Statistical Bioinformatics, Makerere
More on Markov Chains
Timo Koski

TK

02.08.2018

KTH Matematik

Lecture

The lecture expands on Markov chains.

Discrete Time Markov Chain

- A *discrete time Markov chain* (MC) $\{X_n | n = 0, 1, \ldots\}$ is a discrete time, random sequence with values in a discrete state space S_X, such that

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

- p_{ij} is the transition probability
- *State* X_n summarizes the past history needed to predict X_{n+1}

KTH Matematik

Discrete Time Markov Chain

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

The condition is known as the *Markov property*. MC is assumed time homogeneous, i.e.,

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)$$

KTH Matematik

Markov Chain: Lack of Memory

The significance of an MC lies in the fact that if $X_n = j$ is a future event, then the conditional probability of this event given the past history $X_0 = j_0, X_1 = j_1, \ldots, X_{n-1} = j_{n-1}$ depends only upon the immediate past $X_{n-1} = j_{n-1}$ and not upon the remote past $X_0 = j_0, X_1 = j_1, \ldots, X_{n-2} = j_{n-2}$.

KTH VETENSKAP OCH KONST
KTH Matematik

Transition Probabilities Properties

- $p_{ij} \geq 0$
- $\sum_{j \in S_X} p_{ij} = 1$
- Finite Markov chain: $S_X = \{0, 1, \ldots, K\}$

Transition matrix

$$\mathbf{P} = (p_{ij})_{i=0,j=0}^{K,K}$$

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0K} \\ p_{10} & p_{11} & \cdots & p_{1K} \\ \vdots & \vdots & \vdots & \vdots \\ p_{K0} & p_{K1} & \cdots & p_{KK} \end{pmatrix}.$$

$\mathbf{P}$ is an $K+1 \times K+1$ matrix to be called a *transition matrix*.

Binary Autoregression

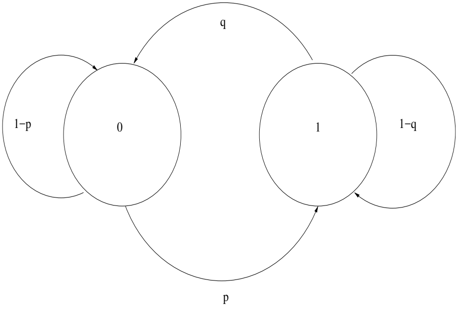
$Y_{n+1} = Y_n +_2 X_n$, $+_2$ is binary addition, X_n I.I.D. $\mathrm{Be}(p)$, Y_0 binary, independent of X_n. This MC has the transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} ; 0 \leq p \leq 1$$

$0 \leq p \leq 1$.

State transition graph

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad 0 \leq p \leq 1, 0 \leq q \leq 1$$



n-step transition probabilities

- The n-step transition probabilities are

$$p_{ij}(n) = P(X_{n+m} = j | X_m = i)$$

  This does not depend on m.

- **Chapman-Kolmogorov equations:**

$$p_{ij}(n+m) = \sum_{k \in S_X} p_{ik}(n) \cdot p_{kj}(m)$$

KTH Matematik

Chapman-Kolmorogorov equations, Matrix Form:

Using a matrix notation we can write the Chapman - Kolmogorov equation as the following matrix multiplication

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(m)} \cdot \mathbf{P}^{(n)}.$$

$$\Rightarrow$$

$$\mathbf{P}^{(n)} = \mathbf{P}^{n}.$$

State probabilities at step n

$$p_j(n) = P\left(X_n = j\right)$$

The probability that the MC visits the state j at time n.

Computation of State probabilities at step n

- One iteration with n-step transition probs:

$$p_j(n) = \sum_i p_{ij}(n) p_i(0)$$

- n iterations with 1-step transition probs:

$$p_j(n) = \sum_i p_i(n-1) p_{ij}$$

KTH Matematik

Matrix Formalism

- n step transition matrix: $\mathbf{P}^{(n)} = \mathbf{P}^n$
- $$p_j(n) := P\left(X_n = j\right)$$
  $$p(n) = \left(p_0(n) \cdots p_K(n)\right)$$
- Recursion for $p(n)$:
  $$p(n) = p(0)\mathbf{P}^n = p(n-1)\mathbf{P}$$

KTH Matematik

Limiting State Probabilities

- MC with states $\{0, 1, 2, \ldots\}$
- initial state probabilities $\{p_j(0)\}$
- The *limiting state probabilities*, when they exist, are $\{\pi_j\}$ s.t.

$$\pi_j = \lim_{n\to\infty} p_j(n) = \lim_{n\to\infty} P(X_n = j)$$

where

$$\sum_{j \in S_X} \pi_j = 1.$$

KTH Matematik

Stationary Probabilities

If the limiting state probabilities exist, then

$$\pi = \lim_{n \to \infty} p(n) = \lim_{n \to \infty} p(n-1)\mathbf{P}$$

$\Leftrightarrow$

$$\pi = \pi \mathbf{P}$$

We shall call a probability distribution on S_X that satisfies the equation (system of equations) above a **stationary distribution**.

KTH Matematik

Stationary Probabilities

If

$$\pi = \pi \mathbf{P}$$

and $p(0) = \pi$, then $p(n) = \pi$ for all $n \geq 1$, since

$$p(n) = p(0)\mathbf{P}^n = \pi \mathbf{P}^n = \pi \mathbf{P}\mathbf{P}^{n-1} = \pi \mathbf{P}^{n-1} =$$

$$\ldots = \pi.$$

KTH Matematik

Three Questions

- Existence of solution to $\pi = \pi\mathbf{P}$?
- Uniqueness of π ?
- Convergence

$$\pi = \lim_{n\to\infty} p(n) =$$

$$\lim_{n\to\infty} p(0)\mathbf{P}^n$$

for any $p(0)$?

KTH Matematik

Three Questions

The questions above have clearly to do with $\mathbf{P}^n$, (the probabilities $p_{ij}(n)$), i.e., with the structure of the transition matrix $\mathbf{P}$. We shall consider first the case of finite state spaces $S_X = \{0, 1, \ldots, K\}$.

Part B: Classification of states of finite MCs

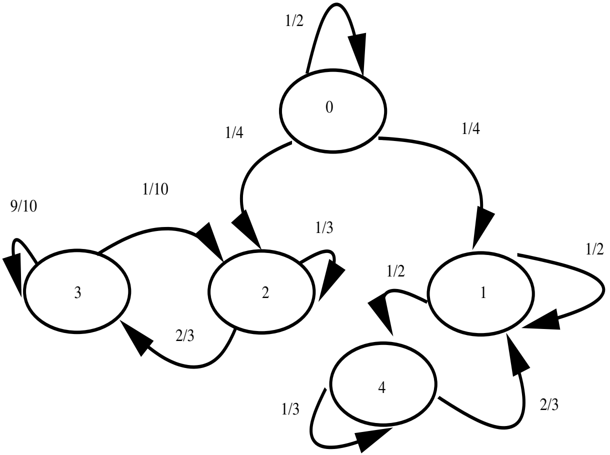
Phenomenon: $\lim_{n\to\infty} \mathbf{P}^n$ does not exist

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

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

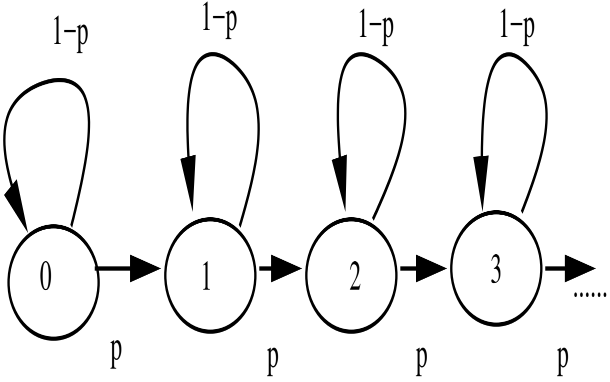
KTH Matematik

Phenomenon: different limits



Phenomenon: The probability mass drifts away

V_n I.I.D. $\mathrm{Be}(p)$, $X_n = V_1 + \ldots + V_n$, $X_0 = 0$ is MC. $S_X = \{0, 1, 2, \ldots\}$, (countably) infinite state space needed for this phenomen



Concepts

We shall now develop concepts for dealing with the phenomena above. We need to study the structure of the transition matrices or the transition graph.

Accessibility

State j in a finite MC is accessible from state i, written $i \to j$, if

$$p_{ij}(n) > 0,$$

for some $n > 0$

Communicating States

- i and j *communicate*, $(i \leftrightarrow j)$, if there is a path from i to j, $i \rightarrow j$, and a path from j to i with positive probability, $j \rightarrow i$.

Communicating States

- i and j *communicate*, $(i \leftrightarrow j)$, if there is a path from i to j and a path from j to i, or there are n and m such that

$$p_{ij}(n) > 0, p_{ji}(m) > 0.$$

KTH Matematik

Communicating Class

- A *communicating class* is a subset C of states $C \subseteq S_X$ that all communicate

Transient States

In a finite MC a state i is **transient**, if there is a state j such that $i \to j$ but $j \not\to i$.

Example: $X_n = V_1 + \ldots + V_n$, $X_0 = 0$, V_n I.I.D. $\mathrm{Be}(p)$, all states of $\{X_n\}$ are transient.

Recurrernt States

In a finite MC a state i is **recurrent**, if there is no state j such that $i \to j$ but $j \not\to i$.

Example: All states of a binary autoregression are recurrent.

- If i is recurrent and $i \leftrightarrow j$, then j is recurrent.

KTH Matematik

Class properties I

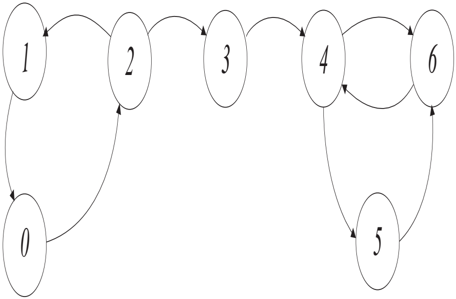
- An MC with a finite number of states always has a set of recurrent states.

Class Properties II

- For a communicating class of a finite Markov chain, one of the following must be true:
  - All states are transient
  - All states are recurrent

Example: Comm Classes

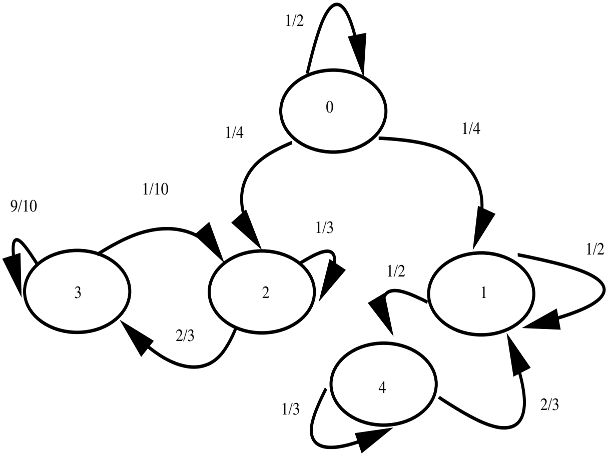
- Positive transition probs shown. What are communicating classes?



Irreducible Markov Chain

A finite Markov chain is *irreducible* if there is only one communicating class.

Example of a Reducible Markov Chain



Periodic and Aperiodic States

State i has **period** d if d is the largest integer such that $p_{ii}^{(n)} = 0$ whenever n is not divisible by d.

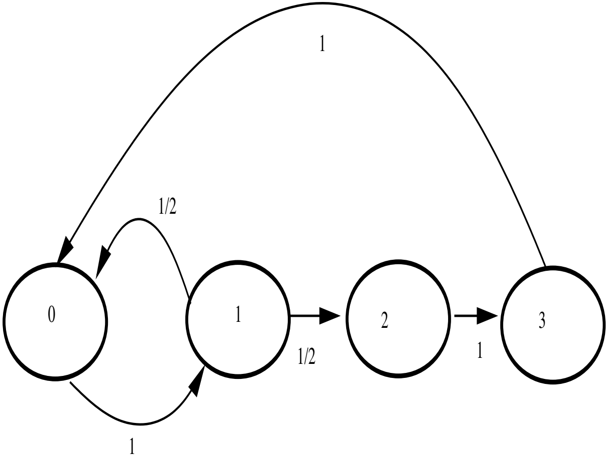
- If $d = 1$, then state i is called **aperiodic**.
- All states in the same communicating class have same period

KTH Matematik

Periodic States in Other Words

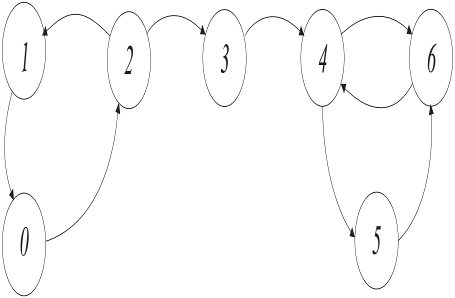
A state has period $d > 1$, if it is possible to return to this state with positive probability only when the number of steps is a multiple of d.

Periodicity



Example: Periods of Communicating Classes

- Positive transition probs shown. What are the periods of the communicating classes?



Finite Irreducible Aperiodic Chain

- For an irreducible, aperiodic, finite MC, $\pi = (\pi_0, \ldots, \pi_K)$ is the unique nonnegative solution of

$$\pi_j = \sum_{i=0}^{K} \pi_i p_{ij} \qquad \sum_{j=0}^{K} \pi_j = 1$$

- It will be seen that π_j is the limiting fraction of time spent in state j

KTH Matematik

Limit Theorems (1) for Finite Irreducible Aperiodic Chain

For an irreducible, aperiodic, finite MC,

$$\lim_{n \to \infty} \mathbf{P}^n =$$

$$\begin{pmatrix} \pi_0 & \cdots & \pi_K \\ \vdots & \vdots & \vdots \\ \pi_0 & \cdots & \pi_K \\ \vdots & \vdots & \vdots \\ \pi_0 & \cdots & \pi_K \end{pmatrix} = \pi \mathbf{1}.$$

where $\mathbf{1}$ is a row vector of ones. $\pi = \pi \mathbf{P}$ $\quad \sum_{j=0}^{K} \pi_j = 1$

KTH Matematik

Limit Theorems (2) for Finite Irreducible Aperiodic Chain

For an irreducible, aperiodic, finite MC, with initial distribution $p(0)$

$$\lim_{n\to\infty} p(n) = \pi.$$

Proof:

$$\begin{aligned}\lim_{n\to\infty} p(n) &= \lim_{n\to\infty} p(0)\mathbf{P}^n \\ &= p(0)\pi\mathbf{1} = \pi\end{aligned}$$

KTH Matematik

Ergodic Markov chain

MC is *ergodic* if it has a stationary distribution and the state probabilities converge to it.

KTH Matematik

Solving $\pi = \pi\mathbf{P}$

- Note that there are many solutions to

$$\pi_j = \sum_{i=0}^{K} \pi_i p_{ij}, \quad j = 0, 1, \ldots, K$$

Uniqueness may be obtained by replacing one of the equations by

$$\sum_{j=0}^{K} \pi_j = 1.$$

KTH Matematik

Stationary Markov chain

An ergodic Markov chain is a stationary stochastic process if $p(X_0) = \pi$, since

$$P\left(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) =$$

$$= \prod_{l=1}^{n} P\left(X_l = i_l | X_{l-1} = i_{l-1}\right) \pi_{i_0} = \prod_{l=1}^{n} p_{i_{l-1} i_l} \pi_{i_0}$$

and this is invariant to the shift of time, $0 \mapsto h$, $l \mapsto l + h$, $l = 1, \ldots, n$.

KTH Matematik

Why worry about stationarity: McCabe's library in a special case

$p_1 = \Pr(\beta)$, $p_2 = \Pr(B)$, $p_3 = \Pr(M)$

	β BM	β MB	B β M	BM β	M β B	M B β
β BM	p_1	0	p_2	0	p_3	0
β MB	0	p_1	p_2	0	p_3	0
B β M	p_1	0	p_2	0	0	p_3
BM β	p_1	0	0	p_2	0	p_3
M β B	0	p_1	0	p_2	p_3	0
M B β	0	p_1	0	p_2	0	p_3

KTH Matematik

McCabe's library: stationary distribution

It is easily seen that this is an aperiodic, irreducible MC. There is an explicit formula for the stationary distribution: Let $k = 1, 2, \ldots, 6$ be a numbering of the permutations of the three books,

$$\{\beta BM, \beta MB, B\beta M, BM\beta, M\beta B, MB\beta\} = \{1, 2, 3, 4, 5, 6\}$$

Then the stationary probability of the kth state (permutation) is

$$\pi_k = \prod_{n=1}^{3} \left(\frac{p_{i_n}}{\sum_{j=n}^{3} p_{i_n}} \right),$$

For example, for $k = 2$, $p_{i_1} = p_1, p_{i_2} = p_3, p_{i_3} = p_2$ and

$$\pi_2 = \pi(\beta MB) = \frac{p_1}{p_1 + p_2 + p_3} \cdot \frac{p_3}{p_3 + p_2} \cdot \frac{p_2}{p_2}$$

$$= \frac{p_1 p_3}{p_2 + p_3}.$$

KTH Matematik

McCabe's library: stationary distribution

We see that the most probable state in the stationary distribution is the one in which the books are ordered from left to right in decreasing order of probability.

McCabe's library

If the library is stationary, then various important quantities like expected search time for an item can be computed.

How to compute $\mathbf{P}^n$

- Let λ_i, $\mathbf{x}_i$ (a row vector $1 \times K + 1$) be the **left** eigenvalues and eigenvectors of a symmetric $\mathbf{P}$, i.e.,

$$\mathbf{x}_i \mathbf{P} = \lambda_i \mathbf{x}_i$$

- $$\mathbf{P} = \mathbf{U}^{-1} \Lambda \mathbf{U}$$

  where Λ is a diagonal with the eigenvalues at the main diagonal, $\mathbf{U}$ is an orthogonal matrix that has the standardized left eigenvectors as rows.

- $$\mathbf{P}^n = \mathbf{U}^{-1} \Lambda^n \mathbf{U}$$

KTH Matematik

Example: Binary Autoregression

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

$$\lambda_1 = 1,\ \lambda_2 = 1-2p,\ \mathbf{x}_1 = \tfrac{1}{\sqrt{2}}(1,1),\ \mathbf{x}_2 = \tfrac{1}{\sqrt{2}}(1,-1)$$

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\mathbf{U}^{-1} = \mathbf{U}$$

Example: Binary MC

For

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

a similar expansion can be done, see example 12.6 on page 449 in (Yates & Goodman Second Edition).

Example: $\mathbf{P}^n$ for Binary Autoregression

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1-2p \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

KTH Matematik

Example: $\mathbf{P}^n$ for Binary Autoregression

$$\mathbf{P}^n = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & (1-2p)^n \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

KTH Matematik

Example: $\lim_{n\to\infty} \mathbf{P}^n$ for Binary Autoregression

$$\Lambda^n = \begin{pmatrix} 1^n & 0 \\ 0 & (1-2p)^n \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

since $\mid (1-2p)^n \mid < 1$, if $0 < p < 1$.

KTH Matematik

Example: $\lim_{n\to\infty} \mathbf{P}^n$ for Binary Autoregression

$$\mathbf{P}^n = U^{-1}\Lambda^n U = \rightarrow \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

KTH Matematik

Limiting State Probabilities

$$\pi_j = \lim_{n\to\infty} p_j(n) = \lim_{n\to\infty} P(X_n = j)$$

$p(n) = (p_0(n) \cdots p_K(n))$, row vector,

$$\lim_{n\to\infty} p(n) = \lim_{n\to\infty} p(0)\mathbf{P}^n$$

$$= \lim_{n\to\infty} p(0)U^{-1}\Lambda^n U$$

Limiting State Probabilities for Binary Autoregression

From the above

$$\lim_{n\to\infty} p(0)U^{-1}\Lambda^n U = p(0)\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$= (p_0(0), p_1(0))\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \left(\frac{1}{2}, \frac{1}{2}\right)$$

since $p_0(0) + p_1(0) = 1$. Hence $(\pi_0, \pi_1) = \left(\frac{1}{2}, \frac{1}{2}\right)$.
Notice that this convergence holds for an arbitrary initial distribution, $(p_0(0), p_1(0))$.

KTH VETENSKAP OCH KONST
KTH Matematik

Limiting State Probabilities for Binary Autoregression are Stationary:

$\pi = (\pi_0, \pi_1) = \left(\frac{1}{2}, \frac{1}{2}\right).$

$$(1/2, 1/2)\,\mathbf{P} = (1/2, 1/2)\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} = (1/2, 1/2)$$

i.e.,

$$\pi\mathbf{P} = \pi$$

Del C: Countable number of states

Auxiliary random variables

- Starting in state i, V_{ij} is the event that the system eventually[1] visits state j.
- For states i and j with $P(V_{ij}) = 1$, the *first transition time* T_{ij} is the number of transitions required to enter state j when starting from state i.

[1] så småningom, omsider

Auxiliary probabilities

We have

$$P\left(T_{ij}=n\right)=P\left(X_n=j,X_k\neq j,1\leq k\leq n-1|X_0=i\right)$$

We denote also $f_{ij}(n):=P\left(T_{ij}=n\right)$. Then

$$P(V_{ij})=\sum_{n=1}^{\infty}f_{ij}(n)$$

$$E\left(T_{ij}\right)=\sum_{n=1}^{\infty}nf_{ij}(n)$$

KTH Matematik

Binary Autoregression

$$f_{jj}(0) = 0, f_{jj}(1) = 1 - p, f_{jj}(n) = p \cdot (1-p)^{n-2} \cdot p, \quad n \geq 2.$$

$$E\left(T_{ij}\right) = \sum_{n=1}^{\infty} n f_{ij}(n) = 2$$

KTH Matematik

Transient and Recurrent States

- A state j is either
  - *Transient* if $P(V_{jj}) < 1$
  - *Null Recurrent* if $P(V_{jj}) = 1$, and $ET_{jj} = \infty$
  - *Positive Recurrent* if $P(V_{jj}) = 1$ and $\mathrm{E}\mathrm{T}_{jj} < \infty$

Recurrent States

- For a communicating class of a Markov chain, one of the following must be true:
    - All states are transient
    - All states are null recurrent
    - All states are positive recurrent

Recurrent States

Positive recurrent and null recurrent states are thus recurrent states.

Limiting State Probabilities

Assume

- States i and j with $P(V_{ij}) = 1$
- $\sum_{n=1}^{\infty} P\left(T_{ij} = nd\right) = 1$ for $d = 1$ but for no larger d.

Then

- the limiting prob of state j given the MC starts in state i is

$$\lim_{n \to \infty} p_{ij}(n) = \frac{1}{E\mathrm{T}_{jj}}$$

KTH Matematik

Transient States

The state j is *transient* if

$$\sum_{n=1}^{\infty} p_{jj}(n) < \infty$$

Transition matrices/graphs

How do we know when $P(V_{ij}) = 1$? It is in most cases difficult to compute $f_{ij}(n) = P\left(T_{ij} = n\right)$.

KTH Matematik

Recurrent States

- We say that the state j is *recurrent*, if $P(V_{ij}) = 1$.
- The state j is *recurrent* $\Leftrightarrow$

$$\sum_{n=1}^{\infty} p_{jj}(n) = \infty,$$

where $p_{jj}(n)$ is the element in position j, j in $\mathbf{P}^{(n)}$.

KTH Matematik

Transition matrices/graphs

How do we know when $P(V_{ij}) = 1$? It is also in most cases difficult to calculate $\sum_{n=1}^{\infty} p_{jj}(n)$.

Aperiodic Irreducible Chain

For an aperiodic irreducible Markov chain, either

- States are all transient or all null recurrent, $\lim_{n\to\infty} p_{ij}(n) = 0$. No stationary probabilities
- All states are positive recurrent, and

$$\pi_j = \frac{1}{E\mathrm{T}_{jj}} = \lim_{n\to\infty} P_{ij}^n > 0$$

are unique stationary probabilities satisfying

$$\pi_j = \sum_{i \in S_X} \pi_i p_{ij}$$

KTH Matematik

Stationary distribution

The components in the stationary distribution can be interpreted as the asymptotic percentages of time the chain spends in each of the states.

- To see this, let $T_j(1), T_j(1) + T_j(2), T_j(1) + T_j(2) + T_j(3), \cdots$ be the times, when the MC returns to j.
- Then
  $T_j(k) =$ the time between the k:th and $k - 1$:th visits j.
- It can be shown that $T_j(k)$ are I.I.D. random variables, if $P(V_{ij}) = 1$.

KTH Matematik

Stationary distribution

Hence the portion of time the MC has spent in state j after k returns

$$\frac{\text{number of visits}}{\text{total time}} = \frac{k}{T_j(1) + T_j(2) + \ldots + T_j(k)}$$

$$= \frac{1}{\frac{T_j(1)+T_j(2)+\ldots+T_j(k)}{k}} \rightarrow \frac{1}{E\mathrm{T}_{jj}}$$

by the law of large numbers.

KTH Matematik

Renewal Equation

$$p_{ij}(n) = \sum_{k=1}^{n} f_{ij}(k) p_{jj}(n-k)$$

From this renewal equation we can get

$$\lim_{n\to\infty} p_{ij}(n) = \frac{1}{\sum_{n=1}^{\infty} n f_{ij}(n)}$$

(Renewal theorem)

Binary Autoregression

$$E\left(T_{ij}\right) = \sum_{n=1}^{\infty} n f_{ij}(n) = 2$$

$$\lim_{n \to \infty} p_{ij}(n) = \frac{1}{\sum_{n=1}^{\infty} n f_{ij}(n)} = \frac{1}{2}.$$

KTH Matematik

- If MC irreducible, then

$$\lim_{n \to \infty} p_{ij}(n) = \pi_j$$

and the limit is the same for all i and depends only on j.

KTH Matematik