{array}{rcrcrcl}
-0.3\pi_1 & + & 0.3\pi_2 & + & 0.2\pi_3 & = & 0 \\
0.2\pi_1 & + & -0.5\pi_2 & + & 0.4\pi_3 & = & 0 \\
\pi_1 & + & \pi_2 & + & \pi_3 & = & 1
\end{array} \tag{6}
$$

For the social mobility chain (cont.)

$$\boldsymbol{\pi}^T \cdot \mathbf{A} = (0, 0, 1),$$

where $\boldsymbol{\pi}^T$ is a row vector and

$$\mathbf{A} := \begin{bmatrix} -0.3 & 0.2 & 1 \\ 0.3 & -0.5 & 1 \\ 0.2 & 0.4 & 1 \end{bmatrix}$$

So

$$\boldsymbol{\pi}^T = (0, 0, 1) \cdot A^{-1} \tag{7}$$

Steps of computing vector π:

For the social mobility chain (cont.)

1. Start with the transition matrix $\mathbf{P}$,
2. subtract 1 from its diagonal elements,
3. replace the last column with the vector whose all elements are equal to 1.
4. The matrix that we obtained is called A.
5. By formula (7): The last row of matrix A^{-1} is $\boldsymbol{\pi}$.

For the social mobility chain (cont.)

In the case of the social mobility chain:

$$A^{-1} = \begin{pmatrix} -\frac{90}{47} & \frac{20}{47} & \frac{70}{47} \\ -\frac{10}{47} & -\frac{50}{47} & \frac{60}{47} \\ \frac{22}{47} & \frac{16}{47} & \frac{9}{47} \end{pmatrix}.$$

And from it: $\boldsymbol{\pi} = \left(\frac{22}{47}, \frac{16}{47}, \frac{9}{47}\right)$.

Multistep transition probabilities

Let $p^m(i,j)$ be the probability that the Markov chain with transition matrix $\mathbf{P} = p(i,j)$, starting from state i is in state j after m steps.

$$p^m(i,j) \overset{\text{in general}}{\neq} \underbrace{p(i,j) \cdots p(i,j)}_{m} \tag{8}$$

Multistep transition probabilities (cont.)

We would like to compute the m-step transition matrix with $\mathbf{P}$.
First observe that

$$p^{m+n}(i,j) = \sum_k p^m(i,k) \cdot p^n(k,j). \tag{9}$$

This is called the Chapman-Kolmogorov equation.
The proof is obvious from the following Figure:

Multistep transition probabilities (cont.)

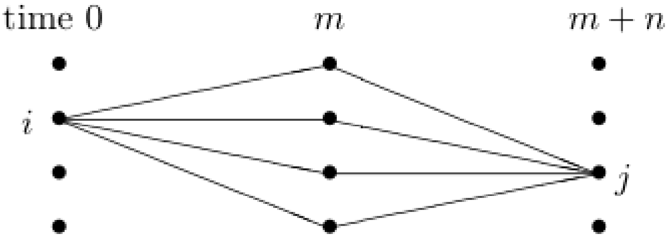


Figure: The Figure is from [3]

Multistep transition probabilities (cont.)

Theorem 3.1

The m-step transition probability $\mathbb{P}\left(X_{n+m} = j | X_n = i\right)$ *is the* (i, j)*-th element of the m-th power of the transition matrix.*

Multistep transition probabilities (cont.)

In the Gambler's ruin example, where the transition matrix was:

P	**0**	**1**	**2**	**3**	**4**
0	1	0	0	0	0
1	0.6	0	0.4	0	0
2	0	0.6	0	0.4	0
3	0	0	0.6	0	0.4
4	0	0	0	0	1

Multistep transition probabilities (cont.)

The $\lim_{n\to\infty} \mathbf{P}^n$ limit also exists, and we will see that it equals to:

$\lim_{n\to\infty} \mathbf{P}^n$	**0**	**1**	**2**	**3**	**4**
0	1	0	0	0	0
1	57/65	0	0	0	8/65
2	45/65	0	0	0	20/65
3	27/65	0	0	0	38/65
4	0	0	0	0	1

In the Ehrenfest chain example, where the transition matrix was:

	0	**1**	**2**	**3**	**4**
0	0	1	0	0	0
1	1/4	0	3/4	0	0
2	0	2/4	0	2/4	0
3	0	0	3/4	0	1/4
4	0	0	0	1	0

The $\lim_{n\to\infty} \mathbf{P}^n$ limit also exists, and we will see that it is the matrix on the next slide. Namely, the limit is a 5×5 matrix such that all of its rows are the stationary distribution vector π cf. slide 26.

Multistep transition probabilities (cont.)

$\lim_{n\to\infty} \mathbf{P}^n$	**0**	**1**	**2**	**3**	**4**
0	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$
2	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$
3	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$
4	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

- A square matrix **P** is a stochastic matrix if all elements are non negative and all the row-sums are equal to 1.
- For a stochastic matrix **P** we obtain the corresponding adjacency matrix A_P by replacing all non-zero elements of P by 1. So, if

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.2 & 0.1 & 0.7 \\ 0.7 & 0.3 & 0 \end{pmatrix} \text{ then } A_P = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

- We are given a Markov Chain (MC) X_n with (finite or countably infinite) state space S and transition matrix $\mathbf{P} = (p(i,j))_{i,j \in S}$ (which is always a stochastic matrix).
- We write

$$\mathbb{P}_x(A) := \mathbb{P}\left(A | X_0 = x\right).$$

$\mathbb{E}_x$ notates the expected value for the probability $\mathbb{P}_x$.
The time of the first visit to y:

$$T_y := \min\left\{n \geq 1 : X_n = y\right\}$$

So, even if we start from y, $T_y \neq 0$.

- Let $i, j \in S$, where S is the state space. We say that i and j communicate if there exists an n and an m such that $p^n(i,j) > 0$ and $p^m(j,i) > 0$.
- Observe that "communicates with" is an equivalence relation. The classes of the corresponding partition of S are called communication classes or simply classes.
- If there is only one communication class (everybody communicates with everybody) then we say that the Markov Chain (MC) is irreducible.

- Consider the MC with $S := \{1,2,3,4\}$ and
$$\mathbf{P} := \begin{pmatrix} 0 & 0.4 & 0 & 0.6 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0.1 & 0 & 0.9 & 0 \end{pmatrix}.$$
Then
$$\mathbf{P}^2 = \begin{pmatrix} 0.26 & 0. & 0.74 & 0. \\ 0. & 0.35 & 0. & 0.65 \\ 0.22 & 0. & 0.78 & 0. \\ 0. & 0.31 & 0. & 0.69 \end{pmatrix}$$
This chain is irreducible because for every $i, j \in S$ either $p(i,j) > 0$ or $p^2(i,j) > 0$ (here $p^2(i,j)$ is the (i,j)-th element of $\mathbf{P}^2$.

- The corresponding adjacency matrices for every n are:

$$A_{P^{2n-1}} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, A_{P^{2n}} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

- For the chain above the greatest common divisor (gcd):

$$\gcd\{n : p^n(i,i) > 0\} = 2 \text{ for } \forall i \in S. \tag{10}$$

Then we say that the period of every state is 2. In general, the **period of state i** is

$$d_i := \gcd\{n : p^n(i,i) > 0\}.$$

We will see that in a communication class all elements have the same period. So, for an irreducible MC all elements have the same period. If this period is equal to 1 then we say that the irreducible chain is **aperiodic**.

- We say that a state $i \in S$ is **transient** if the MC returns to i finitely many times almost surely.

- We say that a state $i \in S$ is recurrent if the MC returns to i infinitely many times almost surely. Every state is either recurrent or transient.
- If an element of a communication class is recurrent then all other elements of this class are also recurrent. These classes are the recurrent classes, while the other classes are the transient classes.
- If a communication class is closed (no arrow goes out of the class) then it is recurrent class. The non-closed communication classes are the transient class.

- Let $i \in S$ be a recurrent state. We say that i is positive recurrent if the expected time of the first return to i (starting from i) is finite.
- Let $i \in S$ be a recurrent state. We say that i is null recurrent if the expected time of the first return to i (starting from i) is infinite.
- A state $i \in S$ is ergodic if i aperiodic and positive recurrent.
- A Marov chain is ergodic if all of it states are ergodic. In particular, a Markov chain is ergodic if there is an N_0 such that for every $m \geq N_0$ for every $i, j \in S$ the state j can be reached from i in m steps.

- A state $i \in S$ is absorbing if $p_{ii} = 1$ (we cannot go anywhere from this state, it is a trap).
- A Markov Chain is absorbing if every state can reach an absorbing state.
- Stationary distribution π is a probability measure on S ($\pi(i) \geq 0$ and $\sum_{i \in S} \pi(i) = 1$) which satisfies:

$$\pi^T \cdot \mathbf{P} = \pi^T \tag{11}$$

Convention: every vector is a column vector. When I need a row vector, I write transpose of the vector as above.

An example of irreducible classes

Example 4.1

	1	2	3	4	5	6	7
1	0.7	0	0	0	0.3	0	0
2	0.1	0.2	0.3	0.4	0	0	0
3	0	0	0.5	0.3	0.2	0	0
4	0	0	0	0.5	0	0.5	0
5	0.6	0	0	0	0.4	0	0
6	0	0	0	0	0	0.2	0.8
7	0	0	0	1	0	0	0

An example of irreducible classes (cont.)

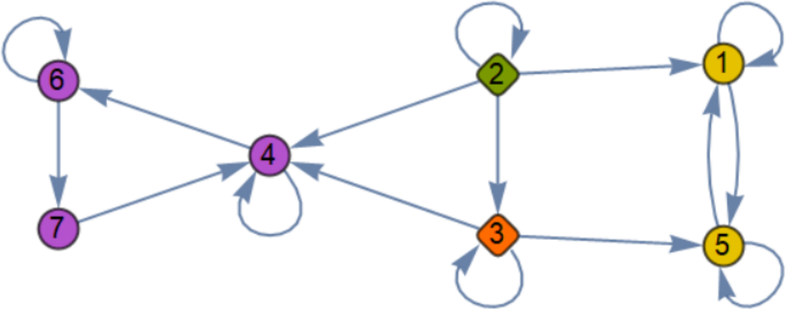


Figure: The graph corresponding to Example 4.1

An example of irreducible classes (cont.)

Let us create a graph whose vertices are the elements of state space $S = \{1, \ldots, 7\}$ and it has directed edge (i, j) if $p(i, j) > 0$. $A \subset S$ is **closed** if it is impossible to get out. So

$$i \in A \text{ and } j \notin A \text{ then } p(i, j) = 0.$$

In the example above: sets $\{1, 5\}$ and $\{4, 6, 7\}$ are closed, so is their union, and even $\{1, 5, 4, 6, 7, 3\}$ and S itself are closed too.

An example of irreducible classes (cont.)

$B \subset S$ is **irreducible** if any two of its elements communicate with one another: $\forall i, j \in B, \quad i \rightsquigarrow j$. So, in the graph that is shown above (slide 57) we can get from every element of B to any other through directed edges; and the irreducible and closed sets are: $\{1, 5\}$ and $\{4, 6, 7\}$. That is the irreducible classes are: $\{1, 5\}$ and $\{4, 6, 7\}$.

Definitions I

Here we follow Senata's book [8, Section 1.2]. For a $k \geq 1$ we use the shorthand notation

$$[k] := \{1, \ldots, k\}.$$

We consider here only square matrices with non-negative elements. If we replace all positive elements of such a matrix to get its adjacency matrix. That is the adjacency matrix is a $0-1$ matrix. Let $A = (a_{i,j})_{i,j=1}^{n}$ be an $n \times n$ adjacency matrix. Then $a_{i,j} \in \{0, 1\}$.
We say that $i, i_1, \ldots, i_{k-1}, j$ is a chain of length of k between i and j if

$$a_{i,i_1} \cdot a_{i_1,i_2} \cdots a_{i_{k-1}j} = 1.$$

Definitions II

We can associate a directed graph $\mathcal{G}_A = (E, V)$ with the adjacency matrix A such that

1. the set of vertices $V = [n]$ and
2. the set of edges E is defined as follows: there is directed edge between vertices i, j if and only if $a_{i,j} = 1$.

In this way $i, i_1, \ldots, i_{k-1}, j$ is a chain of length of k between i and j if and only if $i, i_1, \ldots, i_{k-1}, j$ is a chain of length of k in the directed graph $G_{\mathcal{A}}$.

Definitions III

Definition 5.1

We write

1. $i \to j$ if there is a chain between i and j. Then i and j communicate. If $i \not\to j$ then i and j does not communicate.
2. $i \leftrightarrow j$ if $i \to j$ and $j \to i$.
3. i is transient if $\exists j$ such that $i \to j$ but $j \not\to i$
4. recurrent states are does which are NOT transient.
5. For a $C \subset [n]$ we say that
   1. C is irreducible if $i \leftrightarrow j$ for all $i, j \in [n]$.
   2. C is closed if $\forall i, j \in [n]$ $i \in C, j \notin C$ implies that $i \not\to j$.

If i is a recurrent state and $i \leftrightarrow j$ then j is also a recurrent state.

1. The recurrent states form classes in which everybody communicates with everybody and a member of such a clas does not communicate to anyone out of the clas. These classes are the recurrent self-communication classes.
2. Those transient states which communicate with some other states can be divided into transient classes such that any two member of such a class communicate. These are the transient communication classes.
3. There can be transient states that do not communicate with any one. They together form a class let us call it inessential class.

Path-diagram I

The path diagram for the incidens matrix $A = (a_{i,j})_{i,j=1}^{n}$:

1. Start with index 1. This is the first stage , and determine all j for which $a_{1,j} = 1$. These j's form the second stage .
2. Starting from all such j repeat the previous procedure to form stage 3 and so on.
3. Stop when an index appears second time.
4. The diagram terminates when every index which appears in the diagram has been repeated.
5. If some indices were left over start with any of them and draw a similar diagram regarding the indices of the previous diagrams as "occoured in a previous stage".

Path-diagram III

Now we follow all of these on an axample:

	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	0	0
2	1	1	1	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	1
5	0	0	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	1	0	0	0	1	0	1	0
9	0	0	0	1	0	0	0	0	1

Diagram 1

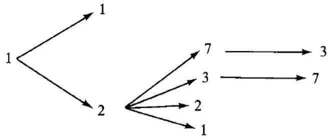
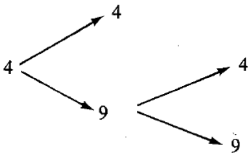


Diagram 2

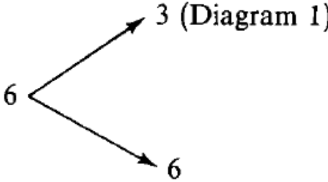


Path-diagram IV

	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	0	0
2	1	1	1	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	1
5	0	0	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	1	0	0	0	1	0	1	0
9	0	0	0	1	0	0	0	0	1

Diagram 3

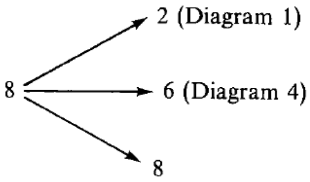
Diagram 4



Path-diagram V

	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	0	0
2	1	1	1	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	1
5	0	0	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	1	0	0	0	1	0	1	0
9	0	0	0	1	0	0	0	0	1

Diagram 5



Recurrent and tarnsient self-communication classes

1. Diagram 1 $\Longrightarrow \{3,7\}$ recurrent class, $\{1,2\}$ transient class.
2. Diagram 2 $\Longrightarrow \{4,9\}$ recurrent class,
3. Diagram 3 $\Longrightarrow \{5\}$ recurrent class,
4. Diagram 4 $\Longrightarrow \{6\}$ transient class,
5. Diagram 5 $\Longrightarrow \{8\}$ transient class,

The recurrent self-communication classes:
$\{5\}, \quad \{4,9\} \quad \{3,7\}.$
The transient self-communication classes:
$\{1,2\}, \quad \{6\}, \quad \{8\}.$

Canonical form I

So, the canonical form of the matrix on the left-hand side is the matrix on the righ-hand side.

	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	0	0
2	1	1	1	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	1
5	0	0	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	1	0	0	0	1	0	1	0
9	0	0	0	1	0	0	0	0	1

	5	4	9	3	7	1	2	6	8
5	1	0	0	0	0	0	0	0	0
4	0	1	1	0	0	0	0	0	0
9	0	1	1	0	0	0	0	0	0
3	0	0	0	0	1	0	0	0	0
7	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	1	0	0
2	0	0	0	1	1	1	1	0	0
6	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	0	1	1	1

.

Canonical form II

Assume that a matrix T has canonical form:

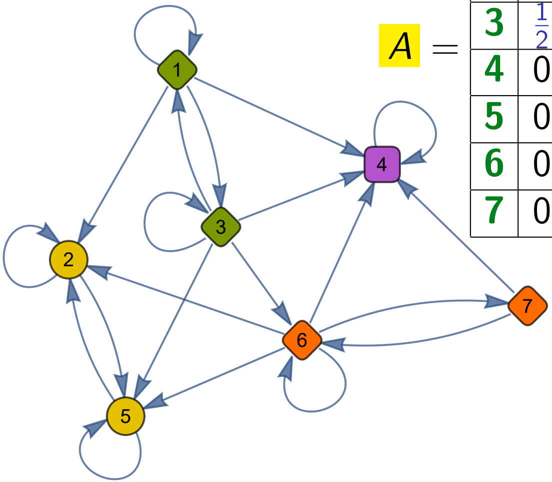
$$T = \left[\begin{array}{cccccc|c} T_1 & 0 & \dots & 0 & \dots & & 0 \\ 0 & T_2 & & & & & \vdots \\ & 0 & & & & & \\ \vdots & & & & & & \\ 0 & 0 & \dots & & T_z & & 0 \\ \hline & R & & & & & Q \end{array}\right] \qquad Q = \begin{bmatrix} Q_1 & & & \\ & Q_2 & & \\ & & \ddots & 0 \\ S & & & Q_w \end{bmatrix}$$

Canonical form III

Then the k-th power T^k of T is of the form:

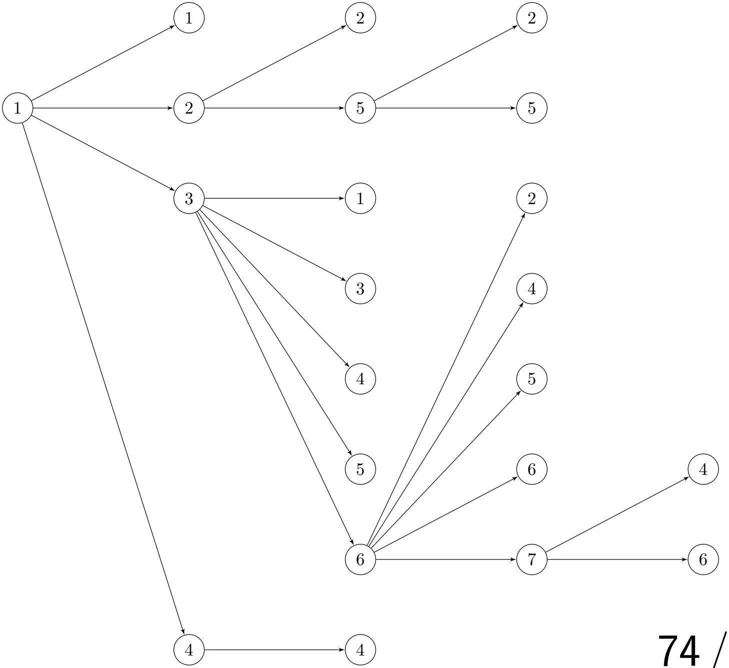
$$
T^k = \left[\begin{array}{cccc|c}
T_1^k & & & & \\
 & T_2^k & & 0 & 0 \\
0 & & \ddots & & \\
 & & & T_z^k & \\
\hline
 & R_k & & & Q^k
\end{array}\right], \quad
Q^k = \begin{bmatrix}
Q_1^k & & & \\
 & Q_2^k & & \\
 & & \ddots & 0 \\
S_k & & & Q_w^k
\end{bmatrix}
$$

Example 5.2



$A =$

	1	2	3	4	5	6	7
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0
2	0	$\frac{5}{6}$	0	0	$\frac{1}{6}$	0	0
3	$\frac{1}{2}$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	0
4	0	0	0	1	0	0	0
5	0	$\frac{1}{3}$	0	0	$\frac{2}{3}$	0	0
6	0	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
7	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0


1
1
2
3
4
2
5
1
3
4
5
6
4
2
5
2
4
5
6
7
4
6
4

Recurrent Classes: $\{2, 5\}$ and $\{4\}$

Transient Classes: $\{1, 3\}$ and $\{6, 7\}$

$T =$

	4	**2**	**5**	**6**	**7**	**3**	**1**
4	1	0	0	0	0	0	0
2	0	$\frac{5}{6}$	$\frac{1}{6}$	0	0	0	0
5	0	$\frac{1}{3}$	$\frac{2}{3}$	0	0	0	0
6	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	0	0
7	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	0	0
3	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$	$\frac{1}{4}$

$$= \left(\begin{array}{c||c} U & \mathbf{0}_{3,4} \\ \hline\hline V & W \end{array} \right)$$

That is

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \; V = \begin{pmatrix} \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix}, \; W = \begin{pmatrix} \frac{1}{2} & \frac{1}{8} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix},$$

and $\mathbf{0}_{3,4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. Let $I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, and

$$\mathbf{0}_{4,4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

$$\pi := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 2 & 6 & 1 & 3 & 4 & 5 \end{pmatrix}, \pi^{-1} := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 5 & 6 & 7 & 3 & 1 \end{pmatrix}$$

$$\Pi = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \Pi^{-1} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In the matrix Π in the i-th row the 1 is at the position $\pi(i)$. With this notation:

$$T(i,j) = A(\pi^{-1}(i), \pi^{-1}(j)), \quad A(i,j) = T(\pi(i), \pi(j)). \tag{12}$$

By the definition of matrix products we get

$$T = \Pi^{-1} \cdot A \cdot \Pi. \tag{13}$$

We know that

$$T^n = \left(\begin{array}{c||c} U^n & \mathbf{0}_{3,4} \\ \hline\hline S_n & W^n \end{array} \right). \tag{14}$$

Moreover,

1. Using that the matrix W corresponds to the transient states we get that $\lim_{n\to\infty} W^n = \mathbf{0}$.
2. We learned that $\lim_{n\to\infty} U^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix} =: B$.

So $T_\infty := \lim_{n\to\infty} T^n = \left(\begin{array}{c||c} B & \mathbf{0}_{3,4} \\ \hline\hline X & \mathbf{0}_{4,4} \end{array} \right)$, where $X = \lim_{n\to\infty} S_n$.

Using that

$$\left(\begin{array}{c||c} U & \mathbf{0} \\ \hline\hline V & W \end{array}\right) \cdot \left(\begin{array}{c||c} B & \mathbf{0}_{3,4} \\ \hline\hline X & \mathbf{0}_{4,4} \end{array}\right) = T \cdot T_\infty = T_\infty = \left(\begin{array}{c||c} B & \mathbf{0}_{3,4} \\ \hline\hline X & \mathbf{0}_{4,4} \end{array}\right)$$

We get that

$$V \cdot B + W \cdot X = X = I \cdot X. \tag{15}$$

On slide 76 we defined the matrices W, I, V, B we can compute:

$$X = (W - I)^{-1} \cdot (-V \cdot B) = \begin{pmatrix} \frac{3}{7} & \frac{8}{21} & \frac{4}{21} \\ \frac{5}{7} & \frac{4}{21} & \frac{2}{21} \\ \frac{58}{119} & \frac{122}{357} & \frac{61}{357} \\ \frac{59}{119} & \frac{40}{119} & \frac{20}{119} \end{pmatrix}. \tag{16}$$

Hence,

$$T_\infty = \left(\begin{array}{c||c} B & \mathbf{0}_{3,4} \\ \hline\hline X & \mathbf{0}_{4,4} \end{array}\right) = \left(\begin{array}{ccc||cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \hline\hline \frac{3}{7} & \frac{8}{21} & \frac{4}{21} & 0 & 0 & 0 & 0 \\ \frac{5}{7} & \frac{4}{21} & \frac{2}{21} & 0 & 0 & 0 & 0 \\ \frac{58}{119} & \frac{122}{357} & \frac{61}{357} & 0 & 0 & 0 & 0 \\ \frac{59}{119} & \frac{40}{119} & \frac{20}{119} & 0 & 0 & 0 & 0 \end{array}\right).$$

Finally we get for $A_\infty := \lim_{n\to\infty} A^n$ that

$$A_\infty = \Pi \cdot T_\infty \cdot \Pi^{-1} = \begin{pmatrix} 0 & \frac{40}{119} & 0 & \frac{59}{119} & \frac{20}{119} & 0 & 0 \\ 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{122}{357} & 0 & \frac{58}{119} & \frac{61}{357} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{8}{21} & 0 & \frac{3}{7} & \frac{4}{21} & 0 & 0 \\ 0 & \frac{4}{21} & 0 & \frac{5}{7} & \frac{2}{21} & 0 & 0 \end{pmatrix},$$

where the permutation matrices Π and Π^{-1} were defined on slide 77. This implies for example that starting from 5 after very many steps, the probability that we are at 2 is appriximately $\frac{2}{3}$ and that we are at 5 is approximately $\frac{1}{3}$.

On the following slides we state the limit theorems.

One of the important consequence of the following theorems is that under some not restrictive conditions, the same thing happens as on slide 45. That is $\lim_{n\to\infty} P^n$ exists and equal to a matrix whose all rows are equal to π.

Limit Theorems (Preparation)

- Given a Markov Chain (X_n) on a
- state space S (finite or countably infinite)
- transition matrix $\mathbf{P} = (p(i,j))_{i,j \in S}$.
- $p^m(i,j)$: the probability that starting from i we will be in j after m steps.

Definition 6.1 (Abbreviations used below)

- $\mathcal{I}$: irreducible,
- $\mathcal{A}$: aperiodic,
- $\mathcal{R}$: all states are recurrent,
- $\mathcal{S}$: $\exists \pi$ stationary distribution.

The Limit theorems below hold for countable state spaces. This means that the state space S is either countably infinite of finite.

Theorem 6.2 (Convergence Theorem)

$\mathcal{I}$ and $\mathcal{A}$ and $\mathcal{S}$ implies that Then

(a) *The MC is positive recurrent,*

(b) $\lim_{n\to\infty} p^n(i,j) = \boldsymbol{\pi}(j), \forall i, j$

(c) $\forall j, \boldsymbol{\pi}(j) > 0.$

(d) *The stationary distribution is unique.*

Theorem 6.3 (Asymptotic frequency)

$\mathcal{I}$ *and* $\mathcal{R}$ $\Longrightarrow \lim_{n\to\infty} \frac{\#\{k\leq n : X_k = j\}}{n} = \frac{1}{\mathbb{E}_j[T_j]}, \ \forall j \in S,$
where $\mathbb{E}_j\left[T_j\right]$ *is the expected time of the first return to* j*, starting from* j*.*

Theorem 6.4 (π is unique)

$\mathcal{I}$ *and* $\mathcal{S}$ $\Longrightarrow \pi(j) = \frac{1}{\mathbb{E}_j[T_j]}, \ \forall j \in S.$
In particular, π *is unique.*

Theorem 6.5

Let $f : S \to \mathbb{R}$, *s.t.* $\sum_{i \in S} |f(i)| \cdot \pi(i) < \infty$. *Then*

$$(17)\quad \mathcal{I} \text{ and } \mathcal{S} \Longrightarrow \lim_{n \to \infty} \frac{1}{n} \cdot \sum_{m=1}^{n} f(X_m) = \sum_{i \in S} f(i) \cdot \pi(i).$$

The Limit theorems below hold for finite state spaces.

Theorem 6.6 (Finite state space I)

$\#S < \infty$ and $\mathcal{I}$ and $\mathcal{A}$ then

(a) *π exists and unique,*

(b) *$\pi_i > 0$ for all $i \in S$.*

(c) *For every initial distribution $\boldsymbol{\alpha}$ on S we have*
$$\lim_{n\to\infty} \boldsymbol{\alpha}^T \cdot P^n = \pi^T$$

The proof is [7, p. 19]. If $\#S < \infty$ then the assumptions of the theorem are equivalent to P is primitive: ($\exists k$ s.t. $P^k > 0$ that is all elements of P^k are positive.)

Theorem 6.7 (Finite state space II)

$\#S < \infty$ and $\mathcal{I}$ then

- *(a) π exists and unique,*
- *(b) $\pi_i > 0$ for all $i \in S$.*
- *(c) But it is not necessarily true that for every initial distribution $\boldsymbol{\alpha}$ on S we have $\lim_{n\to\infty} \boldsymbol{\alpha}^T \cdot P^n = \pi^T$*

Notation

Let $A = (a_{ij})$ be matrix of $N \times N$. We are assuming from now on that A is nonnegative.
Hence $a_{ij} \geq 0$.

$a_{ij}^{(m)}$ denotes element (i, j) of matrix A^m.

Notation (cont.)

Definition 7.1 (Adjacency matrix of directed graphs)

Let $G = (V, E)$ be a directed graph. We denote the set of vertices by V and the set of edges by E.
The adjacency matrix of graph G (the matrix of its vertices): $A_G = (a_{ij})$

$$a_{ij} = \begin{cases} 1, & \text{if } (i,j) \in E; \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

Notation (cont.)

It is easy to see that

$$a_{ij}^{(m)} = \#\{\text{paths with length } m \text{ from } i \text{ to } j\}. \tag{19}$$

On the other hand, for every nonnegative $N \times N$ matrix A there exists a directed graph G_A in which $V(G) := \{1, \ldots, N\}$ and

$$(i,j) \in E(G) \text{ if and only if } a_{ij} > 0.$$

Notation (cont.)

Definition 7.2 (irreducible matrices)

Matrix A is irreducible, if $\forall(i,j)$, $\exists m = m(i,j)$, so that $a_{ij}^{(m)} > 0$

It's obvious that A is irreducible if and only if G_A is strongly connected, so there is a path in each direction between each pair of vertices of the graph.

Notation (cont.)

Definition 7.3 (Primitive matrices)

We say that a nonnegative matrix A is primitive, if
$\exists M : \forall i, j,\ a_{ij}^{(M)} > 0$

- If a matrix is irreducible and aperiodic then this matrix is primitive (see [7, p. 19]).
- It is easy to see that if a nonnegative matrix is irreducible and at least one of its diagonal elements is nonzero, then it is primitive.

Perron-Frobenius Theorem I

Theorem 7.4

Let A be a $N \times N$ nonnegative matrix. Then

(i) *A has eigenvalue $\lambda \in \mathbb{R}_0^+$ (so called as Perron-Frobenius eigenvalue) such that no other eigenvalues of A are greater than λ in absolute value.*

(ii) $\min_i \sum_{j=1}^{N} a_{ij} \leq \lambda \leq \max_i \sum_{j=1}^{N} a_{ij}$.

(iii) *We can choose the left and right eigenvectors $\mathbf{u}$ and $\mathbf{v}$ of λ so that all of their components are nonnegative.*

$$\mathbf{u}^T \cdot A = \lambda \mathbf{u}^T, \; A \cdot \mathbf{v} = \lambda \cdot \mathbf{v}.$$

Perron-Frobenius Theorem II

From now on we normalize **u** and **v** so that

(20) $$\sum_{i=1}^{N} u_i = 1 \quad \text{and} \quad \sum_{i=1}^{N} u_i v_i = 1.$$

If we additionally assume that *A* is irreducible, then:

(iv) λ is eigenvalue with multiplicity 1 and all elements of **u** and **v** are strictly positive.

(v) λ is the only eigenvalue for which there exists an eigenvector with only nonnegative elements.

Perron-Frobenius Theorem III

And if we assume that A is primitive, then:

(vi) $\forall i, j$:

$$\lim_{n\to\infty} \lambda^{-n} a_{ij}^{(n)} = u_j v_i, \tag{21}$$

where $\mathbf{u}, \mathbf{v}$ are the left and right eigenvectors with positive components corresponding to λ which satisfy condition (20).

Part (vi) of Perron-Frobenius Theorem comes from the Renewal Theorem.

Application for Markov chains

In our case the matrix A is te transition matrix P which is a stochastic matrix. Then all row sums are equal to 1.This implies that

- $\lambda = 1$ according to (ii) on slide 96 and
- $\mathbf{v} = (1, \ldots, 1)$.
- $\mathbf{u}^T \cdot P = \mathbf{u}^T$ by (iii) on slide 96 and by (20). That is the stationary distribution $\pi = \mathbf{u}$.

Application for Markov chains (cont.)

Then (vi) on slide 98 reads like: $\forall i, j \in S$

$$\lim_{n\to\infty} p_{i,j}^n = u_j = \pi_j, \tag{22}$$

here $p_{i,j}^n$ was defined on slide 38. So, Theorem 6.6 is a corollary of the Peron-Frobenius Therem.

Moreover, let Π be an $|S| \times |S|$ matrix, (where $|S|$ is the cardinality of S) such that all **rows** of Π are equal to π. Then

$$\lim_{n\to\infty} P^n = \Pi. \tag{23}$$

Observe that (22) is the same as (23) in terms of components. Speed of convergence:see [6, Theorem 4.9].

$\#S < \infty$, irreducible with period d

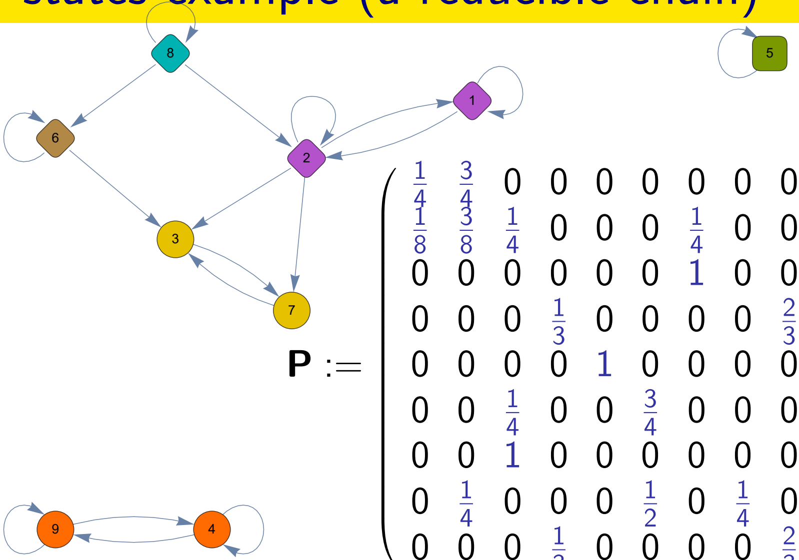
Theorem 7.5

Assume that $\#S < \infty$, P is irreducible, periodic with period $d > 1$. Then P has d eigenvalues with absolute value 1, each of them is simple. In particular 1 is a simple eigenvalue that is there is a unique invariant probability vector π corresponding to the eigenvalue 1. Let $\boldsymbol{\alpha}$ be a probability distribution on S. That is $\boldsymbol{\alpha} = (\alpha_i)_{i \in S}$ with $\sum_{i \in S} \alpha_i = 1$ and $\alpha_i \geq 0$. Then

$$\lim_{n \to \infty} \frac{1}{d} \left(\boldsymbol{\alpha}^T \cdot P^{n+1} + \cdots + \boldsymbol{\alpha}^T \cdot P^{n+d} \right) = \pi. \tag{24}$$

This Theorem is a corollary of Theorem 6.2.

A 9-states example (a reducible chain)



$$\mathbf{P} := \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{2}{3} \end{pmatrix}$$

The graph on the previous slide was prepared by Mathematics 11 using the first code. The second one gives the properties shown on the next slide

```
Graph[DiscreteMarkovProcess[2,
  {{1/4, 3/4, 0, 0, 0, 0, 0, 0, 0}, {1/8, 3/8, 1/4, 0, 0, 0, 1/4, 0, 0}, {0, 0, 0, 0, 0, 0, 1, 0, 0},
   {0, 0, 0, 1/3, 0, 0, 0, 0, 2/3}, {0, 0, 0, 0, 1, 0, 0, 0, 0}, {0, 0, 1/4, 0, 0, 3/4, 0, 0, 0},
   {0, 0, 1, 0, 0, 0, 0, 0, 0}, {0, 1/4, 0, 0, 0, 1/2, 0, 1/4, 0}, {0, 0, 0, 1/3, 0, 0, 0, 0, 2/3}}]]
```

```
In[57]:= MarkovProcessProperties[
    DiscreteMarkovProcess[2, {{1/4, 3/4, 0, 0, 0, 0, 0, 0, 0}, {1/8, 3/8, 1/4, 0, 0, 0, 1/4, 0, 0},
      {0, 0, 0, 0, 0, 0, 1, 0, 0}, {0, 0, 0, 1/3, 0, 0, 0, 0, 2/3}, {0, 0, 0, 0, 1, 0, 0, 0, 0},
      {0, 0, 1/4, 0, 0, 3/4, 0, 0, 0}, {0, 0, 1, 0, 0, 0, 0, 0, 0}, {0, 1/4, 0, 0, 0, 1/2, 0, 1/4, 0},
      {0, 0, 0, 1/3, 0, 0, 0, 0, 2/3}}]]
```

The properties of the MC in the last example

Structural Properties	
CommunicatingClasses	{3, 7}, {1, 2}, {4, 9}, {5}, {6}, {8}
RecurrentClasses	{3, 7}, {4, 9}, {5}
TransientClasses	{1, 2}, {6}, {8}
AbsorbingClasses	{5}
PeriodicClasses	{3, 7}
Periods	{2}
Irreducible	False
Aperiodic	False
Primitive	False

Continuation

This shows that $\{1,2\}$, $\{6\}$ and $\{8\}$ are transient classes. This implies that their measure by the stationary distribution must be zero. On each of the recurrent classes we have different stationary distributions which have nothing to do with each other. On the class $\{3,7\}$, $\{4,9\}$ and $\{5\}$ the stationary distributions in this order are: $\widehat{\pi} := \left(0,0,\frac{1}{2},0,0,0,\frac{1}{2},0,0\right)$.

$\widetilde{\pi} := \left(0,0,0,\frac{1}{3},0,0,0,0,\frac{2}{3}\right)$,

$\overline{\pi} := (0,0,0,0,1,0,0,0,0)$.

Let $\pi := \alpha_1\widehat{\pi} + \alpha_2\widetilde{\pi} + \alpha_3\overline{\pi}$, where $\alpha_i \geq 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Then π is one of the uncountably many stationary distributions of the chain.

Continuation

We obtained $\tilde{\pi}$ on the previous slide by the Mathematica

```
In[78]:= PDF[StationaryDistribution[
  DiscreteMarkovProcess[4, {{1/4, 3/4, 0, 0, 0, 0, 0, 0, 0}, {1/8, 3/8, 1/4, 0, 0, 0, 1/4, 0, 0},
   {0, 0, 0, 0, 0, 0, 1, 0, 0}, {0, 0, 0, 1/3, 0, 0, 0, 0, 2/3}, {0, 0, 0, 0, 1, 0, 0, 0, 0},
   {0, 0, 1/4, 0, 0, 3/4, 0, 0, 0}, {0, 0, 1, 0, 0, 0, 0, 0, 0}, {0, 1/4, 0, 0, 0, 1/2, 0, 1/4, 0},
   {0, 0, 0, 1/3, 0, 0, 0, 0, 2/3}}]], 9]

Out[78]= 2/3
```

Explanation: The very first number in the code is 4. It says that we are in the recurrence class that contains 4. The very last number is 9. This gives the measure of state 9 for that stationary distribution which is supported by the recurrence class that contains 4.

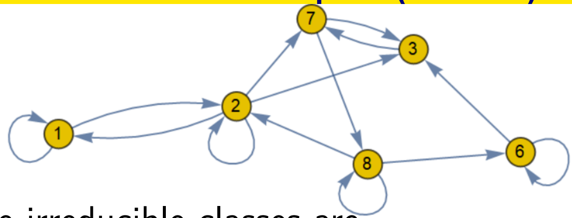
Another 9 states example

Example 7.6

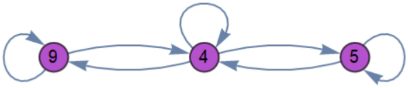
Find the all of the stationary distributions for the Markov chain given by P, where P is:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Another 9 states example (cont.)



That is the irreducible classes are $\{1, 2, 3, 6, 7, 8\}$ (above) and, $\{4, 5, 9\}$ (below). Then we run the Mathematica code on the next slide. The only thing missing from this code is the definition of the matrix p which should be defined first as P.



Another 9 states example (cont.)

```
In[43]:= pbig = p[[{1, 2, 3, 6, 7, 8}, {1, 2, 3, 6, 7, 8}]]

In[44]:= psmall = p[[{4, 5, 9}, {4, 5, 9}]]

In[47]:= invmatrep =
          Inverse[ReplacePart[pbig - IdentityMatrix[Length[pbig[[1]]]],
            {i_, Length[pbig[[1]]]} :> 1]]
         invmatrep[[Length[pbig[[1]]]]]
```

Out[48]=

$$\left\{\frac{2}{31}, \frac{4}{31}, \frac{7}{31}, \frac{4}{31}, \frac{8}{31}, \frac{6}{31}\right\}$$

```
In[49]:= invmatrep =
          Inverse[ReplacePart[psmall - IdentityMatrix[Length[psmall[[1]]]],
            {i_, Length[psmall[[1]]]} :> 1]]
         invmatrep[[Length[psmall[[1]]]]]
```

Out[50]=

$$\left\{\frac{3}{7}, \frac{2}{7}, \frac{2}{7}\right\}$$

Another 9 states example (cont.)

That is let

$$\pi^{(1)} := \left(\frac{2}{31}, \frac{4}{31}, \frac{7}{31}, 0, 0, \frac{4}{31}, \frac{8}{31}, \frac{6}{31}, 0\right),$$

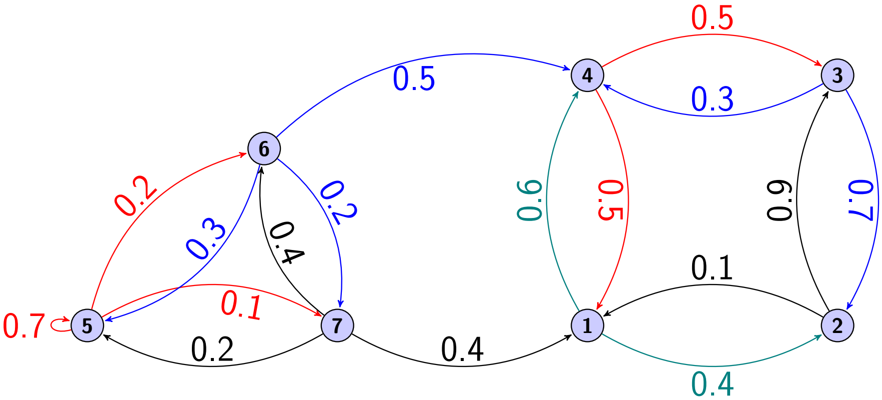
$$\pi^{(2)} := \left(0, 0, 0, \frac{3}{7}, \frac{2}{7}, 0, 0, 0, \frac{2}{7}, \right).$$

Then for every $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ with $\alpha_1, \alpha_2 > 0$ and $\alpha_1 + \alpha_2 = 1$ the vector

$$\pi = \alpha_1 \cdot \pi^{(1)} + \alpha_2 \cdot \pi^{(2)} \tag{25}$$

is a stationary distribution and all stationary distributions π can be presented of the form as in (25) for suitable α_1, α_2.

Example 7.7 (Triangle-square chain)



Transient states: $1, 2, 3$, Recurrent states: $4, 5, 6, 7$.

Example 7.7 (cont.)

It is enough to focus on the right hand side square shaped part. That is the subgraph of vertices $\{1, 2, 3, 4\}$. The transition matrix is:

$$\mathbf{P} = \begin{pmatrix} 0 & 0.4 & 0 & 0.6 \\ 0.1 & 0 & 0.9 & 0 \\ 0 & 0.7 & 0 & 0.3 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

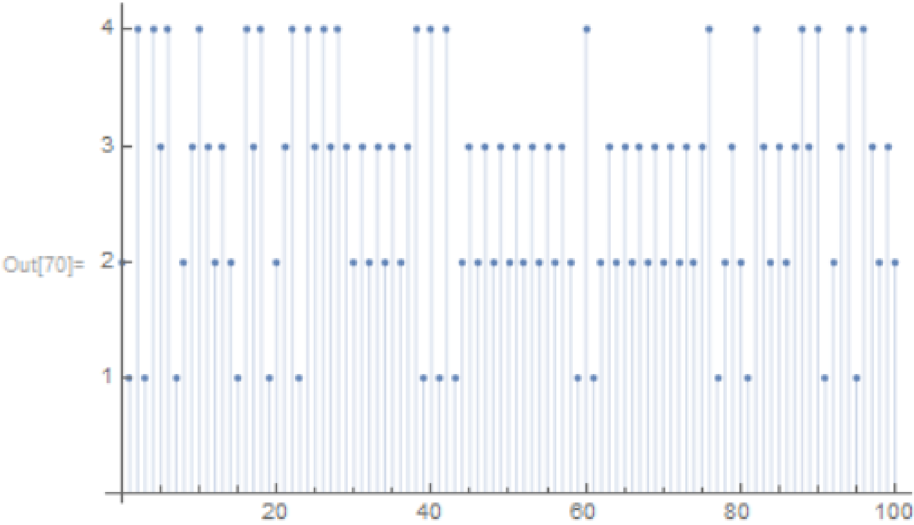
Example 7.7 (cont.)

Using the following Mathematica 11 code:

```
                                                  ( 0    0.4  0    0.6 )
In[59]:= 𝒫 = DiscreteMarkovProcess[{0, 1, 0, 0}, ( 0.1  0    0.9  0   ) ]
                                                  ( 0    0.7  0    0.3 )
                                                  ( 0.5  0    0.5  0   )

In[69]:= data = RandomFunction[𝒫, {0, 100}]

Out[69]= TemporalData[ Time: 0 to 100
                       Data points: 101    Paths: 1 ]

In[70]:= ListPlot[data, Filling → Axis, Ticks → {Automatic, {1, 2, 3, 4}}]
```

We get:

Example 7.7 (cont.)



Example 7.7 (cont.)

We get the stationary distribution

$$\pi = \left(\frac{1}{8}, \frac{5}{16}, \frac{3}{8}, \frac{3}{16}\right) \tag{26}$$

by Mathematica 11 on the next slide:

Stationary distribution with Mathematica

In[74]:= `Clear[p]`

In[75]:= $p = \begin{pmatrix} 0 & \frac{4}{10} & 0 & \frac{6}{10} \\ \frac{1}{10} & 0 & \frac{9}{10} & 0 \\ 0 & \frac{7}{10} & 0 & \frac{3}{10} \\ \frac{5}{10} & 0 & \frac{5}{10} & 0 \end{pmatrix}$

Out[75]= $\left\{\left\{0, \frac{2}{5}, 0, \frac{3}{5}\right\}, \left\{\frac{1}{10}, 0, \frac{9}{10}, 0\right\}, \left\{0, \frac{7}{10}, 0, \frac{3}{10}\right\}, \left\{\frac{1}{2}, 0, \frac{1}{2}, 0\right\}\right\}$

In[76]:=
```
invmatrep =
  Inverse[ReplacePart[p - IdentityMatrix[Length[p[[1]]]], {i_, Length[p[[1]]]} :> 1]]
```

Out[76]= $\left\{\left\{-\frac{61}{88}, -\frac{25}{176}, \frac{17}{88}, \frac{113}{176}\right\}, \left\{\frac{5}{11}, -\frac{25}{22}, -\frac{5}{11}, \frac{25}{22}\right\}, \left\{\frac{39}{88}, -\frac{85}{176}, -\frac{83}{88}, \frac{173}{176}\right\}, \left\{\frac{1}{8}, \frac{5}{16}, \frac{3}{8}, \frac{3}{16}\right\}\right\}$

In[77]:= `invmatrep[[Length[p[[1]]]]]`

Out[77]= $\left\{\frac{1}{8}, \frac{5}{16}, \frac{3}{8}, \frac{3}{16}\right\}$

$P = S \cdot D \cdot S^{-1}$

$$\mathrm{S}=\begin{pmatrix} -1 & 1 & -\frac{3\sqrt{3}}{5} & \frac{3\sqrt{3}}{5} \\ 1 & 1 & -\frac{3}{5} & -\frac{3}{5} \\ -1 & 1 & \frac{\sqrt{3}}{5} & -\frac{\sqrt{3}}{5} \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \mathrm{D}=\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{\sqrt{3}}{5} & 0 \\ 0 & 0 & 0 & \frac{\sqrt{3}}{5} \end{pmatrix}$$

Let $D_{odd}=\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; D_{even}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

$$\lim_{n\to\infty} D^{2n+1} = D_{odd}, \; \lim_{n\to\infty} D^{2n} = D_{even}$$

$\lim_{n\to\infty} P^{2n+1} = S \cdot D_{odd} \cdot S^{-1}$, and
$\lim_{n\to\infty} P^{2n} = S \cdot D_{even} \cdot S^{-1}$.

Cont.

This yields that

$$\lim_{n\to\infty} P^{2n+1} = \begin{pmatrix} 0 & \frac{5}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{5}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \end{pmatrix} \quad \lim_{n\to\infty} P^{2n} = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{5}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{5}{8} & 0 & \frac{3}{8} \end{pmatrix}$$

and $\lim_{n\to\infty} \frac{1}{2}\left(P^{2n+1} + P^{2n+1}\right) = \begin{pmatrix} \frac{1}{8} & \frac{5}{16} & \frac{3}{8} & \frac{3}{16} \\ \frac{1}{8} & \frac{5}{16} & \frac{3}{8} & \frac{3}{16} \\ \frac{1}{8} & \frac{5}{16} & \frac{3}{8} & \frac{3}{16} \\ \frac{1}{8} & \frac{5}{16} & \frac{3}{8} & \frac{3}{16} \end{pmatrix}$. Note that all row vectors are the same and identical to π.

Inventory chain Durrett, Example 1.6

s, S **storage strategy**:

- Given $s < S$
- Let X_n be the amount of stock on hand at the end of day n.

Strategy:

- If $X_n \leq s$ we fill up the stock during the night so that the stock at the beginning of day $n+1$ is S.
- If $X_n > s$ we do not do anything.

Inventory chain Durrett, Example 1.6 (cont.)

Let D_{n+1} be the demand of this item on day $n+1$.

Using the $x^+ := \max\{x, 0\}$ notation:

$$X_{n+1} = \begin{cases} (X_n - D_{n+1})^+, & \text{if } X_n > s; \\ (S - D_{n+1})^+, & \text{if } X_n \leq s. \end{cases}$$

Inventory chain Durrett, Example 1.6 (cont.)

In an example with $s = 1, S = 5$ and
$\mathbb{P}(D_{n+1} = 0) = 0.3,\ \mathbb{P}(D_{n+1} = 1) = 0.4$
$\mathbb{P}(D_{n+1} = 2) = 0.2,\ \mathbb{P}(D_{n+1} = 3) = 0.1$

	0	**1**	**2**	**3**	**4**	**5**
0	0	0	0.1	0.2	0.4	0.3
1	0	0	0.1	0.2	0.4	0.3
2	0.3	0.4	0.3	0	0	0
3	0.1	0.2	0.4	0.3	0	0
4	0	0.1	0.2	0.4	0.3	0
5	0	0	0.1	0.2	0.4	0.3

Inventory chain Durrett, Example 1.6 (cont.)

For $s = 1$ and $S = 5$ the stationary distribution is:

$$\pi = \left\{ \frac{177}{1948}, \frac{379}{2435}, \frac{225}{974}, \frac{105}{487}, \frac{98}{487}, \frac{1029}{9740} \right\}$$

Assume that the profit of every single item is \$12, but the daily storage fee is \$2.

Question:

- What is the long-term profit on this item for the previous choice of s, S?

Inventory chain Durrett, Example 1.6 (cont.)

- How should we choose values of s, S to maximize the profit?

Repair chain

A machine has 3 critical components which can go wrong, but the machine operates until all of them stops working. If at least two components are broken, they get repaired for the next day. We assume that on a single day maximum 1 component can go wrong, and the probability of component 1, 2 and 3 failing is (in order) 0.01, 0.02 and 0.04.

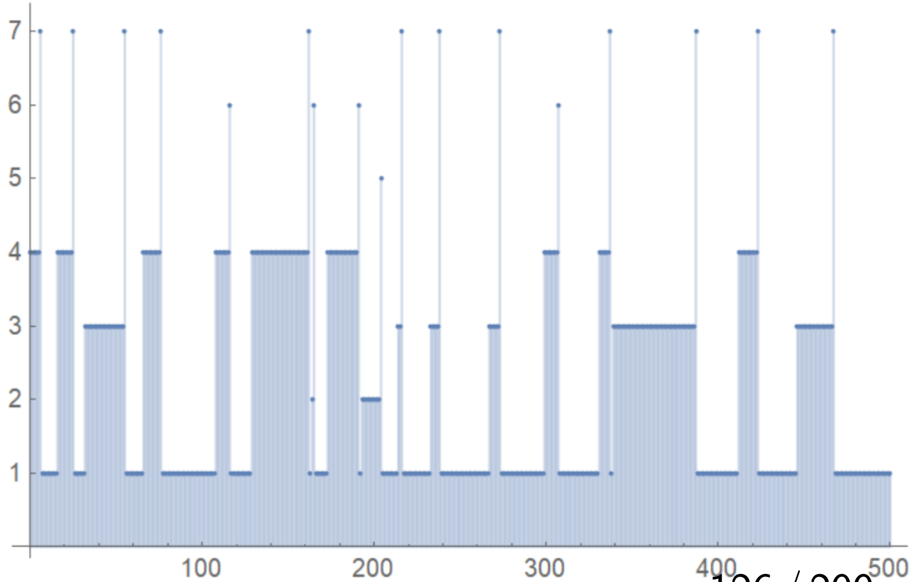
If we are to model this process with a Markov chain, it is recommended to use state space of broken parts: $\{0, 1, 2, 3, 12, 13, 23\}$. The transition matrix is:

Repair chain (cont.)

	0	**1**	**2**	**3**	**12**	**13**	**23**
0	0.93	0.01	0.02	0.04	0	0	0
1	0	0.94	0	0	0.02	0.04	0
2	0	0	0.95	0	0.01	0	0.04
3	0	0	0	0.97	0	0.01	0.02
12	1	0	0	0	0	0	0
13	1	0	0	0	0	0	0
13	1	0	0	0	0	0	0

Question: How many components are used of type $1, 2$ and 3 in 1000 days?

Repair chain (cont.)



Repair c

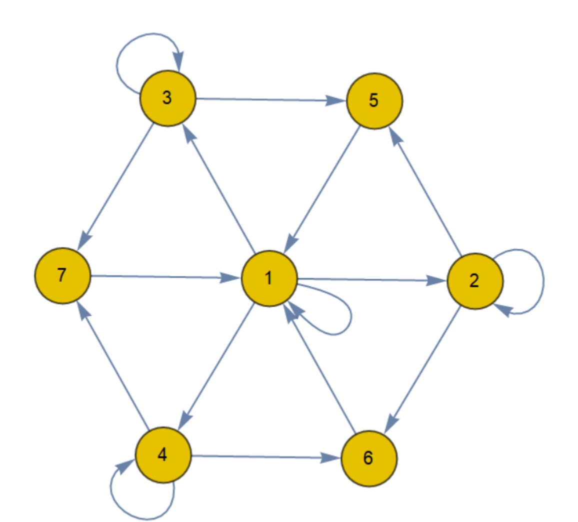


Figure: Prepared with Wolfram mathematica

Repair cha

Structural Properties	
CommunicatingClasses	{1, …, 7}
RecurrentClasses	{1, …, 7}
TransientClasses	None
AbsorbingClasses	None
PeriodicClasses	None
Periods	{}
Irreducible	True
Aperiodic	True
Primitive	True

Figure: Prepared with Wolfram mathematica

Repair chain (cont.)

Stationary distribution:

$$\pi = (0.336, 0.056, 0.134, 0.448, 0.002, 0.006, 0.014)$$

Mean first passage matrix:

$$\begin{pmatrix} 0. & 279.333 & 127.5 & 39.9167 & 404. & 147.5 & 68.6094 \\ 17.6667 & 0. & 145.167 & 57.5833 & 286.667 & 66.1667 & 86.276 \\ 21. & 300.333 & 0. & 60.9167 & 344. & 168.5 & 33.9219 \\ 34.3333 & 313.667 & 161.833 & 0. & 438.333 & 132.333 & 56.5365 \\ 1. & 280.333 & 128.5 & 40.9167 & 0. & 148.5 & 69.6094 \\ 1. & 280.333 & 128.5 & 40.9167 & 405. & 0. & 69.6094 \\ 1. & 280.333 & 128.5 & 40.9167 & 405. & 148.5 & 0. \end{pmatrix}$$

Wright-Fisher model

Example 7.8

A (fixed size) generation consists of $2N$ genes with type either a or A. If there are $j \in \{0, \dots, 2N\}$ a-type gene in the parent population, then the next generaton's building will be determined with $2N$ independent binomial trials, with probabilities $p_j = \frac{j}{2N}$, $q_j = 1 - \frac{j}{2N}$. So, if X_n is the number of a-type genes in the n^{th} generation, then the appropriate Markov-chain is:
$\mathbb{P}(X_{n+1} = k | X_n = j) = p(j, k) = \binom{2N}{k} p_j^k q_j^{2N-k}$.

The transition matrix for the Wright-Fisher model when $2N = 6$

$$\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{15625}{46656} & \frac{3125}{7776} & \frac{3125}{15552} & \frac{625}{11664} & \frac{125}{15552} & \frac{5}{7776} & \frac{1}{46656} \\
\frac{64}{729} & \frac{64}{243} & \frac{80}{243} & \frac{160}{729} & \frac{20}{243} & \frac{4}{243} & \frac{1}{729} \\
\frac{1}{64} & \frac{3}{32} & \frac{15}{64} & \frac{5}{16} & \frac{15}{64} & \frac{3}{32} & \frac{1}{64} \\
\frac{1}{729} & \frac{4}{243} & \frac{20}{243} & \frac{160}{729} & \frac{80}{243} & \frac{64}{243} & \frac{64}{729} \\
\frac{1}{46656} & \frac{5}{7776} & \frac{125}{15552} & \frac{625}{11664} & \frac{3125}{15552} & \frac{3125}{7776} & \frac{15625}{46656} \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

In the Wright-Fisher model above we have **absorbing states** when $x = 0$ and $x = 2N$. This means that if the process ever reaches one of these states, it remains there forever.

We modify the model so that there will be no absorbing state:

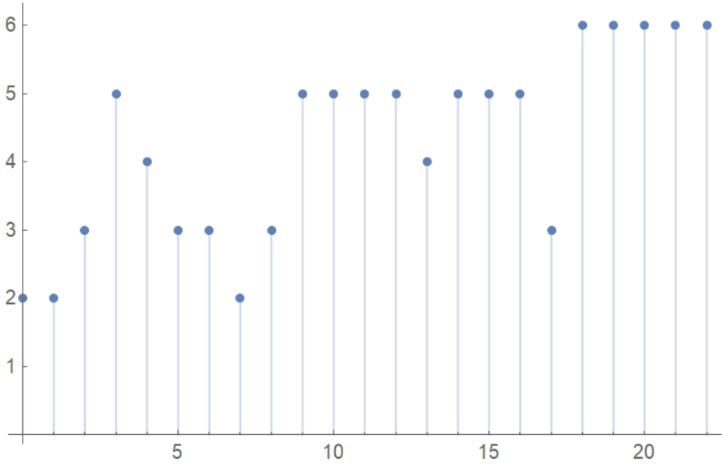


Figure: Simulation for the Wright-Fisher model, $2N = 6$, starting from 2

```
𝒫 = DiscreteMarkovProcess[3,
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{15625}{46656} & \frac{3125}{7776} & \frac{3125}{15552} & \frac{625}{11664} & \frac{125}{15552} & \frac{5}{7776} & \frac{1}{46656} \\ \frac{64}{729} & \frac{64}{243} & \frac{80}{243} & \frac{160}{729} & \frac{20}{243} & \frac{4}{243} & \frac{1}{729} \\ \frac{1}{64} & \frac{3}{32} & \frac{15}{64} & \frac{5}{16} & \frac{15}{64} & \frac{3}{32} & \frac{1}{64} \\ \frac{1}{729} & \frac{4}{243} & \frac{20}{243} & \frac{160}{729} & \frac{80}{243} & \frac{64}{243} & \frac{64}{729} \\ \frac{1}{46656} & \frac{5}{7776} & \frac{125}{15552} & \frac{625}{11664} & \frac{3125}{15552} & \frac{3125}{7776} & \frac{15625}{46656} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

```
]

In[218]:=
data = RandomFunction[𝒫, {0, 22}]
Out[218]=
TemporalData[ Time: 0 to 22  Data points: 23  Paths: 1 ]

In[219]:=
ListPlot[data - 1, Filling → Axis,
 Ticks → {Automatic, {0, 1, 2, 3, 4, 5, 6, 7}}]
```

Figure: Mathematica code for the Wright-Fisher model, $2N = 6$, starting from 2

Wright-Fisher model with mutations

Example 7.9

In this model every gene can mutate before creating the new generation. An a can mutate into A with probability α_1 and the reverse side has probability α_2.
In this case the transition matrix is the same, but now, for the mutation, the probabilities are modified.

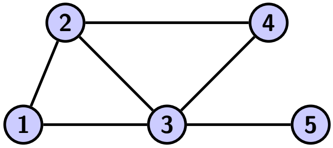
$$p_j = \frac{j}{2N}(1-\alpha_1) + \left(1 - \frac{j}{2N}\right)\alpha_2,$$

and

$$q_j = \frac{j}{2N}\alpha_1 + \left(1 - \frac{j}{2N}\right)(1-\alpha_2).$$

Simple RW on simple graphs

Example 7.10



	1	2	3	4	5
1	0	1/2	1/2	0	0
2	1/3	0	1/3	1/3	0
3	1/4	1/4	0	1/4	1/4
4	0	1/2	1/2	0	0
5	0	0	1	0	0

Simple graph and the transition matrix of the corresponding simple random walk (RW) on this graph. From every vertex we move to a uniformly chosen neighbour. (Described more precisely on the next slide.)

Let $G = (V, E)$ be a simple graph (no loops, no double edges), where as usual, V is the set of vertices and E is the set of edges. We denote the degree of vertex $x \in V$ by $\deg(x)$. The simple random walk on G is Markov chain on state space S which is defined by the following transition matrix:

$$p(x,y) = \begin{cases} \frac{1}{\deg(x)}, & (x,y) \in E; \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

Example 7.10 (Cont.)

Using the mathematica 11 code on the next slide we obtain that the stationary distribution:
$\pi = \left(\frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{1}{6}, \frac{1}{12}\right)$ (the last command on the next slide results the 5-th component of π). The mean first passage matrix is $M = (m_{i,j})_{i,j=1}^{5}$, where $m_{i,j}$ is the expected number (≥ 1) of steps to get from i to j for the first time.

$$M = \begin{pmatrix} 6 & \frac{11}{4} & \frac{9}{4} & 6 & \frac{53}{4} \\ \frac{19}{4} & 4 & \frac{5}{2} & \frac{19}{4} & \frac{27}{2} \\ \frac{21}{4} & \frac{7}{2} & 3 & \frac{21}{4} & 11 \\ 6 & \frac{11}{4} & \frac{9}{4} & 6 & \frac{53}{4} \\ \frac{25}{4} & \frac{9}{2} & 1 & \frac{25}{4} & 12 \end{pmatrix}$$

Mean First Passage Time Matrix

$M = (m_{i,j})$ and we know the diagonal: $m_{i,i} = \frac{1}{\pi_i}$. In general we need to solve the system of equations for all $i \neq j$:

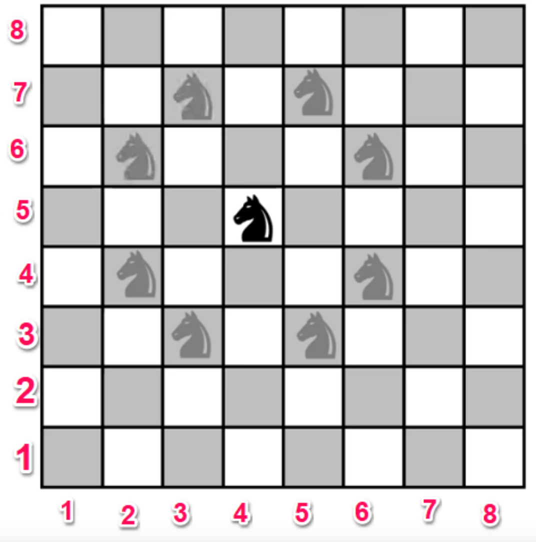
$$m_{i,j} = p_{i,j} \cdot 1 + \sum_{k \neq j} p_{i,k} \cdot (1 + m_{k,j}) = 1 + \sum_{k \neq j} p_{i,k} \cdot m_{k,j}.$$

Example 7.10 (Cont.)

```
In[173]:= 𝒫 = DiscreteMarkovProcess[{0, 0, 1, 0, 0}, 
  ( 0    1/2  1/2  0    0   )
  ( 1/3  0    1/3  1/3  0   )
  ( 1/4  1/4  0    1/4  1/4 )
  ( 0    1/2  1/2  0    0   )
  ( 0    0    1    0    0   ) ]

Out[173]= DiscreteMarkovProcess[{0, 0, 1, 0, 0},
  {{0, 1/2, 1/2, 0, 0}, {1/3, 0, 1/3, 1/3, 0}, {1/4, 1/4, 0, 1/4, 1/4}, {0, 1/2, 1/2, 0, 0}, {0, 0, 1, 0, 0}}]

In[174]:= 𝒟 = FirstPassageTimeDistribution[𝒫, 4];
          Mean[𝒟]

Out[175]= 21/4

In[176]:= PDF[StationaryDistribution[𝒫], 5]

Out[176]= 1/12
```

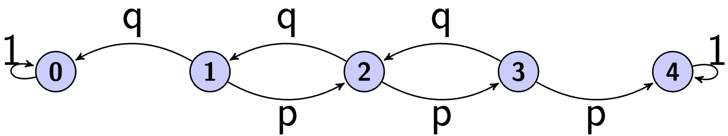
The second and third commands computes the value $m_{3,4} = \frac{21}{4}$. The last command yields that $\pi(5) = \frac{1}{12}$.

Knight moves on chessboard



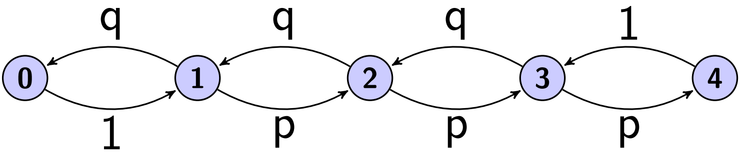
Simple RW on the graph G, where $G = (E, V)$:
$V := \{1, \ldots, 8\}^2$,
and for $(i_1, j_1), (i_2, j_2) \in V$
$((i_1, j_1), (i_2, j_2)) \in E$
iff either:
$|i_1 - i_2| = 2 \& |j_1 - j_2| = 1$
or
$|j_1 - j_2| = 2 \& |i_1 - i_2| = 1$

RW with absorbing boundary :



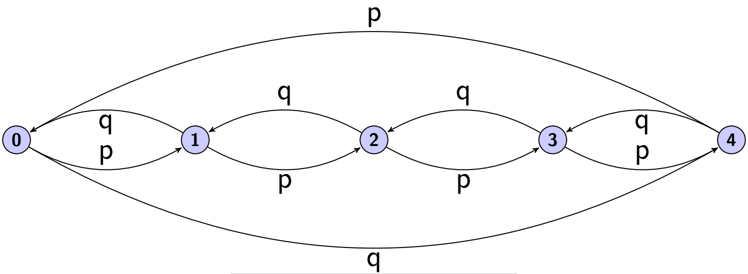
	0	1	2	3	4
0	1	0	0	0	0
1	q	0	p	0	0
2	0	q	0	p	0
3	0	0	q	0	p
4	0	0	0	0	1

RW with reflecting boundary



	0	1	2	3	4
0	0	1	0	0	0
1	q	0	p	0	0
2	0	q	0	p	0
3	0	0	q	0	p
4	0	0	0	1	0

RW with periodic boundary conditions



	0	**1**	**2**	**3**	**4**
0	0	p	0	0	q
1	q	0	p	0	0
2	0	q	0	p	0
3	0	0	q	0	p
4	p	0	0	q	0

The history of Branching Processes

In 1873 Francis Galton asked in Educational Times: what is the probability of dying off of a name, a family dying agnatically? Reverend Henry William Watson answered it and they published a paper together in 1874: On the probability of extinction of families. Thus the correspondent MC is called Galton-Watson process. So we only regard the number of sons in various generations, because they carry on the name.

Branching processes

Let's regard a population, in which the 0^{th} generation only consists of one person and in the n^{th} generation one gives birth to k children (who will be counted in the $(n+1)^{st}$ generation) with probability p_k (independently of each other); with $k = 0, 1, 2, \dots$

Let X_n be the number of individuals in the n^{th}generation. The state space is $\mathbb{N} = \{0, 1, 2, \dots\}$. If $Y_1, Y_2, \dots$ are i.i.d. random variables for which $\mathbb{P}(Y_m = k) = p_k$, then the transition matrix is $p(0,0) = 1$ and

Branching processes (cont.)

$$p(i,j) := \mathbb{P}(Y_1 + \cdots + Y_i = j) \text{ if } i > 0 \text{ and } j \geq 0,$$

Special case: The number of children has geometric distribution.

$$p_\ell := \mathbf{P}(\text{number of children} = \ell) = q^\ell p.$$

Then element (k, l) of the transition matrix:

$$p(k,\ell) = \binom{k+\ell-1}{\ell} p^n q^k.$$

Random walks on $\mathbb{Z}^d$

Simple symmetric random walk on $S = \mathbb{Z}^d$:

$$p(x,y) := \begin{cases} \frac{1}{2d}, & \text{ha } \|\mathbf{x}-\mathbf{y}\| = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

General random walk on $S = \mathbb{Z}^d$:
$\mathfrak{p} : \mathbb{Z}^d \to [0,1]$; $\sum_{\mathbf{x}\in\mathbb{Z}^d} \mathfrak{p}(\mathbf{x}) = 1$, and the transition matrix $\mathbf{P} = (p(x,y))$:

$$p(x,y) := \mathfrak{p}(\mathbf{x}-\mathbf{y}).$$

Two stage Markov chains

In this example X_{n+1} is dependent of (X_{n-1}, X_n).

Basketball chain

Consider a basketball player who makes a shot with the following probabilities:

1/2, if both of his previous shots are missed
2/3, if he has hit one of his last two shots
3/4, if he has hit both of his last two shots.

So let $X_n = S$ denote the success and $X_n = M$ denote the miss.

The state space is: $\{SS, SM, MS, MM\}$ and the transition matrix is:

Two stage Markov chains (cont.)

	SS	SM	MS	MM
SS	$3/4$	$1/4$	0	0
SM	0	0	$2/3$	$1/3$
MS	$2/3$	$1/3$	0	0
MM	0	0	$1/2$	$1/2$

Explanation: If $(X_{n-1}, X_n) = (S, M)$, then the probability of $(X_n, X_{n+1}) = (M, S)$ is equal to $2/3$.

Stationary distribution for the Basketball chain

Following the rule shown above to compute stationary distribution $\boldsymbol{\pi}$, we subtract 1 from transition matrix $\mathbf{P}$'s diagonal elements and replace the last column with ones.

$$A = \begin{bmatrix} -1/4 & 1/4 & 0 & 1 \\ 0 & -1 & 2/3 & 1 \\ 2/3 & 1/3 & -1 & 1 \\ 0 & 0 & 1/2 & 1 \end{bmatrix}$$

Stationary distribution for the Basketball chain (cont.)

Then $A^{-1} = \begin{pmatrix} -\frac{13}{6} & -\frac{5}{16} & \frac{11}{16} & \frac{43}{24} \\ -\frac{1}{6} & -\frac{17}{16} & -\frac{1}{16} & \frac{31}{24} \\ -1 & -\frac{3}{8} & -\frac{3}{8} & \frac{7}{4} \\ \frac{1}{2} & \frac{3}{16} & \frac{3}{16} & \frac{1}{8} \end{pmatrix}$.

Its last row is $\boldsymbol{\pi}$. Hence,

$$\boldsymbol{\pi} = \left(\frac{1}{2}, \frac{3}{16}, \frac{3}{16}, \frac{1}{8}\right).$$

Stationary distribution for the Basketball chain (cont.)

Reminder: the order of components is (SS,SM,MS,MM). (S: success, M: miss.) So, in the long term the ratio of successes is:

$$\pi_{SS} + \pi_{KS} = \pi_1 + \pi_3 = \frac{1}{2} + \frac{3}{16} = \frac{11}{16}.$$

Stationary distribution with Mathematica

In[51]:= `Clear[p, n]`

In[52]:= $p = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

Out[52]= $\left\{\left\{\frac{3}{4}, \frac{1}{4}, 0, 0\right\}, \left\{0, 0, \frac{2}{3}, \frac{1}{3}\right\}, \left\{\frac{2}{3}, \frac{1}{3}, 0, 0\right\}, \left\{0, 0, \frac{1}{2}, \frac{1}{2}\right\}\right\}$

In[53]:= `n = Length[p[[1]]]`

Out[53]= 4

In[54]:= `matrep = ReplacePart[p - IdentityMatrix[n], {i_, n} :> 1]`

Out[54]= $\left\{\left\{-\frac{1}{4}, \frac{1}{4}, 0, 1\right\}, \left\{0, -1, \frac{2}{3}, 1\right\}, \left\{\frac{2}{3}, \frac{1}{3}, -1, 1\right\}, \left\{0, 0, \frac{1}{2}, 1\right\}\right\}$

In[55]:= `invmatrep = Inverse[matrep]`

Out[55]= $\left\{\left\{-\frac{13}{6}, -\frac{5}{16}, \frac{11}{16}, \frac{43}{24}\right\}, \left\{-\frac{1}{6}, -\frac{17}{16}, -\frac{1}{16}, \frac{31}{24}\right\}, \left\{-1, -\frac{3}{8}, -\frac{3}{8}, \frac{7}{4}\right\}, \left\{\frac{1}{2}, \frac{3}{16}, \frac{3}{16}, \frac{1}{8}\right\}\right\}$

In[56]:= `invmatrep[[n]]`

Out[56]= $\left\{\frac{1}{2}, \frac{3}{16}, \frac{3}{16}, \frac{1}{8}\right\}$

Ehrenfest chain: Stationary distribution

Example 7.11 (π for the Ehrenfest chain)

Recall the definition of the Ehrenfest chain: Consider the Markov Chain with state space $S := \{0, 1, 2, \ldots, n\}$ and

1. It jumps from 0 to 1 and from n to $n-1$ with probability 1.
2. For any $0 < i < n$, it jumps from i to $i-1$ with probability i/n and from i to $i+1$ with probability $1 - \frac{i}{n}$.

Ehrenfest chain: Stationary distribution (cont.)

Now compute the stationary state for this chain. The transition matrix:

$$\mathbf{P} := \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{n} & 0 & \frac{n-1}{n} & 0 & \cdots & 0 \\ 0 & \frac{2}{n} & 0 & \frac{n-2}{n} & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \frac{n-1}{n} & 0 & \frac{1}{n} \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Ehrenfest chain: Stationary distribution (cont.)

For $\boldsymbol{\pi}^T \cdot \mathbf{P} = \boldsymbol{\pi}^T$, thus using notation $\pi_{-1} := \pi_{n+1} := 0$ we obtain that:

$$\pi_{k-1}\left(1 - \frac{k-1}{n}\right) + \pi_{k+1}\frac{k+1}{n} = \pi_k, \quad k = 0, 1, \dots, n. \tag{29}$$

We introduce the generating function:

$$g(x) = \sum_{k=0}^{n} x^k \pi_k. \tag{30}$$

Ehrenfest chain: Stationary distribution (cont.)

Multiply both sides of (29) by n and x^k, then sum it up for k from 1 to n:

$$\sum_{k=1}^{n}(n-k+1)x^k\pi_{k-1}+\sum_{k=0}^{n-1}\pi_{k+1}(k+1)x^k=n\underbrace{\sum_{k=0}^{n}x^k\pi_k}_{g(x)}.$$

By obvious manipulations of this formula we obtain:

$$(1+x)g'(x)=ng(x).$$

Ehrenfest chain: Stationary distribution (cont.)

After solving this differential equation we get:

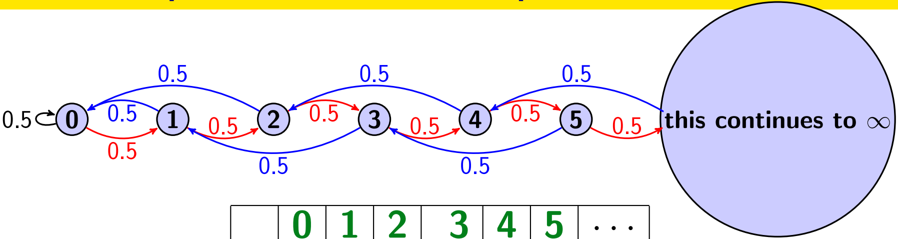
$$g(x) = C(1+x)^n.$$

Using that π is a probability vector we get $g(1) = 1$. Hence $C = 2^{-n}$. That is:

$$g(x) = 2^{-n}(1+x)^n = 2^{-n}\sum_{k=1}^{n}\binom{n}{k}\cdot x^k \tag{31}$$

Compare this to (30) to realize that $\pi_k = 2^{-n}\binom{n}{k}$.

Two steps back, one step forward chain



	0	**1**	**2**	**3**	**4**	**5**	$\cdots$
0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	$\cdots$
1	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0	$\cdots$
2	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	0	$\cdots$
3	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	$\cdots$
4	0	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$\boldsymbol{\pi}$ for the two steps back one step ahead chain:

From the equation $\boldsymbol{\pi}^T \cdot \mathbf{P} = \boldsymbol{\pi}^T$: $\pi_0 = \frac{1}{2}(\pi_0 + \pi_1 + \pi_2)$ and $\forall k \geq 1 : \pi_k = \frac{1}{2}(\pi_{k-1} + \pi_{k+2})$. From these two equations it comes by induction that

$$\forall k \geq 0 : \pi_k = \pi_{k+1} + \pi_{k+2}. \tag{32}$$

It is obviously satisfied by $\pi_k = (1-\rho)\rho^k$, $k \geq 0$, where ρ is the golden ratio: $\rho = \frac{\sqrt{5}-1}{2}$. Homework: there is no other stationary distribution. So the process spends most of its time (more than 99%) in the set $\{0, 1, \ldots, 9\}$.

Definition 7.12

A MC is doubly stochastic if its probability matrix's column sum equals to 1. $\sum_j p(i,j) = 1$, $\forall j$.

Theorem 7.13

A MC with finite state space is doubly stochastic iff its stationary distribution is the uniform distribution.

Proof.

Let us assume that $\#S = N$, then

$$\sum_x \pi(x) p(x,y) = \frac{1}{N} \sum_x p(x,y) = \frac{1}{N} = \pi(x).$$

Examples

Example 7.14 (Random walk with periodic boundary conditions)

Recall the definition of the random walk with periodic boundary conditions from slide 144. It is obviously doubly stochastic.

Modulo 6 jumps on a circle

Example 7.15

We roll the finite number series $0, 1, 2, \ldots, 5$ on to a circle so that 5 and 0 be neighbours. Then we use such a regular dice which has number

- 1 on three sides,
- 2 on two sides,
- 3 on one side.

We move forward as much as we scored (modulo 6).

Modulo 6 jumps on a circle (cont.)

The transition matrix is:

	0	**1**	**2**	**3**	**4**	**5**
0	0	1/2	1/3	1/6	0	0
1	0	0	1/2	1/3	1/6	0
2	0	0	0	1/2	1/3	1/6
3	1/6	0	0	0	1/2	1/3
4	1/3	1/6	0	0	0	1/2
5	1/2	1/3	1/6	0	0	0

It can easily be seen that the elements of transition matrix's third power $\mathbf{P}^3$ are positive. Thus we see that

Modulo 6 jumps on a circle (cont.)

the chain is irreducible and aperiodic, so the conditions of Theorem 6.2 are satisfied (obviously $\pi(i) = 1/6, \forall i$).

Recurrence of the simple Symmetric random walk in $\mathbb{Z}^d$-ben

Theorem 8.1

In $\mathbb{R}^d$ the simple symmetric random walk is recurrent (zero recurrent) if $d = 1$ or $d = 2$ but transient for $d \geq 3$.

For the proof in the case of $d = 1$ see the discussion starting on slide **??** in 2018_File_BB.

One needs to be careful

In the case of countably infinite state space it can happen that there are no recurrent states as the following trivial example shows

Example 8.2 (Monotone increasing MC)

Let S be the set of non-negative integers and $p(i, i+1) := 1$ for all $i \in S$.

Detailed balance condition

π satisfies detailed balance condition, if $\forall x, y$

$$\pi(x)p(x,y) = \pi(y)p(y,x) \tag{33}$$

If we sum both sides for y, we get that

$$\sum_y \pi(y)p(y,x) = \pi(x)\underbrace{\sum_y p(x,y)}_{=1} = \pi(x).$$

So, if a probability measure satisfies formula (33), then it is a stationary distribution. There exist stationary

Detailed balance condition (cont.)

distributions which do not satisfy the detailed balance condition (33). For example, consider the MC whose probability matrix is:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.3 & 0.1 & 0.6 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}.$$

Then the stationary distribution π of P does not satisfy (33). To get contradiction, assume that π satisfies (33). From this and from the fact that $p(1,3) = 0$ we get

Detailed balance condition (cont.)

$\pi(3) = 0$. This and formula (33) yield that $\pi(2) = \pi(1) = 0$ which is impossible. On the other hand, $\mathbf{P}$ is a doubly stochastic matrix for which there is a stationary distribution (the uniform distribution): $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

So, it can happen that there is a stationary distribution but it does not satisfy (33). In spite of this, if we have a guess about a probability vector that it could be the stationary distribution, we can check it easily by substituting it into formula (33).

Reversible MC

Now we use [6, chapter 1.6]. **Notation**: For the MC (X_n) we introduce

$$X_0^n := (X_0, \dots, X_n). \tag{34}$$

So for $\mathbf{x} := (x_0, \dots, x_n)$

$$\{X_0^n = \mathbf{x}\} = \{X_0 = x_0, \dots, X_n = x_n\} \tag{35}$$

and for an $\mathbf{x} = (x_0, \dots, x_n)$ let

$$\overleftarrow{\mathbf{x}} := (x_n, x_{n-1}, \dots x_1, x_0). \tag{36}$$

Reversible MC (cont.)

It comes easily from formula (33) that:

(37)

$$\boldsymbol{\pi}(x_0)p(x_0, x_1)\cdots p(x_{n-1}, x_n) = \boldsymbol{\pi}(x_n)p(x_n, x_{n-1})\cdots p(x_1, x_0).$$

Using notation $\mathbf{x} = (x_0, \dots, x_n)$ this implies that:

$$\mathbb{P}_{\boldsymbol{\pi}}\left(X_0^n = \mathbf{x}\right) = \mathbb{P}_{\boldsymbol{\pi}}\left(X_0^n = \overleftarrow{\mathbf{x}}\right). \tag{38}$$

So if MC (X_n) has stationary distribution, **and** it satisfies **detailed balance condition**, then the distribution of $(X_0, \dots, X_n)$ is the same as the distribution of $(X_n, \dots, X_0)$.

Reversible MC (cont.)

Definition 9.1 (reversible MC)

A MC X_n is reversible if it has stationary distribution $\boldsymbol{\pi}$ and $\boldsymbol{\pi}$ satisfies the detailed balance condition, that is formula (33) holds.

Reversible MC (cont.)

Example 9.2 (Simple random walk on graphs, slide 137)

Let us regard a simple random walk on graph $G = (V, E)$. Using notation of slide 137, the stationary distribution is: $\pi(y) = \deg(y)/2\#E$. It can be easily seen (homework) that it satisfies detailed balance condition:

$$\pi(x)p(x,y) = \pi(y)p(y,x), \quad \forall x, y \in S.$$

Reversible MC (cont.)

Example 9.3 (Random walk with periodic boundary condition (slide 144))

Reminder: finite state space (with cardinality N) rolled onto a circle. We jump 1 clockwise with probability p and anticlockwise with probability $q = 1 - p$. The chain is double stochastic, so $\boldsymbol{\pi} = (\frac{1}{N}, \ldots, \frac{1}{N})$. But

$\pi(k)p(k, k+1) = \frac{p}{N}$ and $\frac{q}{N} = \pi(k+1)p(k+1, k)$ and they are equal only if $p = q$. So in other instances the detailed balance condition is not satisfied.

Definition 9.4 (Chain with reversed time)

Given an irreducible MC X_n with transition matrix $\mathbf{P}$ and stationary distribution $\boldsymbol{\pi}$. Let us define the matrix $\widehat{\mathbf{P}} = (\widehat{p}(x,y))$:

$$\widehat{p}(x,y) := \frac{\pi(y)p(y,x)}{\pi(x)}. \tag{39}$$

Then $\widehat{\mathbf{P}}$ is a stochastic matrix (every element is non-negative, the row-sums are 1.) So $\widehat{\mathbf{P}}$ determines a MC $(\widehat{X}_n)$, which we call time reversal of (X_n).

Obviously, if (X_n) is reversible, then $\mathbf{P} = \widehat{\mathbf{P}}$.

Time reversal

Theorem 9.5

Using notation of Definition 9.4:

(a) π *is stationary distribution not only for* (X_n) *but for* $(\widehat{\mathbf{X}}_n)$, *too, and*

(b) *for all* $\mathbf{x}$*:*

$$\mathbf{P}_\pi\left(X_0^n = \mathbf{x}\right) = \mathbf{P}_\pi\left(\widehat{X}_0^n = \overleftarrow{\mathbf{x}}\right), \tag{40}$$

where $\overleftarrow{\mathbf{x}}$ *was defined in* (36).

Time reversal (cont.)

Proof.

Firstly we prove part (a):

$$\sum_y \pi(y)\widehat{p}(y,x) = \sum_y \pi(y)\frac{\pi(x)p(x,y)}{\pi(y)} = \pi(x).$$

Now we see part (b):

$$\begin{aligned} \mathbb{P}_\pi\left(X_0^n = \mathbf{x}\right) &= \pi(x_0)p(x_0,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n) \\ &= \pi(x_n)\widehat{p}(x_n,x_{n-1})\cdots\widehat{p}(x_2,x_1)\widehat{p}(x_1,x_0) \\ &= \mathbb{P}_\pi\left(\widehat{X}_0^n = \overleftarrow{\mathbf{x}}\right). \end{aligned}$$

Birth and death processes

Birth and death processes are those MCs, whose state space are

$$S := \{k, k+1, \dots, n\}.$$

and we cannot jump more than 1. So the possible jumps are: $-1, 0, 1$. The transition probability:

$$p(x, y) = 0 \text{ if } |x - y| > 1 :$$

Birth and death processes (cont.)

Then the transition matrix **P** is:

$$\begin{aligned} p(x, x+1) &= p_x \text{ if } x < n \\ p(x, x-1) &= q_x \text{ if } x > k \\ p(x, x) &= 1 - p_x - q_x \text{ if } k \le x \le n. \end{aligned}$$

and all other $p(x, y) = 0$. Warning: $p + q \neq 1$ is possible here!

Theorem 9.6

All birth and death processes are reversible.

Birth and death processes (cont.)

Proof

We need to see that we can find a probability measure π on S which satisfies formula (33), thus for $x < n$ it must be true for π:

$$\pi(x+1)\underbrace{p(x+1,x)}_{q_{x+1}} = \pi(x)\underbrace{p(x,x+1)}_{p_x}$$

So, for (33), it is needed that

$$\pi(x+1) = \frac{p_x}{q_{x+1}}\pi(x). \tag{41}$$

Iterating this for every $1 \leq i \leq n-k$

Proof Cont.

$$\pi(k+i) = \pi(k) \cdot \underbrace{\frac{p_{k+i-1} \cdot p_{k+i-2} \cdots p_{k+1} \cdot p_k}{q_{k+i} \cdot q_{k+i-1} \cdots q_{k+2} \cdot q_{k+1}}}_{r_i} \tag{42}$$

It is easy to see that if we choose $\pi(k)$ such way that

$$\pi(k) \cdot \left(1 + \sum_{i=1}^{n-k} r_i\right) = 1, \tag{43}$$

then π is a stationary distribution which satisfies the detailed balance condition, so the chain is reversible.

We have computed the stationary distribution for the Ehrenfest Chain (see slides 12 and 155). We got that $\pi(k) = 2^{-N}\binom{N}{k}$, but we needed an unpleasant reduction involving generator functions. Now we can easily get this from formula (42) because the Ehrenfest Chain is obviously a birth and death process.

Example 9.7 (π for the Ehrenfest chain (see file A, slide 12))

Here: $S = \{0, 1, \ldots, N\}$. From formula (42) we get that

$$r_i = \binom{N}{i} \text{ if } 1 \le i \le N.$$

Using that $1 + \sum_{i=1}^{N} r_i = 2^N$ we obtain that for $i = 0, \ldots, N$:

$$\pi(i) = 2^{-N} \binom{N}{i}.$$

Two year collage

Example 10.1 (Two year collage)

At a two year collage the first year students are called freshmen the second year students are the sophomores.

- **Freshmen:** 60% of them become sophomores , 25% of them remain freshmen, 15% of them exit (E) so leave the school.
- **Sophomores:** 70% of them complete the courses with Success (S), 20% of them remain sophomores and 10% of them exit.

Two year collage (cont.)

Then if $S = \{1, 2, G, D\}$ (freshmen, sophomores, Graduate, Drop out) and X_n shows that a student is in which state after n years, then X_n is a MC whose state space is S and its transition matrix:

	1	2	G	D
1	0.25	0.6	0	0.15
2	0	0.2	0.7	0.1
G	0	0	1	0
D	0	0	0	1

Two year collage (cont.)

Let $h(x)$, $x \in S$ be the probability that a student in state x eventually graduates. Then we apply the one step reasoning method. Namely, we do not know $h(1)$ and $h(2)$ but after making one step on the chain the following equations hold:

$$h(1) = 0.25h(1) + 0.6h(2)$$
$$h(2) = 0.2h(2) + 0.7.$$

From this $h(2) = 7/8$ and $h(1) = 0.7$.

Theorem 10.2

A MC is given with a finite state space S. Let $a, b \in S$ and $C := S \setminus \{a, b\}$. Let $h : S \to \mathbb{R}^+$ be a function satisfying:

(44)

$$h(a) = 1, \; h(b) = 0, \quad \forall x \in C : \; h(x) = \sum_{y \in S} p(x,y) h(y).$$

Put

$$V_y = \min\{n \geq 0 : X_n = y\}.$$

Assume that $\forall x \in C$*:* $\mathbb{P}_x(V_a \wedge V_b < \infty) = 1$*. Then*

$$h(x) = \mathbb{P}_x(V_a < V_b).$$

Proof

We frequently use the shorthand notation

$$a \wedge b := \min\{a, b\}.$$

Let $T := V_a \wedge V_b$. By assumption

$$\forall x \in C,\ \mathbb{P}_x(T < \infty) = 1. \tag{45}$$

First we express the probability $\mathbb{P}_x(V_a < V_b)$ in terms of the expectation of a random variable. Namely, note that by definition,

$$h(X_T) = \begin{cases} 1, & \text{if } V_a < V_b\ ; \\ 0, & \text{if } V_b < V_a. \end{cases}$$

Proof (cont.)

That is

$$h(X_T) = \mathbb{1}_{\{V_a < V_b\}} \tag{46}$$

Hence, for all $x \in C$ we have

$$\mathbb{P}_x\left(V_a < V_b\right) = \mathbb{E}_x\left[h(X_T)\right] \tag{47}$$

Now we prove that

$$\mathbb{E}_x\left[h(X_T)\right] = \lim_{n\to\infty} \mathbb{E}_x\left[h(X_{T\wedge n})\right]. \tag{48}$$

Proof (cont.)

To see this, recall that we assumed that the state space $\#S < \infty$. So, $M := \max_{x \in S} h(x) < \infty$, That is, on the one hand, for all $x \in C$,

$$h(X_{T \wedge n}) < M \text{ holds for all } n. \tag{49}$$

On the other hand, using (45) (which says that T is almost surely finite) we have that

$$\lim_{n \to \infty} h(X_{T \wedge n}) = h(X_T). \tag{50}$$

Proof (cont.)

Putting together (50) and (49), we obtain that (48) holds by Lebesgue Dominated Convergence Theorem. Finally, we verify that

$$\mathbb{E}_x h(X_{T\wedge n}) = h(x), \ \forall n > 1, \ \forall x \in C. \tag{51}$$

$u_1(x,a) := \mathbb{P}_x\left(X_1 = a\right) = p(x,a)$ and for $k \geq 2$

$$\begin{aligned} u_k &= \mathbb{P}_x\left(X_k = a, T = k\right) \cdot \underbrace{h(a)}_{1} \\ &= \sum_{x_1,\dots,x_{k-1}\in C} p(x,x_1)p(x_1,x_2)\cdots p(x_{k-1},a). \end{aligned}$$

Proof (cont.)

Moreover, let $S_0 := h(x)$ and

$$S_k := \sum_{x_1,\dots,x_k \in C} p(x,x_1)p(x_1,x_2)\cdots p(x_{k-1},x_k)h(x_k).$$

A careful case analysis yields that by (44) for $k \geq 1$:

(52) $$S_k = S_{k-1} - u_k.$$

Observe that for a $k \leq n$ we have

(53) $$\mathbb{E}_x\left[h(X_{T\wedge n}), T = k\right] = \mathbb{P}_x\left(X_k = a, T = k\right).$$

Proof (cont.)

Using (52), (53), a telescoping sum in the third step and the fact that $S_0 = h(x)$ we obtain:

$$\mathbb{E}_x[h(X_{T\wedge n})] = \mathbb{E}_x[h(X_{T\wedge n}); T > n] + \sum_{k=1}^{n} \mathbb{E}_x\left[h(X_{T\wedge n}), T=k\right]$$

$$(54) \qquad = S_n + \sum_{k=1}^{n} u_k$$

$$(55) \qquad = \underbrace{h(x)}_{S_0} + \sum_{k=1}^{n} (\underbrace{S_k - S_{k-1}}_{-u_k}) + \sum_{k=1}^{n} u_k$$

$$(56) \qquad = h(x). \blacksquare$$

Wright-Fisher model, see slide 130

The state space: $S = \{0, 1, \ldots, 2N\}$. The absorbing states: 0 and $2N$. Question: what is the probability of ending up in $2N$, or in the model's language: what is the probability that once every gene becomes type a?
The transition matrix:

$$p(x, y) = \underbrace{\binom{2N}{y} \left(\frac{x}{2N}\right)^{y} \left(1 - \frac{x}{2N}\right)^{N-y}}_{\text{Binomial}(2N, x/2N)}.$$

That is: the distribution of $y \in \{0, 1, \ldots, 2N\}$ where the Markov chain jumps to from $x \in \{0, 1, \ldots, 2N\}$ is a

Wright-Fisher model, see slide 130 (cont.)

$\mathrm{Binomial}(2N, x/2N)$ random variable. We know that expected value of a $\mathrm{Binomial}(2N, x/2N)$ r.v. is equal to x. The same in formula:

$$x = \sum_{y=0}^{2N} p(x,y) \cdot y \tag{57}$$

Let us define a function: $h(t) := \frac{t}{2N}$, then by (57):

$$h(x) = \sum_{y=0}^{2N} p(x,y) h(y).$$

Wright-Fisher model, see slide 130 (cont.)

Let $a = 2N$ and $b = 0$. Then $h(a) = 1$ and $h(b) = 0$. Obviously:

$$\mathbb{P}_x(V_a \wedge V_b < \infty) > 0, \quad \forall 0 < x < N.$$

So, we can use Theorem 10.2, thus we get:

$$\mathbb{P}_x(V_{2N} < V_0) = h(x) = \frac{x}{N}. \blacksquare$$

In summary: here we guessed the exit probability function $h(x)$ and to verify our guess we used Theorem 10.2.

Example: Gambler's ruin, unfair case

Now we use the notation introduced on slide 4, where the Gambler's ruin example was introduced with the modification that now $p \neq 1/2$ is arbitrary. Let

$$h(x) = \mathbb{P}_x\left(V_N < V_0\right).$$

That is $h(x)$ is the probability that a gambler starting with \$$x$ eventually wins, that is reaches \$$N$ earlier than \$0. Obviously, $h(N) = 1$ and $h(0) = 0$. As usual, let

Example: Gambler's ruin, unfair case (cont.)

$q := 1 - p$ and let $0 < x < N$. Yet again we use the one-step argument: After one step:

$$X_{n+1} = \begin{cases} x+1, & \text{with probability } p; \\ x-1, & \text{with probability } q. \end{cases}$$

So, for $0 < x < N$:

$$h(x) = ph(x+1) + qh(x-1). \tag{58}$$

Example: Gambler's ruin, unfair case (cont.)

Obvious manipulations yield:

$$p\left(h(x+1)-h(x)\right)=q\left(h(x)-h(x-1)\right).$$

Hence,

$$h(x+1)-h(x)=\frac{q}{p}\left(h(x)-h(x-1)\right) \tag{59}$$

Example: Gambler's ruin, unfair case (cont.)

Let $c := h(1) - h(0)$. So, from formula (59) for $x \geq 1$

$$h(x) - h(x-1) = c \left(\frac{q}{p}\right)^{x-1}. \tag{60}$$

Using that $h(N) = 1$, $h(0) = 0$ and a telescopic sum in the second step and (60) in the last step:

$$1 = h(N) - h(0) = \sum_{x=1}^{N} h(x) - h(x-1) = c \sum_{x=1}^{N} \left(\frac{q}{p}\right)^{x-1}.$$

Example: Gambler's ruin, unfair case (cont.)

Put $\theta = q/p$. Then $c = (1-\theta)/(1-\theta^N)$. So

$$h(x) = h(x) - h(0) = c \sum_{i=0}^{x-1} \theta^i = \frac{1-\theta^x}{1-\theta^N}. \tag{61}$$

From here if $N \to \infty$ we get that

$$p > \frac{1}{2} \Rightarrow \mathbb{P}_x\left(V_0 = \infty\right) = 1 - \left(\frac{q}{p}\right)^x. \tag{62}$$

Example: Gambler's ruin, unfair case (cont.)

Corollary 10.3

Consider a random walk on $\mathbb{Z}$, in which starting from all $x > 0$ we go forward one step with probability $p > \frac{1}{2}$ and we go backward one step with probability $q = 1 - p$. Then the probability that starting from an arbitrary $x > 0$ we never reach 0 is $1 - \left(\frac{q}{p}\right)^x > 0$. That is every state is transient.

Example: Gambler's ruin, fair case

We consider the Gambler's ruin example with $p = 1/2$.
We use the unfair case $(p \neq 1/2)$'s notation. The argument is the same until formula (59).
But in case of $p = 1/2$ formula (59) shows that the gradient of function $h(x)$ is constant and $h(0) = 0$, $h(N) = 1$ so if $p = 1/2$

$$\mathbb{P}_x(V_N < V_0) = h(x) = x/N.$$

Tennis

The following problem is from [3, p.44].

Tennis (cont.)

Example 10.4

In tennis a player wins the game if either she gets 4 points when the other player has not more than 2 points. If the score is $4-3$ then the winner is the player who makes a two pints advantage first. Assume that

- The server wins the point with 0.6 probability,
- Successive points are independent.

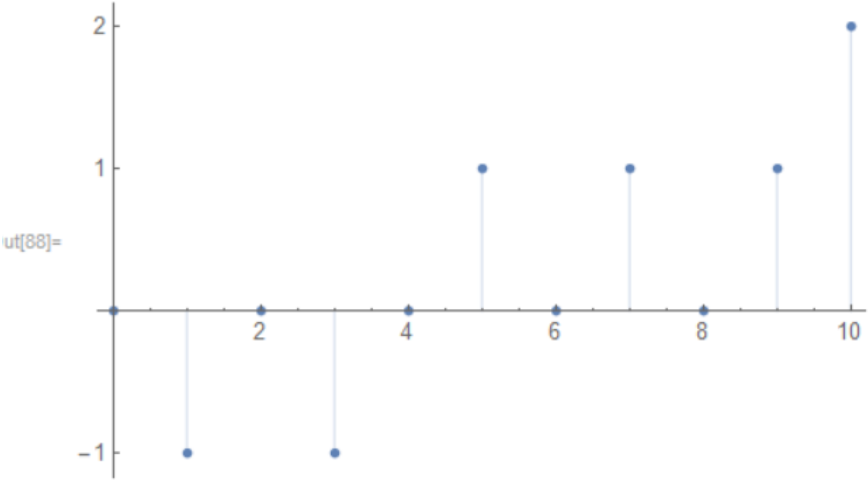
Question: What is the probability that the server wins if the score now is $3-3$?

Tennis (cont.)

Solution: Let X_n be the difference of the points scored from the point of the server after $3-3$ until one of the player has a 2 point advantage so that the game ends. That is the state space is $S := \{-2, -1, 0, 1, 2\}$. Then the transition matrix:

	2	**1**	**0**	**-1**	**-2**
2	1	0	0	0	0
1	0.6	0	**0.4**	0	0
0	0	**0.6**	0	**0.4**	0
-1	0	0	**0.6**	0	0.4
-2	0	0	0	0	1

A simulation for the Tennis starting from 0



A simulation for the Tennis starting from 0

```
Clear[𝒫, p]
```

$$p = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{4}{10} & 0 & \frac{6}{10} & 0 & 0 \\ 0 & \frac{4}{10} & 0 & \frac{6}{10} & 0 \\ 0 & 0 & \frac{4}{10} & 0 & \frac{6}{10} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

```
𝒫 = DiscreteMarkovProcess[3, p]
```

```
In[87]:= data = RandomFunction[𝒫, {0, 10}]

Out[87]= TemporalData[ Time: 0 to 10  Data points: 11  Paths: 1 ]
```

```
In[88]:= ListPlot[data - 3, Filling → Axis, Ticks → {Automatic, {-2, -1, 0, 1, 2}}]
```

Tennis

Let $h(x)$ be the probability that the server wins when staring from $X_0 = x$. Obviously now the absorbing states are $\{-2, 2\}$ and $C = \{-1, 0, 1\}$. Clearly,

$$h(2) = 1 \text{ and } h(-2) = 0.$$

From the one-step reasoning:

$$h(x) = \sum_y p(x, y)h(y), \quad \forall x \in C. \tag{63}$$

Tennis (cont.)

$$
\begin{aligned}
(64) \qquad h(1) &= 0.6 \cdot \underbrace{h(2)}_{1} + 0.4h(0) = 0.4h(0) + 0.6 \\
h(0) &= 0.6h(1) + 0.4h(-1) \\
h(-1) &= 0.6h(0) + 0.4 \cdot \underbrace{h(-2)}_{0} = 0.6h(0).
\end{aligned}
$$

Let $\mathbf{R} = (r(x, y))_{x,y \in C}$ be the restriction of matrix $\mathbf{P}$ to rows and columns of C, and let $\widehat{\mathbf{h}}$ be the vector which

Tennis (cont.)

we get by ignoring those coordinates of **h** which are outside C. Then formula (64):

$$\widehat{\mathbf{h}} - \mathbf{R} \cdot \widehat{\mathbf{h}} = \begin{bmatrix} 0.6 \\ 0 \\ 0 \end{bmatrix} \quad (65)$$

Which is:

$$\underbrace{\begin{bmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 0 \end{bmatrix}}_{I-R} \cdot \begin{bmatrix} h(1) \\ h(0) \\ h(-1) \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0 \\ 0 \end{bmatrix}$$

Tennis (cont.)

So

$$\begin{bmatrix} h(1) \\ h(0) \\ h(-1) \end{bmatrix} = (I - R)^{-1} \cdot \begin{bmatrix} 0.6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.8769 \\ 0.6923 \\ 0.4154 \end{bmatrix}$$

Exit time from the two year collage

Consider the two-year collage example on slide 189. There,we asked what was the probability of a $k = 1, 2$-year-student of graduating ever. Now, for the same example we ask:

Question: On average, how much time is needed for a student to get out of the school either by completing it successfully or drop out (unsuccessfully).

Let $g(x)$ be the expected number of years that an $x \in \{1, 2\}$-year student leaves the school either because

Exit time from the two year collage (cont.)

she graduates or because she drops out. We define $g(G) = g(D) = 0$. Again, we use the one-step reasoning:

$$\begin{aligned} g(1) &= 1 + 0.25g(1) + 0.6g(2) \\ g(2) &= 1 + 0.2g(2). \end{aligned}$$

This yields: $g(2) = 1.25$ and $g(1) = 2.333$.

Exit time

Theorem 10.5

Let X_n be a MC with a finite state space S. Let $A \subset S$ and $C := S \setminus A$, and $V_A := \min \{n \geq 0 : X_n \in A\}$. Let $g : S \to \mathbb{R}^+$ be a function which satisfies:

(a) $\mathbb{P}_x (V_A < \infty) > 0$, $\forall x \in C$,

(b) $g(a) = 0$, $\forall a \in A$,

Exit time (Cont.)

Theorem 10.5 (Cont.)

(c) $\forall x \in C$

$$g(x) = 1 + \sum_{y} p(x, y) g(y). \tag{66}$$

Then this function g is the expected exit time. That is

$$g(x) = \mathbb{E}_x [V_A]. \tag{67}$$

Proof.

The proof goes similarly as the proof of Theorem 10.2. □

Waiting time for TT

Example 10.6

We flip a fair coin until we get two Tails (TT) in a row. Question: what is the expected value of the number of flips?

Solution: We call T the Tails and H the Heads. Let T_{TT} be the (random) number of flips until we get the two Tails (the TT). Now we associate a MC (X_n) with state space $S := \{0, 1, 2\}$, where X_n is the number of consecutive Tails after the n^{th} flip. So, if the n^{th} flip

Waiting time for TT (cont.)

results in a Head, then $X_n = 0$, if it is a Tail, then $X_n = 1$ or $X_n = 2$ depending on X_{n-1} (if it was Head or Tail). State 2 is absorbing because we only flip the coin until this happens. So, the transition matrix:

	0	**1**	**2**
0	1/2	1/2	0
1	1/2	0	1/2
2	0	0	1

Waiting time for TT (cont.)

Let

$$V_2 := \min\{n \geq 0 : X_n = 2\} \text{ and } g(x) := \mathbb{E}_x[V_2].$$

Then from the one-step reasoning:

$$\begin{aligned} g(0) &= 1 + 0.5g(0) + 0.5g(1) \\ g(1) &= 1 + 0.5g(0). \end{aligned} \tag{68}$$

Waiting time for TT (cont.)

Let $\mathbf{1}$ be the vector in $\mathbb{R}^2$, having both components equal to 1. Then $g(0) = 0$ by formula (68):

(69) $$(I - R) \cdot \hat{\mathbf{g}} = \mathbf{1},$$

where, as before, R is the matrix we get from $\mathbf{P}$ by deleting the rows and columns corresponding to the absorbing states (now the only absorbing state is 2) and $\hat{\mathbf{g}}$ is the vector we get from vector $\mathbf{g}$ by deleting the

Waiting time for TT (cont.)

components belonging to the absorbing states which is 2 as mentioned before. Hence from (69) we get

$$\hat{\mathbf{g}} = \begin{pmatrix} g(0) \\ g(1) \end{pmatrix} = (I - R)^{-1} \cdot \mathbf{1} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \cdot \mathbf{1} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

So, by Theorem 10.5, we have

$$\mathbb{E}_0 [V_2] = g(0) = \hat{g}(0) = 6.$$

Tennis at $3-3$

Consider the Tennis problem on slide 209 again.
Question: How long the game lasts if now the score is $4-3$, $3-3$ and $3-4$ from the point of the server?
Solution: Let $g(x)$ be the expected time of the game if $x \in \{1, 0, -1\}$. As we discussed, the absorbing states are $A := \{-2, 2\}$ and the state space is $S := \{-2, -1, 0, 1, 2\}$. So, $C : A \setminus A = \{1, 0, -1\}$.
Using notation analogue to the previous problem:

$$\mathbf{R} = \begin{bmatrix} 0 & 0.4 & 0 \\ 0.6 & 0 & 0.4 \\ 0 & 0.6 & 0 \end{bmatrix}$$

Tennis at $3-3$ (cont.)

and from here:

$$I - R = \begin{bmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 1 \end{bmatrix}$$

So, like the previous problem:

$$\begin{pmatrix} g(1) \\ g(0) \\ g(-1) \end{pmatrix} = (I-R)^{-1}\mathbf{1} = \begin{bmatrix} 19/13 & 10/13 & 4/13 \\ 15/13 & 25/13 & 10/13 \\ 9/13 & 15/13 & 19/13 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Tennis at $3-3$ (cont.)

So, at $3-3$ the expected play-time:

$$g(0)=\frac{15+25+10}{13}=3.846. \tag{70}$$

Tennis at $3-3$ (cont.)

Remark 10.7

Consider an absorbing MC with state space S, absorbing states A and transient states $C := S \setminus A$. Let $y \in C$ and we denote the total number of visit to y including the time 0 if we started from y by $N(y)$. The $N(y) = \sum_{n=0}^{\infty} \mathbb{1}_{X_n=y}$. In this way

Tennis at $3-3$ (cont.)

Remark 10.7 (Cont.)

$$\mathbb{E}_x\left[N(y)\right] = \sum_{n=0}^{\infty} R^n(x,y) = (I-R)^{-1}(x,y). \tag{71}$$

Let T be the duration until the chain gets into an absorbing state. This is equal to the total time the MC spends at all of the transient states together. That is

$$T = \sum_{y \in C} N(y).$$

Hence by (71)

Tennis at $3-3$ (cont.)

Remark 10.7 (Cont.)

$$(72)\, \mathbb{E}_x[T] = \sum_{y\in C} \mathbb{E}_x[N(y)] = \sum_{y\in C} (I-R)^{-1}(x,y),$$

which is the x-th component of the vector

$$(I-R)^{-1} \cdot \mathbf{1}.$$

With this argument we proved that $(I-R)^{-1}(x,y)$ is equal to the expectation of the number of visits to y (counting the initial state if $x=y$) starting from x.

Tennis at $3-3$ (cont.)

As a Corollary of this Remark we can see that in (70) the summands

$$\frac{15}{13}, \quad \frac{25}{13}, \quad \frac{10}{13}$$

are the expected number of cases when the score is $1, 0, -1$ respectively, before the game ends.

Gambler's ruin, $p = 1/2$: How long does it last?

So: $p(i, i+1) = p(i, i-1) = 1/2$. $A := \{0, N\}$,

$$V_A := \min\{n \geq 0 : X_n \in A\}.$$

Let $g(x) := \mathbb{E}_x[V_A]$. Obviously

(73) $$g(0) = g(N) = 0.$$

If $0 < x < N$:

$$g(x) = 1 + \frac{1}{2}g(x+1) + \frac{1}{2}g(x-1)$$

Gambler's ruin, $p = 1/2$: How long does it last? (cont.)

$$g(x+1) - g(x) = g(x) - g(x-1) - 2.$$

If $c = g(1) = g(1) - g(0)$, then

$$g(k) - g(k-1) = c - 2(k-1) \tag{74}$$

Gambler's ruin, $p = 1/2$: How long does it last? (cont.)

Using that $g(N) = 0$ and summing the previous equations for $1 \leq k \leq N$, we get telescopic sums. From these:

$$\begin{aligned} 0 &= g(N) = \sum_{k=1}^{N} (g(k) - g(k-1)) \\ &= \sum_{k=1}^{N} (c - 2(k-1)) = cN - 2\frac{N(N-1)}{2}. \end{aligned}$$

Gambler's ruin, $p = 1/2$: How long does it last? (cont.)

Hence, $c = N - 1$. Substituting this back to formula (74) and summing it up we obtain that:

$$g(x) = x(N - x).$$

Gambler's ruin, $p \neq 1/2$: How long does it last?

So, in this case: $p(i, i+1) = p \neq 1/2$ and $p(i, i-1) = 1 - p =: q$. Let $A := \{0, N\}$, $C := \{1, \ldots, N-1\}$,

$$V_A := \min\{n \geq 0 : X_n \in A\}.$$

Let $g(x) := \mathbb{E}_x [V_A]$. Obviously

(75) $$g(0) = g(N) = 0.$$

Gambler's ruin, $p \neq 1/2$: How long does it last? (cont.)

From the one-step reasoning:

$$g(x) = 1 + p \cdot g(x+1) + q \cdot g(x-1), \quad x \in C. \tag{76}$$

These are $N-1$ equations for the $N-1$ unknowns: $(g(1), \ldots, g(N-1))$. This system of equation is the same that appeared in formula (66). Thus, from Theorem 10.5 its solution can only be $g(x) = \mathbb{E}_x[V_A]$.

Gambler's ruin, $p \neq 1/2$: How long does it last? (cont.)

We can easily check that $g(1), \ldots, g(N-1)$ is the solution of the system of equation (76) if

$$g(x) = \frac{x}{q-p} - \frac{N}{q-p} \cdot \frac{1-(q/p)^x}{1-(q/p)^N}, \quad 0 < x < N-1.$$

So, the expected time of the game for $0 < x < N-1$:

$$\mathbb{E}_x\left[V_A\right] = \frac{x}{q-p} - \frac{N}{q-p} \cdot \frac{1-(q/p)^x}{1-(q/p)^N}. \tag{77}$$

Gambler's ruin, $p \neq 1/2$: How long does it last? (cont.)

From now we always assume that $x \in C$. We use (77) and distinguish two cases: if $p < q$, then

$$\lim_{N \to \infty} \frac{N}{1-(q/p)^N} = 0 \text{ thus } g(x) \approx \frac{x}{q-p}. \tag{78}$$

On the other hand, if $p > q$, then $(q/p)^N \to 0$, thus

$$g(x) \approx \frac{N-x}{p-q}\left[1-(q/p)^x\right] + \frac{x}{p-q}(q/p)^x. \tag{79}$$

This Subsection is based on Charles M. Grinstead, J. Laurie Snell's book. [4]. Click here for the book.

In this Section (unless we say otherwise) X_n is supposed to be an absorbing MC on a finite state space S with

- transition matrix P,
- absorbing states $A \subset S$ and
- transient states $C := S \setminus A$.

We write $a := \#A$ and $c := \#C$.

We will answer the following questions in general terms:

Questions answered on this Subsection in general terms

(Q1) What is the probability that the process will end up in a given absorbing state? (Theorem 10.11.)

(Q2) What is expected exit time (expectation of the time to get to any of the absorbing states)? (Theorem 10.9.)

(Q3) What is the expected number of visits to a transient state before finally getting to an absorbing state. (Theorem 10.8.)

We always assume that the $c + a$ states of S are arranged as follows: the first c states are the transient states and the last a states are the absorbing states. Then the transition matrix P is in the canonical form:

(80)

	C	**A**
C	$\mathbf{R}$	$\mathbf{Q}$
A	$\mathbf{0}_{a,c}$	$\mathbf{I}_a$

, that is $P = \begin{pmatrix} \mathbf{R} & \mathbf{Q} \\ \mathbf{0}_{a,c} & \mathbf{I}_a \end{pmatrix}$.

where

- R is a $c \times c$ matrix,
- Q is a non-zero $c \times a$ matrix
- $\mathbf{0}_{a,c}$ is an $a \times c$ zero matrix (all elements are zero),
- $\mathbf{I}_a$ is an $a \times a$ identity matrix,

The powers of P

Clearly,

(81)
$$P^n = \begin{pmatrix} \mathbf{R}^n & \star \\ \mathbf{0}_{a,c} & \mathbf{I}_a \end{pmatrix},$$

where $\star$ is a $c \times a$ matrix. We have actually proved that

(82)
$$\lim_{n \to \infty} \mathbf{R}^n = \mathbf{0}_{c,c}$$

The following Theorem answers question Q3. In special cases we have already seen its proof. Alternatively, for the proof see [4, p. 418, Theorem 11.4].

The fundamental matrix

Theorem 10.8

As always in this Subsection, we assume that X_n is an absorbing MC. Then

(a) $\mathbf{I}_c - \mathbf{R}$ *has an inverse* $\mathbf{N} := (\mathbf{I}_c - \mathbf{R})^{-1}$ *which is called the fundamental matrix.*

(b) $\mathbf{N} = \mathbf{I}_c + \mathbf{R} + \mathbf{R}^2 + \mathbf{R}^3 + \cdots$.

(c) $\mathbf{N} = (n_{i,j})_{i,j=1}^{c}$ *then* $n_{i,j}$ *is the expected values of the times the chain starting from* $i \in C$ *visits* $j \in C$ *before the absorbtion happens. Initial state is counted if* $i = j$.

Time to absorption

Let X_n be as in Theorem 10.8. We write

$$V_A := \min \{n \geq 0 : X_n \in A\}.$$

We define the vector $\mathbf{g} = (g(x))_{x \in C}$, where

$$g(x) := \mathbb{E}_x [V_A]. \quad \text{where } x \in C$$

That is the $x \in C$-th component $g(x)$ of the vector $\mathbf{g}$ is the expected number of steps until the absorbtion happens if the MC starts from x.

Time to absorption (cont.)

Theorem 10.9

Let X_n be an absorbing MC. We denote the column vector with all components equal to 1 *by* $\mathbf{1} \in \mathbb{R}^c$. *Then*

$$\mathbf{g} = \mathbf{N} \cdot \mathbf{1}. \tag{83}$$

We have actually proved this in the previous subsection in special cases. For a proof see [4, p. 420, Theorem 11.5]. This theorem answers Question Q2. In the following slides we will answer Question Q1.

An auxiliary lemma

We often need the following simple lemma.

Lemma 10.10

Let X be a non-negative integer valued r.v.. Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k). \tag{84}$$

Proof.

Observe that $X = \sum_{k=1}^{\infty} \mathbb{1}_{\{X \geq k\}}$. Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbb{1}_{\{X \geq k\}}\right] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$ □

Absorption probabilities

Let $\mathbf{B} = (b_{i,j})_{i \in C, j \in A}$ be a $c \times a$ matrix whose elements are defined as follows: for an $i \in C$ and $j \in A$ we write

$$b_{i,j} := \mathbb{P}(\text{the chain starting from } i \text{ is absorbed at } j)$$

Theorem 10.11

Let X_n be an absorbing MC. Then

$$\mathbf{B} = \mathbf{N} \cdot \mathbf{Q}. \tag{85}$$

Now we present the proof in a shorter form the we repeat in a more detailed form.

Proof in short

Proof of Theorem 10.11 in short.

Let $\mathbf{R}^0 := \mathbf{I}$. Then

$$\begin{aligned} b_{i,j} &\overset{(84)}{=} \sum_{n=0}^{\infty} \sum_{k \in C} r_{i,k}^{(n)} \cdot q_{k,j} \\ &= \sum_{k \in C} \sum_{n=0}^{\infty} r_{i,k}^{(n)} \cdot q_{k,j} \\ &= \sum_{k \in C} n_{i,k} \cdot q_{k,j} \\ &= (\mathbf{N} \cdot \mathbf{R})_{i,j}. \end{aligned}$$

□

Proof of Theorem 10.11 with details

Proof of Theorem 10.11 with details

Fix an arbitrary $i \in C$ and $j \in A$. Imagine that we start from i and finally arrive at j on such a such path which stay within C before arriving at j. Let m be the length of this path. Observe that $m = 2$ means no states in between i and j on the path and for $m > 2$ there are $n - 2$ states in between i and j on the path and all of them must be in C. So such a path is describe with $c_1, \ldots, c_{m-2} \in C$.

Proof of Theorem 10.11 with details (cont.)

Proof of Theorem 10.11 with details (cont.)

Let us call the probability that such a path is realized $w_{i,c_1,\dots,c_{m-2},j}$, where the word $c_1, \dots, c_{m-2}$ is the empty word if $m = 2$. Below we write $n = m - 1$ from the two but last step:

Proof of Theorem 10.11 with details (cont.)

Proof (cont.)

$$
\begin{aligned}
b_{i,j} &= \sum_{m=2}^{\infty} \sum_{c_1,\ldots,c_{m-2}\in C} w_{i,c_1,\ldots,c_{m-2},j} \\
&= \sum_{m=2}^{\infty} \sum_{c_1,\ldots,c_{m-2}\in C} p_{i,c_1} \cdot \prod_{k=1}^{m-2} p_{i_k,c_{k+1}} \cdot p_{c_{n-1},j} \cdot p_{c_{m-1},j} \\
&= \sum_{m=2}^{\infty} \sum_{c_1,\ldots,c_{m-2}\in C} r_{i,c_1} \cdot \prod_{k=1}^{m-2} r_{c_k,c_{k+1}} \cdot q_{c_{m-1},j} \\
&= \sum_{n=0}^{\infty} \sum_{k\in C} r_{i,k}^{(n)} \cdot q_{k,j} \\
&= \left(\sum_{n=0}^{\infty} \mathbf{R}^n \cdot \mathbf{Q}\right)_{i,j},
\end{aligned}
$$

Proof of Theorem 10.11 with details (cont.)

Proof (cont.)

where $\prod_{k=1}^{m-2} r_{c_k,c_{k+1}} := 1$ if $m = 2$. Hence

(86) $$B = \sum_{n=0}^{\infty} \mathbf{R}^n \cdot \mathbf{Q}$$

Recall that according to part (b) of Theorem 10.8 we have $\mathbf{N} = \sum_{n=0}^{\infty} \mathbf{R}^n$. Hence, by (86) we obtained that

(87) $$\mathbf{B} = \mathbf{N} \cdot \mathbf{Q}. \qquad \blacksquare$$

The problems considered

In this subsection X_n is an irreducible chain on the finite state space S with transition matrix P and w assume that $\#S \geq 3$. Let $i, j, k \in S$ be three distinct elements of S. We pose the following questions:

(Q4) What is the probability that the chain staring from $i \in S$ visits $j \in S$ earlier than $k \in S$?

(Q5) What is the probability that the chain staring from $j \in S$ returns to j earlier than it visits $k \in S$.

The answer to question Q4

We prepare an absorbing MC from X_n by declaring some of the states absorbing. Namely, let $\mathbf{e}_j$ and $\mathbf{e}_k$ be the coordinate unit vectors in $\mathbb{R}^{\#S}$ which contains a 1 in their j and k-th position respectively, and all other components are zero. We replace of the j-th and k-th rows of P by $\mathbf{e}_j$ and $\mathbf{e}_k$ respectively. The transition probability matrix obtained in this way is denoted by $P^{(j,k)}$ and the corresponding MC is denoted by $X_n^{(j,k)}$.

The answer to question Q4 (cont.)

Clearly, $X_n^{(j,k)}$ is an absorbing MC with absorbing states $A := \{j, k\}$ transient states $C := S \setminus C$. Let

$$P^{(j,k)} = \begin{pmatrix} \mathbf{R}^{(j,k)} & \mathbf{Q}^{(j,k)} \\ \mathbf{0}_{a,c} & \mathbf{I}_a \end{pmatrix}$$

be the canonical form of $P^{(j,k)}$ and let $\mathbf{N}^{(j,k)}$ be the corresponding fundamental matrix:

$$\mathbf{N}^{(j,k)} = \left(I - \mathbf{R}^{(j,k)}\right)^{-1},$$

The answer to question Q4 (cont.)

where I is the $(\#S-2) \times (\#S-2)$ identity matrix. Now we apply Theorem 10.11 for the MC $X_n^{(j,k)}$. That is we define the $(\#S-2) \times 2$ matrix

$$\mathbf{B}^{(j,k)} = \mathbf{N}^{(j,k)} \cdot \mathbf{Q}^{(j,k)}, \tag{88}$$

where the rows are indexed by the elements of C and the columns are indexed by $\{j, k\}$.

Now we can answer question Q4:

We introduce:

The answer to question Q4 (cont.)

(89)
$$\eta_{i,j,k} := \mathbb{P}(\text{the chain starting from } i \text{ visists } j \text{ earlier than } k)$$

Then by Theorem 10.11 and by the definition of matrix B we obtain that

(90)
$$\eta_{i,j,k} = \mathbf{b}_{i,j}^{(j,k)},$$

where $\mathbf{b}_{i,j}^{(j,k)}$ is the j-the element of the i-th row of the matrix $\mathbf{B}_{i,j}^{(j,k)}$ defined in (88) and this answers question Q4.

The answer to question Q5

Fix an arbitrary distinct $j, k \in S$. Let $\tau_{j,k}$ be the probability that the chain staring from $j \in S$ returns to j earlier than it visits $k \in S$. We can use the one-step reasoning. Namely, if the chain starting from j returns to j for the first time before visiting k then the chain starting from j cannot make its first step to k. So, in the first step the chain either remains in j (with probability $p(j, j)$ and then it has arrived back to j without visiting k) or it jumps to an $i \neq\in \{j, k\}$ and then it will continue starting now from $i \notin \{j, k\}$ and visits j earlier than k.

The answer to question Q5 (cont.)

The probability of this is (by definition) $\eta_{i,j,k}$. So, the one-step reasoning yields:

$$\tau_{j,k} = p_{j,j} + \sum_{i \notin \{j,k\}} p(j,i) \cdot \eta_{i,j,k}. \tag{91}$$

Example 10.12 (Exercise 1.13 from Lawler's book [7])

Let X_n be a MC on $S = \{1, 2, 3, 4, 5\}$ with

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{5} & \frac{4}{5} \\ 0 & 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

(a) Is this chain irreducible? Is it aperiodic? **(b)** Find π. **(c)** What is the expected number of steps to return to 1 for the first time if the chain starts from 1?

(d) What is the expected number of steps to get to 4 for the first time, if the chain starts from 1? **(e)** What is the probability that the chain visits 5 earlier than 3 if the chain starts from 1? **(f)** What is the probability that the chain starting from 3 returns to 3 earlier than it visits 5?

In[71]:=

```
Clear[𝒫, p]
```

$$p = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{5} & \frac{4}{5} \\ 0 & 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

```
𝒫 = DiscreteMarkovProcess[1, p]
```

In[74]:=

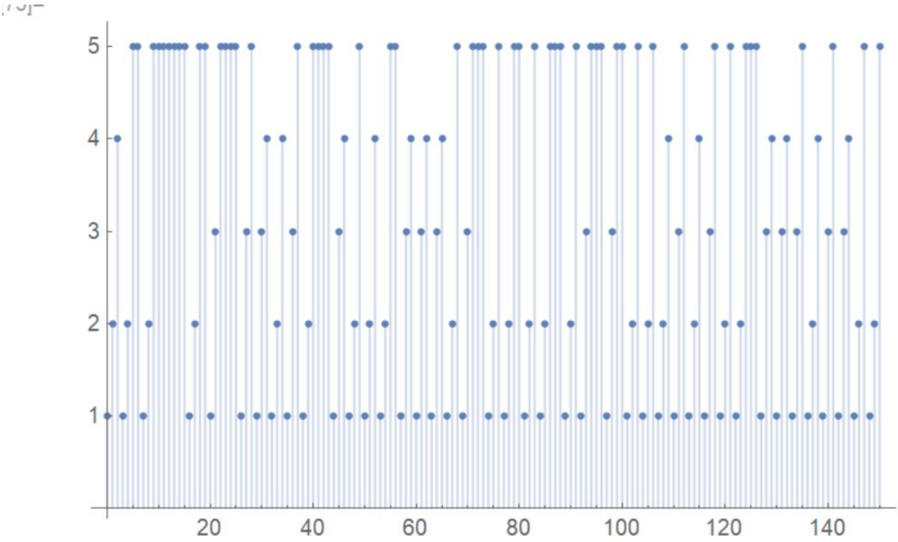
```
data = RandomFunction[𝒫, {0, 150}]
```

Out[74]=

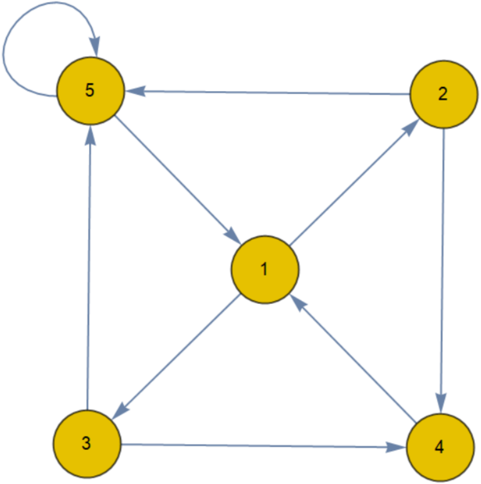
TemporalData[ Time: 0 to 150 Data points: 151 Paths: 1]

In[75]:=

```
ListPlot[data, Filling → Axis, Ticks → {Automatic, {1, 2, 3, 4, 5}}]
```


5
4
3
2
1
20
40
60
80
100
120
140

```
Graph[𝒫]
```



In[102]:= `MarkovProcessProperties[𝒫]`

Out[102]=

Basic Properties	
InitialProbabilities	
TransitionMatrix	
HoldingTimeMean	{0, 0, 0, 0, 1}
HoldingTimeVariance	{0, 0, 0, 0, 2}
Structural Properties	
CommunicatingClasses	{1, 2, 3, 4, 5}
RecurrentClasses	{1, 2, 3, 4, 5}
TransientClasses	None
AbsorbingClasses	None
PeriodicClasses	None
Periods	{}
Irreducible	True
Primitive	True
Aperiodic	True
Limiting Properties	
	

The chain is irreducible and aperiodic. This answers (a)

```
invmatrep =
 Inverse[ReplacePart[p - IdentityMatrix[Length[p[[1]]]], {i_, Length[p[[1]]]} :> 1]]
```

```
invmatrep[[Length[p[[1]]]]]
```

$$\left\{\frac{10}{37}, \frac{5}{37}, \frac{5}{37}, \frac{3}{37}, \frac{14}{37}\right\}$$

```
f[i_, j_] := Mean[FirstPassageTimeDistribution[DiscreteMarkovProcess[i, p], j]]
```

```
Array[f, {5, 5}] // MatrixForm
```

atrixForm=

$$\begin{pmatrix} \frac{37}{10} & \frac{23}{5} & \frac{24}{5} & \frac{34}{3} & \frac{23}{7} \\ \frac{14}{5} & \frac{37}{5} & \frac{38}{5} & \frac{35}{3} & \frac{13}{7} \\ \frac{13}{5} & \frac{36}{5} & \frac{37}{5} & 9 & \frac{19}{7} \\ 1 & \frac{28}{5} & \frac{29}{5} & \frac{37}{3} & \frac{30}{7} \\ 2 & \frac{33}{5} & \frac{34}{5} & \frac{40}{3} & \frac{37}{14} \end{pmatrix}$$

So, **(b)**: $\pi = \left(\frac{10}{37}, \frac{5}{37}, \frac{5}{37}, \frac{3}{37}, \frac{14}{37}\right)$. **(c)**: $37/10$. **(d)**: $\frac{34}{3}$.

Out[92]//MatrixForm=

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{5} & \frac{4}{5} \\ 0 & 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

```
In[94]:= r = p[[{1, 2, 4}, {1, 2, 4}]]

In[95]:= q = p[[{1, 2, 4}, {3, 5}]]

In[96]:= id = IdentityMatrix[3]

In[97]:= n = Inverse[id - r]

In[99]:= b = n.q // MatrixForm
```

Out[99]//MatrixForm=

$$\begin{pmatrix} \frac{5}{9} & \frac{4}{9} \\ \frac{1}{9} & \frac{8}{9} \\ \frac{5}{9} & \frac{4}{9} \end{pmatrix}$$

(e) The answer in 4/9. (The element in the first row since we start from 1 and the column which corresponds to 5 (this is a the second column).

$B^{3,5} =$

	3	5
1	$\frac{5}{9}$	$\frac{4}{9}$
2	$\frac{1}{9}$	$\frac{8}{9}$
4	$\frac{5}{9}$	$\frac{4}{9}$

That is

$$\eta_{1,3,5} = \frac{5}{9}, \quad \eta_{2,3,5} = \frac{1}{9}, \quad \eta_{4,3,5} = \frac{5}{9}. \tag{92}$$

Now we can answer question **(f)** that is we compute $\tau_{3,5}$ which was defined as the probability that the chain staring from 3 returns to 3 earlier than it visits 5. Namely, by (91) we have

$$\begin{aligned}
\tau_{3,5} &= p_{3,3} + p(3,1)\eta_{1,3,5} + p(3,2)\eta_{2,3,5} + p(3,4)\eta_{4,3,5} \\
&= 0 + 0 \cdot \frac{5}{9} + 0 \cdot \frac{1}{9} + \frac{2}{5} \cdot \frac{5}{9} \\
&= \frac{2}{9}.
\end{aligned}$$

Umbrellas example [3, Excercise 1.37]

Example 10.13

An individual has three umbrellas, some at her office, and some at home. If she is leaving home in the morning (or leaving work at night) and it is raining, she will take an umbrella, if one is there. Otherwise, she gets wet. Assume that independent of the past, it rains on each trip with probability 0.2.

Question 1: Which percentage of time does she get wet?

Umbrellas example (cont.)

We approach this problem in the language of Markov chains. The only idea:
Let $S := \{0, 1, 2, 3\}$ and we write X_n for the number of umbrellas at the current location.
Then the transition matrix P is:

	0	**1**	**2**	**3**
0	0	0	0	1
1	0	0	0.8	0.2
2	0	0.8	0.2	0
3	0.8	0.2	0	0

Umbrellas example (cont.)

```
In[57]:= Clear[𝒫, p]

In[58]:= p = {{0, 0, 0, 1},
              {0, 0, 8/10, 2/10},
              {0, 8/10, 2/10, 0},
              {8/10, 2/10, 0, 0}}

In[59]:= invmatrep =
          Inverse[ReplacePart[p - IdentityMatrix[Length[p[[1]]]],
            {i_, Length[p[[1]]]} :> 1]]
         invmatrep[[Length[p[[1]]]]]

Out[60]=
```

$$\left\{\frac{4}{19}, \frac{5}{19}, \frac{5}{19}, \frac{5}{19}\right\}$$

Umbrellas example (cont.)

This yields that the stationary distribution is

$$\pi = \left(\frac{4}{19}, \frac{5}{19}, \frac{5}{19}, \frac{5}{19}\right). \tag{93}$$

Hence it happens with probability $4/19$ that the individual does not have any umbrellas at her current location. However, she does not necessarily get wet at all of these occasions, since there is a rain only every 5th days (independently of everything). So, she gets wet with probability $4/(19 \cdot 5) = 0.04210526...$. Remark: the stationary distribution could be computed by hands easily since the system of equations is very simple.

Umbrellas example (cont.)

Namely, we want to find a probability vector
$\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ such that
(94)

$$(\pi_0, \pi_1, \pi_2, \pi_3) \cdot \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{8}{10} & \frac{2}{10} \\ 0 & \frac{8}{10} & \frac{2}{10} & 0 \\ \frac{8}{10} & \frac{2}{10} & 0 & 0 \end{pmatrix} = (\pi_0, \pi_1, \pi_2, \pi_3)$$

This yields the system of equations:

Umbrellas: answer of Question 1

$$
\begin{aligned}
0.8\pi(3) &= \pi(0) \\
0.8\pi(2) + 0.2\pi(3) &= \pi(1) \\
0.8\pi(1) + 0.2\pi(2) &= \pi(2) \\
\pi(0) + \pi(1) + \pi(2) + \pi(3) &= 1
\end{aligned} \tag{95}
$$

As on slide 33, we throw away the last equation and substituted it by the condition that the sum of the components of π is equal to one, since the last equation of the original system would give no more information than the retained first three equations do. The solution of the system (95) is really obvious high school mathematics.

Notation used in this Section

- In this Section we always that X is a such r.v. which takes only non-negative integers.
- $\forall k \in \mathbb{N}$-re let $p_k := \mathbb{P}(X = k)$.
- The generator function of the r.v. X is

$$g_X(s) := \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} p_k \cdot s^k.$$

The most basic properties of generator functions (in short: g.f.)

Generator functions

(a) A generator function uniquely determines the cumulative distribution function.

(b) The generator function of the sum of two independent r.v. which take only non-negative integers, is the product of the generator functions of these r.v..

Generator functions (cont.)

(c) Let $g(x)$ be the generator function of the r.v. X. Then

$$\mathbb{E}\left[X(X-1)\cdots(X-k)\right] = g^{(k+1)}(1),$$

where $g^{(k+1)}$ is the $k+1$-th derivative of g. Hence by a simple calculation we get:
(96)

$$\mathbb{E}\left[X\right] = g'(1) \text{ és } \mathbb{E}\left[X^2\right] = g''(1) + g'(1).$$

(d) $g(1) = 1$ since (p_k) is a probability vector.

Generator functions (cont.)

Lemma 11.1

Let X and N be independent non-negative integer valued r.v. with generator functions g_X és g_N. Moreover, let $X_1, X_2, \ldots$ be i.i.d. r.v. having the same distribution as X. We define the r.v.:

$$R := X_1 + \cdots + X_N.$$

Then the generator function of R is:

$$g_R(s) = g_N(g_X(s)). \tag{97}$$

Generator functions (cont.)

Before the proof of the Lemma we remark that an important corollary of Lemma 11.1 is as follows: Using properties (c) and (d) from slide 279 we obtain that (98)

$$\mathbb{E}[R] = g_R'(1) = g_N'(\underbrace{g_X(1)}_{1}) \cdot g_X'(1) = \mathbb{E}[N] \cdot \mathbb{E}[X].$$

Generator functions (cont.)

Proof.

$$
\begin{aligned}
g_R(s) &\overset{\text{def of } g_R}{=} \mathbb{E}\left[s^R\right] \overset{\text{def of } R}{=} \mathbb{E}\left[s^{X_1+\cdots+X_N}\right] \\
&\overset{\text{tower prop.}}{=} \mathbb{E}\left[\mathbb{E}\left[s^{X_1+\cdots+X_N}\right] \middle| N\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[\mathbb{E}\left[s^{X_1+\cdots+X_n}\right] \middle| N = n\right] \cdot \mathbb{P}\left(N = n\right) \\
&= \sum_{n=0}^{\infty} \underbrace{\mathbb{E}\left[s^{X_1+\cdots+X_n}\right]}_{g_X^n(s)} \cdot \mathbb{P}\left(N = n\right) \\
&= \mathbb{E}\left[g_X^N(s)\right] \overset{\text{def of } g_N}{=} g_N\left(g_X(s)\right)
\end{aligned}
$$

Branching Processes with more details

We introduced Branching Processes on slide 146. Given a probability vector $(p_k)_{k=0}^{\infty}$ which we call offspring distribution. A population develops according to the following rule: At the beginning there is one individual on level 0. Then for all $n \geq 0$, each individual on level n independently gives birth to k offsprings with probability p_k. The same with notations:
Let Y be a non-negative integer valued r.v. such that $\mathbb{P}(Y = k) = p_k$. Fix an arbitrary $n \geq 0$. Let X_n denote the number of level n individuals. The level n individuals

Branching Processes with more details (cont.)

$\{1, 2, \ldots, X_n\}$ give birth to $Y_1^{(n)}, \cdots, Y_{X_n}^{(n)}$ individuals.
So, the number of level $n+1$ individuals is:

$$X_{n+1} = Y_1^{(n)} + \cdots + Y_{X_n}^{(n)}. \tag{99}$$

We always assume that $\left\{Y_m^{(n)}\right\}_{m,n}$ are i.i.d. r.v. with

$$Y_m^{(n)} \stackrel{d}{=} Y.$$

Branching Processes with more details (cont.)

That is

$$\mathbb{P}\left(Y_m^{(n)} = k\right) = p_k.$$

We can consider (X_n) as a Markov Chain with state space $S = \{0, 1, 2, \dots\}$ and the transition matrix $P = (p_{i,j})$ is given by

(100)
$$p(i,j) = P\left(Y_1 + \cdots + Y_i = j\right) \qquad \text{for } i > 0 \text{ and } j \geq 0,$$

where $\{Y_i\}_{i=1}^{\infty}$ are i.i.d. with $Y_k \overset{d}{=} Y$.

Branching Processes with more details (cont.)

Let

$$g_n := \mathbb{E}\left[s^{X_n}\right],$$

That is g_n is the generator function of X_n, (which was defined as the number of level n individuals). Let

$$g(s) := g_1(s) := g_Y(s) = \sum_{n=0}^{\infty} p_n \cdot s^n.$$

Branching Processes with more details (cont.)

Clearly, for all m, the generator function of Y_m is the same:

$$g(s) = g_{Y_m}(s) \quad \forall m.$$

To get a better understanding of the generator function g_n we apply Lemma 11.1 with the following substitutions:

$$X_n \to N, Y_i \to X_i, X_{n+1} \to R.$$

Branching Processes with more details (cont.)

The we obtain from Lemma 11.1 that

$$g_{n+1} = g_n(g(s)).$$

From here, we obtain by mathematical induction that

$$g_n(s) = \underbrace{g \circ \cdots \circ g}_{n}(s) =: g^n(s). \tag{101}$$

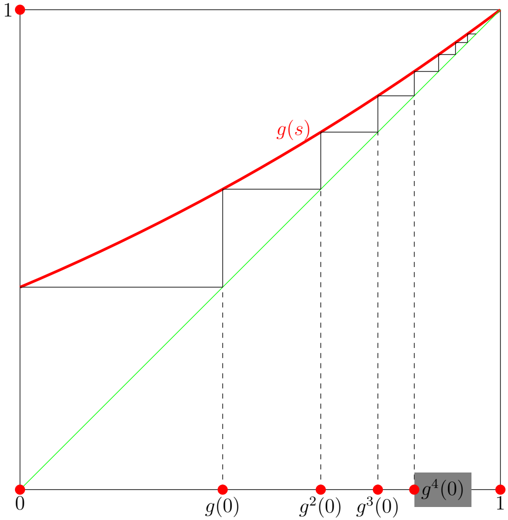
Apply this for $s = 0$ to get:

$$\mathbb{P}(X_n = 0) = g^n(0). \tag{102}$$

Branching Processes with more details (cont.)

Hence $\mathbb{P}(\text{Extinction}) = \lim_{n\to\infty} \mathbb{P}(X_n = 0)$, where Extinction is the event the the Brancing Process dies out in finitely many steps.

$\mathbb{E}[Y] = g'(1) < 1 \implies \lim_{n\to\infty} \mathbb{P}(X_n = 0) = 1$



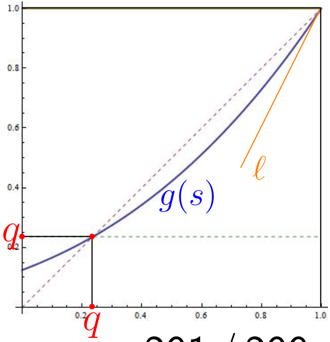
Branching Processes

Summary: p_n is the probability that an individual has exactly n offsprings. Then the expected number of offsprings of an individual is $m := \sum_{n=1}^{\infty} p_n \cdot n$. Consider the generator function:

$$g(s) := \sum_{n=0}^{\infty} p_n \cdot s^n$$

. The graph of g goes through $(1,1)$. Let ℓ be the tangent line to g at $s=1$. The slope of ℓ is $g'(1) = m$. If $m > 1$ then $\ell \cap [0,1]^2$ is below the line $y = x$. Hence $\exists$ a $q \in [0,1)$ with $g(q) = q$. Looking at the Figure:

$0 \le g'(q) < 1$. So, for $g^n := \underbrace{g \circ \cdots \circ g}_{n}$, we have $g^n(0) \to q$. That is by (102) q is the probability of extinction.



[1] Sven Erick Alm
Sven Erick Alm
Simple random walk. Click here for the online version.

[2] Balázs Márton, Tóth Bálint
Valószínűségszámítás 1. jegyzet matematikusoknak és fizikusoknak (Probability 1 *notes for mathematicians and physicists)*
Márton Balzs's website, 2012. Click here for the online version.

[3] R. Durrett
Essentials of Stochastic Processes, Second edition
Springer, 2012. Click here for the almost-ready version.

[4] Charles M. Grinstead, J. Laurie Snell
Itroduction to Probability
Click here for the book.

[5] S. Karlin, H.M. Taylor
Sztochasztikus Folyamatok
Gondolat, Budapest, 1985

[6] D.A. Levin, Y. Peres,E.L. Wilmer
Markov chains and mixing times
American Mathematical Society, 2009.
Click here for the book.

[7] G. Lawler
Intoduction to Stochastic Processes
Chapman & Hall 1995.

[8] E. Senata
Non-negative Matrices and Markov Chains
Springer 1973.

[9] TÓTH BÁLINT *Sztochasztikus folyamatok jegyzet (Stochastic processes notes)*
Click here for the Bálint Tóth notes.

Examples

Examples (cont.)

Examples (cont.)