Imprecise Markov chains

Damjan Škulj

University of Ljubljana
damjan.skulj@fdv.uni-lj.si

4th SIPTA Summer School, Durham, UK
1-6 September 2010

Contents

1. Markov chains
   - Markov chains
   - Calculating probabilities at time n
   - Long term distributions
   - Coefficients of ergodicity
2. Imprecise probabilities
   - Credal sets
   - Lower and upper expectations
   - Coherent lower and upper previsions
3. Imprecise Markov chains
   - Markov chains with probability intervals
   - Markov set chains
   - Operators approach
4. Convergence of imprecise Markov chains
   - Regularity
   - Accessibility relations
   - Coefficients of ergodicity for imprecise Markov chains
   - The uniform coefficient of ergodicity
   - The weak coefficient of ergodicity

Contents

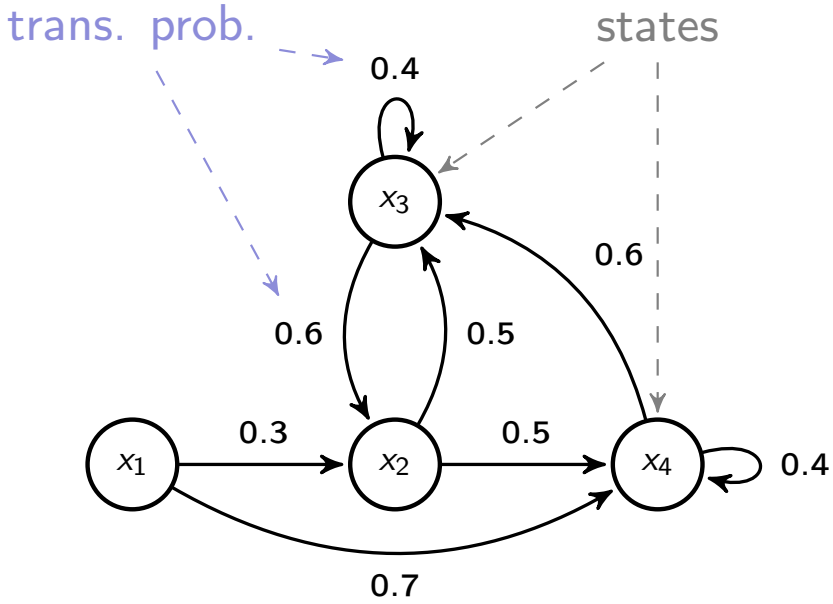
1. Markov chains
   - Markov chains
   - Calculating probabilities at time n
   - Long term distributions
   - Coefficients of ergodicity
2. Imprecise probabilities
   - Credal sets
   - Lower and upper expectations
   - Coherent lower and upper previsions
3. Imprecise Markov chains
   - Markov chains with probability intervals
   - Markov set chains
   - Operators approach
4. Convergence of imprecise Markov chains
   - Regularity
   - Accessibility relations
   - Coefficients of ergodicity for imprecise Markov chains
   - The uniform coefficient of ergodicity
   - The weak coefficient of ergodicity

Discrete time Markov chains

We consider a process which can be in one of the states from a set $\mathcal{X}$ at any time n.

The states are changed in steps.

The probability of moving from a state x_i to another state x_j is denoted by p_{ij} and called transition probability.



The process is assumed to be memoryless which means that the transition probabilities only depend on the current state and not on the history of the process.

A Markov chain is a sequence of states $x_1 x_2 \ldots x_n \ldots$

Examples

1. Switching TV channels. When a program ends the viewer can switch to another channel with some probability depending on the current channel. The sequence of channels viewed is then a Markov chain.
2. Social mobility. Let a state be the highest level of education achieved a person and let the state at the next step be the highest level of education achieved by her first born child.

Mathematical model

Let X_n be the random variable denoting the state at time n.

A Markov chain is the sequence $(X_n)_{n\in\mathbb{N}}$ of random variables.

Given the first n states, all that can be said about the next state is the conditional probability

$$P(X_{n+1} = x_j | X_n = x_i, X_{n-1} = x_{n-1}, \dots X_1 = x_1) = P(X_{n+1} = x_j | X_n = x_i) = p_{ij}$$

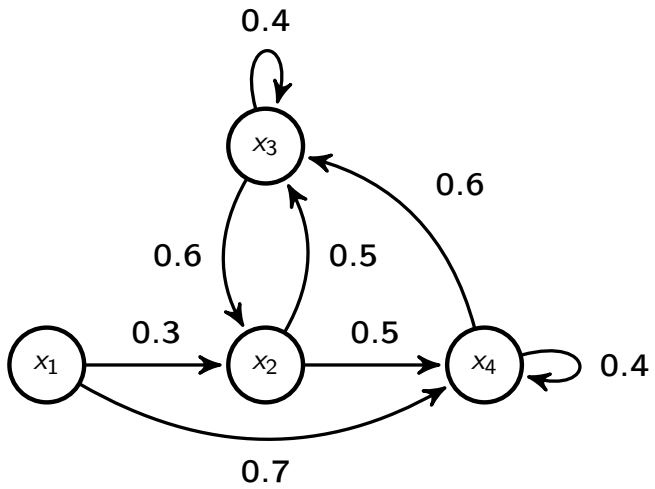
which means that the probability of the value of X_{n+1} only depends on the value of X_n and not on the values at earlier stages.

Transition matrix

Transition probabilities $p_{ij} = P(X_{n+1} = x_j | X_n = x_i)$ form a matrix P that is called transition matrix.

Since the chain will be in exactly one state at time $n+1$ the sum of $\sum_j p_{ij}$ where j runs over all elements in $\mathcal{X}$ must be 1.

Equivalently, the sum of elements in each row of a transition matrix is 1: (row) stochastic matrix.



Transition matrix:

$$P = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{pmatrix} 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{pmatrix} \end{array}$$

Calculating probabilities at time n

Let $(X_i)_{i=0}^n$ be a Markov chain with the transition matrix P.

Let m^n denote the probability mass function over the states at time n: $m_i^n = m^n(x_i) = P(X_n = x_i)$. It is a vector of probabilities

$$m^n = (m_1^n, \ldots, m_s^n).$$

Sometimes the initial state is given, otherwise we have an initial probability distribution m^0 over the states.

The probability mass function corresponding to probability distribution of X_n is then obtained as

$$m^n = m^0 P^n,$$

where P^n denotes the nth matrix power of P and the multiplication is the usual matrix multiplication.

Example

Let

$$P = \begin{bmatrix} 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}$$

and suppose that the chain starts in x_1:

$$m^0 = (1, 0, 0, 0).$$

Then we have the following sequence of distributions at the following times:

n	m_1^n	m_2^n	m_3^n	m_4^n
0	1	0	0	0
1	0	0.3	0	0.7
2	0	0	0.57	0.43
3	0	0.342	0.486	0.172
4	0	0.2916	0.4686	0.2398
5	0	0.28116	0.47712	0.24172

Long term distributions

Let $(X_n)_{n=0}^{\infty}$ be a Markov chain and m^n the probability mass functions corresponding to X_n.

Often we are interested in the limit probability mass function

$$m^{\infty} = \lim_{n \to \infty} m^n.$$

If exists, the above limit may depend on m^0 or not.

If the limit m^{∞} exists for some m^0 then we say that the chain converges.

The case where m^{∞} is independent of m^0 is of a special interest in the analysis of Markov chains.

The limit distribution m^{∞} is always an invariant (stationary) distribution, i.e. has the property that $m^{\infty} P = m^{\infty}$; and if it is the unique limit distribution then it is also the unique invariant distribution.

Examples

1. The Markov chain with the transition matrix

$$P = \begin{bmatrix} 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}$$

   has the unique limit distribution
   $m^{\infty} = (0, 0.285714, 0.47619, 0.238095)$.

2. Let $P = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. Then the chain starting in x_1 moves alternately between x_1 and x_2. So it does not converge. However, it does have an invariant distribution, which is $(0.5, 0.5)$.

3. If $P = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ then every probability distribution is invariant. The convergence is therefore not unique.

The Perron-Frobenius theorem

Theorem (The Perron-Frobenius theorem)

Let P be transition matrix of a Markov chain (X_i) such that all entries of P^r are strictly positive for some $r > 0$. Then there is a unique probability mass function m such that $mP = m$. Moreover, for every initial m^0 we have that

$$m = \lim_{n \to \infty} m^0 P^n$$

and m has all components strictly positive.

Definition

A transition matrix P with the property that P^r has all entries strictly positive for some $r > 0$ is said to be regular.

Example

Let

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \\ 0 & 0.6 & 0.4 \end{bmatrix}$$

Then

$$P^2 = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.24 & 0.46 & 0.3 \\ 0.36 & 0.48 & 0.16 \end{bmatrix}$$

P is therefore regular and has the unique limit distribution $m^\infty = (0.285714, 0.47619, 0.238095)$.

Coefficients of ergodicity

Perron-Frobenius theorem gives sufficient conditions for convergence of a Markov chain.

There are though non-regular Markov chains that also uniquely converge.

A more general way to determine whether a Markov chain is convergent is by using coefficients of ergodicity.

They give necessary and sufficient conditions for convergence.

Properties of coefficients of ergodicity

A coefficient of ergodicity τ assigns a real value $\tau(P)$ to any transition matrix P so that

1. $0 \leq \tau(P) \leq 1$;
2. $\tau(P_1P_2) \leq \tau(P_1)\tau(P_2)$;
3. $\tau(P) = 0$ if and only if $\mathrm{rank}(P) = 1$ or, equivalently, $P = \mathbf{1}v$ for some vector v.

Contraction mappings in euclidean spaces

Let $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a metric.

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ with the property that

$$d(f(x), f(y)) \leq kd(x, y),$$

where $0 \leq k < 1$ is a constant, is said to be a contraction mapping on the space $\mathbb{R}^n$.

Every contraction mapping on a metric space has a unique fixed point, i.e. x such that $f(x) = x$.

Coefficients of ergodicity as measures of contraction

Let P be a transition matrix.

The mapping $x \mapsto xP$ is a non-expansive mapping.

A coefficient of ergodicity $\tau(P)$ is defined as

$$\tau(P) = \sup_{x,y} \frac{d(xP, yP)}{d(x, y)}$$

Different metrics generate different ergodicity coefficients.

Usually the following metric is used

$$d(x, y) = \frac{1}{2} \sum_{i=1}^{s} |x_i - y_i|$$

Coefficients of ergodicity as measures of contraction

By the definition of ergodicity coefficient we have that

$$d(xP, yP) \leq \tau(P)d(x, y)$$

which particularly implies that in the case where $\tau(P^r) < 1$ for some $r > 0$ the Markov chain with the transition matrix P uniquely converges.

It is also easy to see the converse implication, that unique convergence of a Markov chain implies the existence of some $r > 0$ such that $\tau(P^r) < 1$.

Clearly, a regular transition matrix has the coefficient of ergodicity less than 1; however, converse does not necessarily hold. Therefore, coefficients of ergodicity give stronger conditions for unique convergence than regularity.

The calculation of the coefficient of ergodicity

Let P be a transition matrix whose ith row is P_i.

Then its coefficient of ergodicity is equal to the maximal distance between its rows:

$$\tau(P) = \max_{1 \leq i < j \leq s} d(P_i, P_j),$$

where

$$d(P_i, P_j) = \frac{1}{2} \sum_{k=1}^{s} |p_{ik} - p_{jk}|.$$

Example

Let us calculate the coefficient of ergodicity for the matrix:

$$P = \begin{bmatrix} 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}$$

We have

$$d(P_1, P_2) = 0.5 \qquad d(P_2, P_3) = 0.6$$
$$d(P_1, P_3) = 0.7 \qquad d(P_2, P_4) = 0.1$$
$$d(P_1, P_4) = 0.6 \qquad d(P_3, P_4) = 0.6$$

So $\tau(P) = 0.7$, but P is not regular.

Contents

1. Markov chains
   - Markov chains
   - Calculating probabilities at time n
   - Long term distributions
   - Coefficients of ergodicity
2. Imprecise probabilities
   - Credal sets
   - Lower and upper expectations
   - Coherent lower and upper previsions
3. Imprecise Markov chains
   - Markov chains with probability intervals
   - Markov set chains
   - Operators approach
4. Convergence of imprecise Markov chains
   - Regularity
   - Accessibility relations
   - Coefficients of ergodicity for imprecise Markov chains
   - The uniform coefficient of ergodicity
   - The weak coefficient of ergodicity

Models of imprecise probabilities

Imprecision in probability distributions can be modelled in various more or less general ways:

credal sets are closed convex sets of probability distributions which are assumed as candidates for the true distribution;

lower and upper expectations can be assigned to any credal set;

lower and upper prevision denote a generalised notion of lower and upper expectations for a given set of gambles (real valued maps); they are also interpreted as buying and selling prices respectively;

interval probabilities are intervals assigned to collections of events; the interval $P(A) = [L(A), U(A)]$ is assumed to contain the unknown true probability of A;

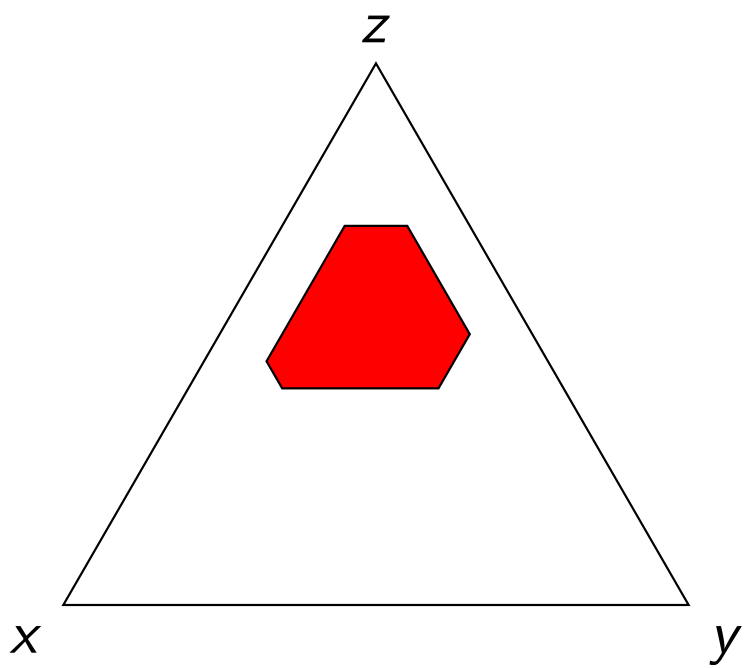
coherence assumptions are conditions imposed on lower and upper previsions (probabilities) in order to ensure a one-to-one correspondence with credal sets.

Credal sets

Credal sets are among most general imprecise probability models.

A credal set $\mathcal{C}$ is a compact convex set of probability distributions.

It can be represented with a subset in the probability simplex.



Sets of transition matrices with separately specified rows

Let $\mathcal{P}$ be a set of matrices formed so that the ith row is any element from a credal set $\mathcal{P}_i$.

Then we say that the set of matrices has separately specified rows, which we usually assume for imprecise transition matrices.

Example

The convex set of matrices

$$\left\{ \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix} : \alpha \in [0.3, 0.4] \right\}$$

does not have separately specified rows. The set with

the same row sets and that has separately specifified rows is

$$\left\{ \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} : \alpha, \beta \in [0.3, 0.4] \right\}$$

Lower and upper expectations

Given a credal set $\mathcal{C}$ the lower expectation with respect to $\mathcal{C}$ is a real valued mapping that assigns a real value $\underline{E}_{\mathcal{C}} f$ to every real valued function f on $\mathcal{X}$ and is defined as

$$\underline{E}_{\mathcal{C}}[f] = \min_{m \in \mathcal{C}} E_q[f]$$

and similarly the upper expectation is defined with

$$\overline{E}_{\mathcal{C}}[f] = \max_{m \in \mathcal{C}} E_q[f].$$

Every convex set is fully specified with the corresponding lower or upper expectation. Also, the lower expectation determines the upper expectation by

$$\overline{E}_{\mathcal{C}}[f] = -\underline{E}_{\mathcal{C}}[-f].$$

Lower and upper expectation operators

Let $\overline{T}_i = \overline{E}_{\mathcal{P}_i}$ be the upper expectation corresponding to the ith row of $\mathcal{P}$.

We may define the upper expectation operator

$$\overline{T} = \begin{bmatrix} \overline{T}_1 \\ \vdots \\ \overline{T}_s \end{bmatrix}.$$

Let a real valued map on $\mathcal{X}$ be given. Then

$$\overline{T}[f] = \begin{bmatrix} \overline{T}_1[f] \\ \vdots \\ \overline{T}_s[f] \end{bmatrix}$$

is again a real valued function on $\mathcal{X}$.

Coherent lower and upper previsions

Let $\mathcal{F}$ be a set of real valued mappings $\mathcal{X} \to \mathbb{R}$.

Then $\underline{P}$ and $\overline{P}$: $\mathcal{F} \to \mathbb{R}$ denote lower and upper prevision respectively.

Coherence ensures the existence of a credal set $\mathcal{M}(\underline{P})$.

The corresponding upper expectation operator $\overline{E}$ restricted to $\mathcal{F}$ coincides with $\overline{P}$ and is called the natural extension of $\overline{P}$.

If they are coherent, it does not matter if upper or lower prevision is given, since always

$$\underline{P}(f) = -\overline{P}(-f).$$

Lower and upper probabilities as lower-upper previsions

Let $\mathcal{A}$ be a collection of subsets of $\mathcal{X}$.

The interval probability P is then a special case of lower-upper prevision defined on the set of indicator functions of the sets in $\mathcal{A}$.

Again it does not matter if upper or lower probability is given, because of

$$L(A) = 1 - U(A^c).$$

Example

Let $\{1,3\} = A \subseteq \{1,2,3\}$ then 1_A is the function corresponding to the vector $(1,0,1)$.

Indicator functions of sets are exactly 0-1 valued functions on $\mathcal{X}$, while the domain of a general upper prevision $\overline{P}$ may contain more general functions, for instance, $f = (1,0,-2)$.

Contents

1. Markov chains
   - Markov chains
   - Calculating probabilities at time n
   - Long term distributions
   - Coefficients of ergodicity
2. Imprecise probabilities
   - Credal sets
   - Lower and upper expectations
   - Coherent lower and upper previsions
3. Imprecise Markov chains
   - Markov chains with probability intervals
   - Markov set chains
   - Operators approach
4. Convergence of imprecise Markov chains
   - Regularity
   - Accessibility relations
   - Coefficients of ergodicity for imprecise Markov chains
   - The uniform coefficient of ergodicity
   - The weak coefficient of ergodicity

Imprecise Markov chains

Given a Markov chain two things can be imprecise:

Initial distribution can be represented as an imprecise probability distribution over the set of states using one of the models described earlier.

Transition matrix: conditional on currently being in a state x_i the probability of transition to a new state is described with the ith row of the transition matrix, which makes a row a conditional probability, and can be imprecise as well. If rows of transition matrix are imprecise then we have an imprecise transition matrix.

Markov chains with probability intervals

Before turning to the more general models it is most intuitive to start with the model with probability intervals.

Recall that the behaviour of a Markov chain $(X_n)_{n=0}^{\infty}$ is fully probabilistically described if the initial and transition probabilities are given. That is if the following probabilities are known

$$P(X_0 = x_i) = m_i$$

and

$$P(X_{n+1} = x_j | X_n = x_i) = p_{ij}.$$

The most basic way how imprecision about these probabilities can be described is by replacing any of the above probabilities by an interval. Thus we have the following imprecise Markov model:

$$P(X_0 = x_i) = [\underline{m}_i, \overline{m}_i]$$

and

$$P(X_{n+1} = x_j | X_n = x_i) = [\underline{p}_{ij}, \overline{p}_{ij}].$$

According to the above model, any probability mass function m that satisfies $m_i \in [\underline{m}_i, \overline{m}_i]$ for every $i = 1, \ldots, s$ can be considered as an initial distribution, and similarly any transition matrix between $\underline{P}$ and $\overline{P}$ can be the transition matrix at some time n.

Thus a set of initial distributions and transition matrices can be assigned to this model. However, more general sets are possible as well that cannot be represented with the constraints of this type.

Markov set chains

A Markov set chain is an imprecise Markov chain where the initial distribution $\mathcal{C}_0$ is an arbitrary compact set of probabilities and the transition set $\mathcal{P}$ is an arbitrary compact set of transition matrices.

The sets of distributions at time n are then obtained as

$$\mathcal{C}_n = \mathcal{C}_0\mathcal{P}^n = \mathcal{C}_{n-1}\mathcal{P} = \{m^0 p^1 p^2 \dots p^n \colon m^0 \in \mathcal{C}_0, p^i \in \mathcal{P}, 1 \leq i \leq n\}.$$

Even if the sets $\mathcal{C}_0$ and $\mathcal{P}$ are both convex, it is not necessary that $\mathcal{C}_n$ are convex without the assumption that rows of $\mathcal{P}$ are separately specified.

This makes general Markov set chains hardly tractable. From now on we will adopt the assumption that all transition sets have separately specified rows.

Markov set chains with separately specified rows

In the case where the rows of the transition set are separately specified, the sets of distributions $\mathcal{C}_n$ are all convex if the initial set and the transition set are convex.

However, the number of extreme points, and consequently the number of constraints needed to represent them may increase rapidly with the number of steps.

Although it is possible to at least obtain approximations with a fixed number of constraints. One convenient way to obtain such approximations is by the use of lower or upper expectation operators.

Approximation with a given set of constraints

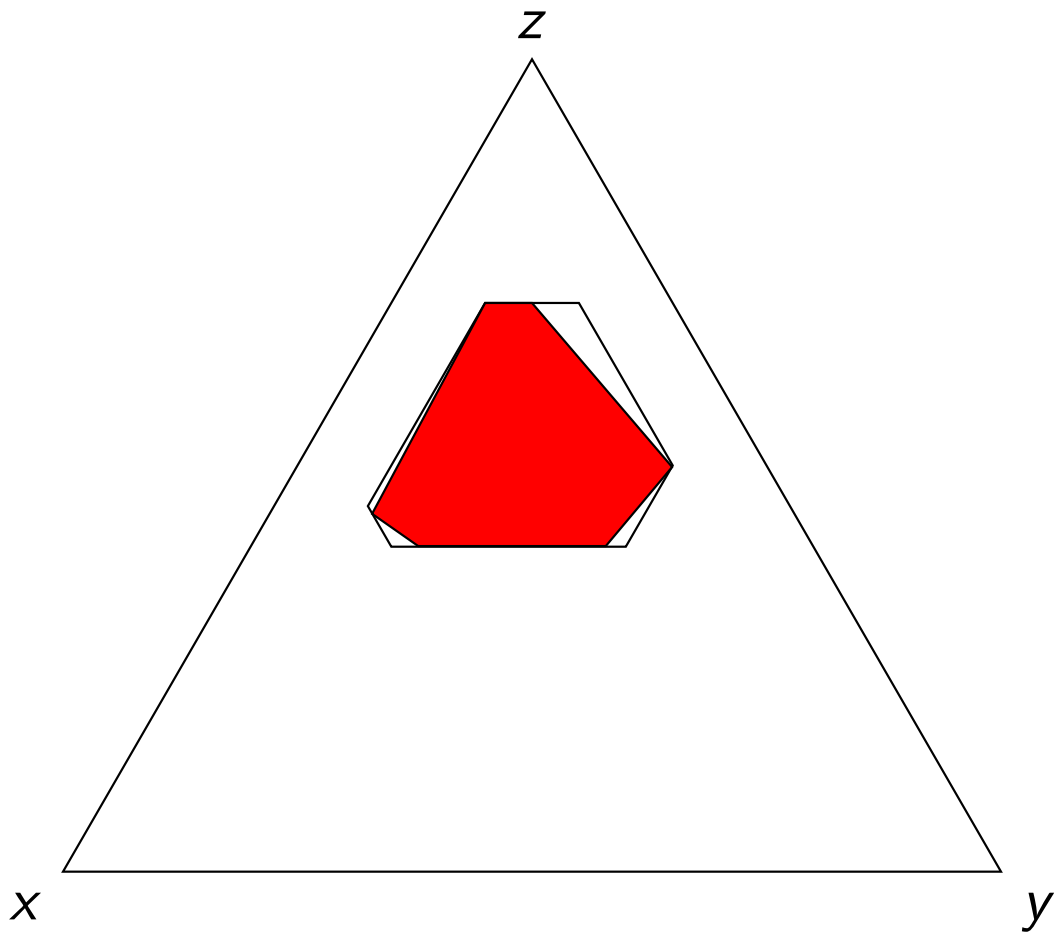


Figure: Approximation of a credal set with an interval probability

Imprecise Markov chains with upper expectation operators

An imprecise Markov chain is uniquely described by the imprecise initial distribution and transition matrix.

If the initial distribution is a credal set (=compact convex set of probability distributions) and if the transition set has separately specified rows, the same can be described by the initial upper expectation operator $\overline{E}_0$ and the upper transition operator $\overline{T}$.

The question remains how the two can be used to calculate the upper expectation operators corresponding to credal sets $\mathcal{C}_n$.

Calculation of upper expectation operators for the sets $\mathcal{C}_n$

Denote the upper expectation operator corresponding to the credal set $\mathcal{C}_n$ by $\overline{E}_n$, and let $f : \mathcal{X} \to \mathbb{R}$ be an arbitrary mapping. Then we have:

$$\overline{E}_n f = \overline{E}_0 \underbrace{\overline{T} \ldots \overline{T}}_{n \text{ instances}} f$$

The above calculation goes as follows: f is a real valued function on $\mathcal{X}$ and $\overline{T}$ maps it to another real valued function on $\mathcal{X}$ and so on n times. Finally the upper expectation $\overline{E}_0$ is applied to obtain $\overline{E}_n f$ which is a single real number.

Calculating interval probabilities with upper expectation operators

Suppose that we have an interval description of an imprecise Markov chain, with $[\underline{m}, \overline{m}]$ and $[\underline{P}, \overline{P}]$.

The sets $\mathcal{C}_n$ are in general not representable with the same type of constraints, but we might still be interested in the nearest probability interval to $\mathcal{C}_n$, i.e.

$$\underline{m}_i = \min_{m \in \mathcal{C}_n} m_i$$

and

$$\overline{m}_i = \max_{m \in \mathcal{C}_n} m_i.$$

Let $A \subseteq \mathcal{X}$ be any set of states. Then the upper probability $\overline{P}_n(A)$ equals to

$$\overline{P}_n(A) = \sup_{m \in \mathcal{C}_n} \sum_{i \in A} m_i = \overline{E}_n 1_A = \overline{E}_0 \overline{T}^n 1_A.$$

Example

Let the following lower and upper transition matrix be given for a Markov chain with three states:

$$\underline{P} = \begin{bmatrix} 0.3 & 0.4 & 0 \\ 0.2 & 0.6 & 0.1 \\ 0 & 0.3 & 0.4 \end{bmatrix} \quad \text{and} \quad \overline{P} = \begin{bmatrix} 0.5 & 0.5 & 0.2 \\ 0.3 & 0.7 & 0.2 \\ 0.1 & 0.4 & 0.6 \end{bmatrix}$$

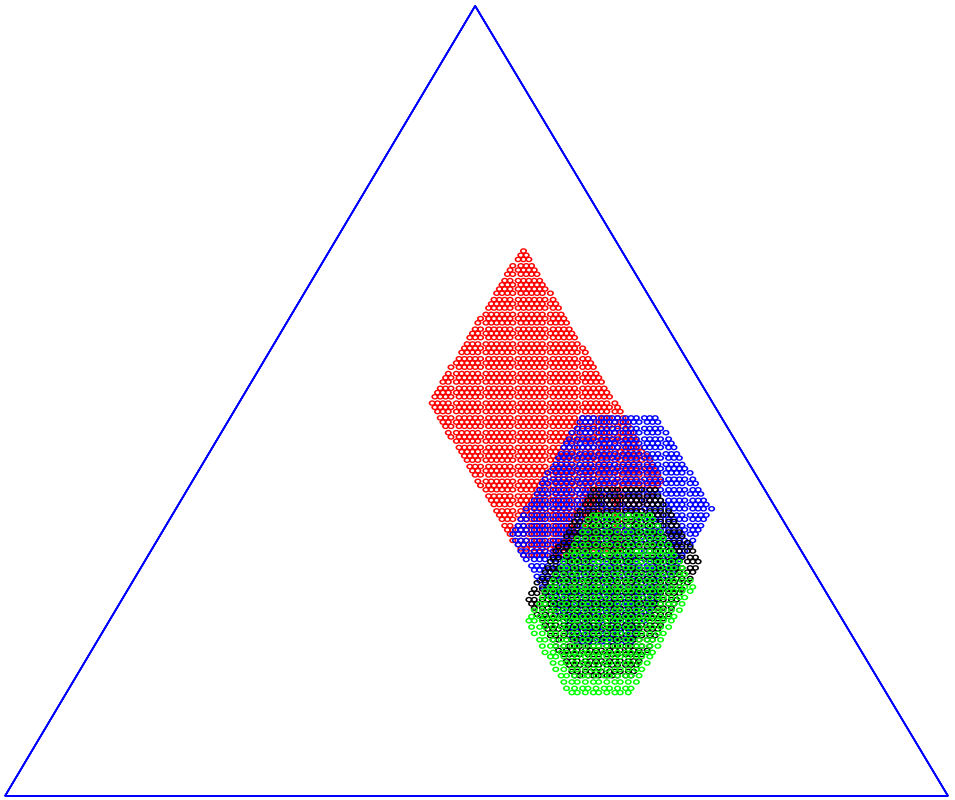
Let the initial lower and upper probability interval be

$$\underline{m} = [0.1, 0.2, 0.3] \quad \text{and} \quad \overline{m} = [0.3, 0.5, 0.7]$$

The lower and upper probabilities after 1, 2 and 3 steps are then:

steps	$\underline{m}_1$	$\underline{m}_2$	$\underline{m}_3$	$\overline{m}_1$	$\overline{m}_2$	$\overline{m}_3$
Initial	0.1	0.2	0.3	0.3	0.5	0.7
1	0.07	0.37	0.19	0.3	0.57	0.48
2	0.111	0.426	0.146	0.322	0.594	0.392
3	0.1327	0.446	0.1262	0.3352	0.6032	0.3568

Graphical representation of the approximating probability intervals



Contents

1. Markov chains
   - Markov chains
   - Calculating probabilities at time n
   - Long term distributions
   - Coefficients of ergodicity
2. Imprecise probabilities
   - Credal sets
   - Lower and upper expectations
   - Coherent lower and upper previsions
3. Imprecise Markov chains
   - Markov chains with probability intervals
   - Markov set chains
   - Operators approach
4. Convergence of imprecise Markov chains
   - Regularity
   - Accessibility relations
   - Coefficients of ergodicity for imprecise Markov chains
   - The uniform coefficient of ergodicity
   - The weak coefficient of ergodicity

Convergence of imprecise Markov chains

Let (X_n) be an imprecise Markov chain with initial credal set $\mathcal{C}_0$ and the corresponding upper expectation operator $\overline{E}_0$ and the transition set $\mathcal{P}$ with the upper transition operator $\overline{T}$.

Assume that all the above sets are convex and that the transition set has separately specified rows.

Then we say that the chain converges whenever $\overline{E}_n f$ converges for every f to some $\overline{E}_\infty f$. Moreover, the limit

$$\overline{E}_\infty f := \lim_{n\to\infty} \overline{E}_n f$$

may be independent of the initial upper expectation $\overline{E}_0$, in which case we have unique convergence.

Regular imprecise Markov chains

A sufficient condition for unique convergence is an obvious generalisation of regularity.

Definition

Let $\mathcal{P}$ be a transition set. If there is some $r > 0$ such that the set

$$\mathcal{P}^r = \{p_1 p_2 \dots p_r : p_i \in \mathcal{P}, 1 \leq i \leq r\}$$

contains only strictly positive matrices, then we say that $\mathcal{P}$ is a regular transition set.

Theorem

Let $\mathcal{P}$ be a closed regular transition set and $\mathcal{C}_0$ any closed set of probability distributions over $\mathcal{X}$. Then the sets $\mathcal{C}_n = \mathcal{C}_0 \mathcal{P}^n$ converge to a limit set $\mathcal{C}_\infty$, independent of the initial set.

Note that convexity of the initial or transition sets is not required in the last theorem.

Accessibility relations

Let x and y be any two states in $\mathcal{X}$. If it is possible that the Markov chain moves from x to y in n steps then we write $x \overset{n}{\rightsquigarrow} y$ and say that y is accessible from x in n steps.

If y is accessible from x in any number of steps then we say that y is accessible from x and write $x \rightsquigarrow y$.

If x and y are both accessible from one another then we say that x and y communicate and write $x \leftrightsquigarrow y$. The relation $\leftrightsquigarrow$ is clearly an equivalence relation. So the set of states is partitioned in the so called communication classes.

Accessibility relations for imprecise Markov chains

Let x and y be two states. Then we say that the state y is accessible from x in n steps if it is possible to move from x to y. This means that the upper probability of moving from x to y in n steps is positive.

In terms of upper transition operator this means the following. Suppose that we are in the state x at time k, which means that we put $E_k f = f(x)$.

The target upper probability is then $\overline{P}^n_{xy} := E_k \overline{T}^n 1_{\{y\}}$.

Classification of communication classes

A communication class D is

maximal if $x \rightsquigarrow y$ implies $y \rightsquigarrow x$ for every $x \in D$;

transient if it is not maximal.

If there is a unique maximal class then it is called the top class and contains the states that are accessible from any other state.

A communication class D is regular under the accessibility relation if there is some $n \in \mathbb{N}$ such that for every $k \geq n$ and for every two elements $x, y \in D$ the relation $x \overset{k}{\rightsquigarrow} y$ holds.

Regularly absorbing imprecise Markov chains

If there is a top class for an accessibility relation $\rightsquigarrow$ which is regular then the relation is top class regular.

An imprecise Markov chain whose accessibility relation is top class regular and for every $y \in \mathcal{X} - R$ there is some $n \in \mathbb{N}$ such that $\underline{T}^n I_R(y) > 0$, where R is the top class, then it is regularly absorbing.

Theorem (Perron-Frobenius theorem, Upper expectation form (by De Cooman, Hermans and Quaeghebeur))

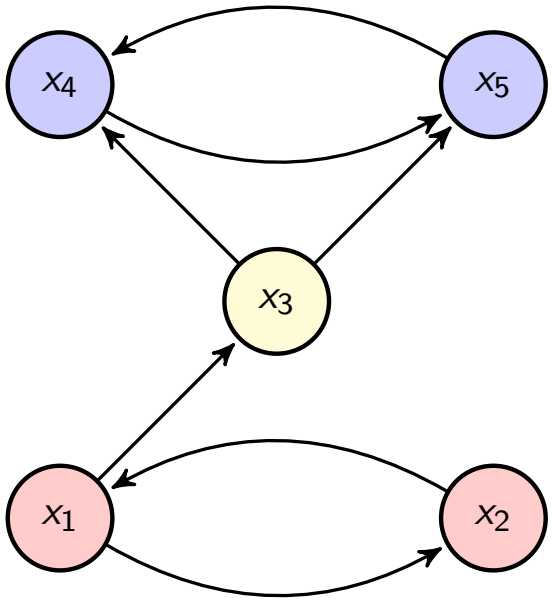
Let an imprecise Markov chain be regularly absorbing. Then for any initial upper expectation $\overline{E}_0$, the upper expectations $\overline{E}_n = \overline{E}_0 \overline{T}^n$ converge pointwise to the same upper expectation $\overline{E}_\infty$:

$$\lim_{n\to\infty} \overline{E}_n f = \lim_{n\to\infty} \overline{E}_0(\overline{T}^n f) =: \overline{E}_\infty f$$

for every real valued function f on $\mathcal{X}$.

Example

Consider the following diagram, where the arrow denotes that it is possible to move from one state to another.



We have three communication classes $C_1 = \{x_1, x_2\}$, $C_2 = \{x_3\}$ and $C_3 = \{x_4, x_5\}$.

C_3 is a regular top class: the chain is regularly absorbing, if the condition $\underline{T}^n I_{C_3}(y)$ is satisfied for some $n \in \mathbb{N}$ for all $y \in C_3^c$.

Coefficients of ergodicity

More about convergence can be said using coefficients of ergodicity that besides giving necessary and sufficient conditions also measure the rate of convergence. That is, how far from the limit distribution is a probability distribution at some time n.

For a precise transition matrix P the coefficient of ergodicity is calculated as the maximal distance $d(p_i, p_j)$ between ith and jth rows.

The coefficients for imprecise Markov chains are based on distances between "imprecise rows" of the imprecise transition matrix.

Distances between expectation operators

The distance between upper expectation operators $\overline{E}_1$ and $\overline{E}_2$ is defined as

$$d(\overline{E}_1, \overline{E}_2) = \max_{0 \leq f \leq 1} |\overline{E}_1[f] - \overline{E}_2[f]|.$$

Another coefficient of ergodicity uses the maximal distance between credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\max_{\substack{p_1 \in \mathcal{M}_1 \\ p_2 \in \mathcal{M}_2}} d(p_1, p_2) = \max_{A \subset S} \max\{\overline{E}_1[1_A] - \underline{E}_2[1_A], \overline{E}_2[1_A] - \underline{E}_1[1_A]\},$$

where $\underline{E}_i$ and $\overline{E}_i$ are the lower and upper expectation operator of the credal set $\mathcal{M}_i$.

The maximal distance between credal sets is the maximal distance between any two members of both sets.

2-alternating interval probabilities

In general it requires a lot of computation to calculate the distance between two expectation operators.

In the case of 2-alternating (concave) upper probabilities, i.e.

$$\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B) - \overline{P}(A \cap B),$$

the distance between the corresponding upper expectation operators turns out to be

$$d(\overline{E}_1, \overline{E}_2) = \max_{A \subseteq \mathcal{X}} |\overline{P}_1(A) - \overline{P}_2(A)|.$$

Every upper probability on a probability space with 3 or less elements is 2-alternating.

The uniform coefficient of ergodicity

Let $\mathcal{P}$ be a set of transition matrices and $\mathcal{P}_i$ it's ith row and $\overline{T}_i$ and $\underline{T}_i$ the corresponding upper and lower expectation operators.

Definition

The uniform coefficient of ergodicity is defined as

$$\tau(\mathcal{P}) = \sup_{p \in \mathcal{P}} \tau(p).$$

Alternatively, the uniform coefficient of ergodicity can be calculated as

$$\tau(\mathcal{P}) = \max_{1 \leq i,j \leq m} \max_{A \subset S} \overline{T}_i(1_A) - \underline{T}_j(1_A).$$

Convergence

A set $\mathcal{P}$ of transition matrices such that $\tau(\mathcal{P}^r) < 1$, for some $r > 0$, is called product scrambling.

Theorem

Let $\mathcal{P}$ be product scrambling. Then

$$d_H(\mathcal{M}_0\mathcal{P}^n, \mathcal{M}_\infty) \leq K\beta^n$$

for some constants K and β; and $\mathcal{M}_\infty$ is a unique credal set, independent of $\mathcal{M}_0$.

The weak coefficient of ergodicity

The uniform coefficient of ergodicity in general gives too strong conditions for convergence for the case of upper transition operators. Necessary and sufficient conditions for this case are obtained using the following weaker coefficient.

Let $\overline{T}$ be an upper transition operator with rows $\overline{T}_i$.

The weak coefficient of ergodicity is defined with

$$\rho(\overline{T}) = \max_{i,j} d(\overline{T}_i, \overline{T}_j).$$

Convergence

An upper transition operator $\overline{T}$ such that $\rho(\overline{T}^r) < 1$, for some $r > 0$, is called weakly product scrambling.

Theorem

Let $\overline{T}$ be weakly product scrambling. Then

$$d(\overline{E}_0 \overline{T}^n, \overline{E}_\infty) \leq K\beta^n$$

for some constants K and $\beta < 1$; and $\overline{E}_\infty$ is the unique upper expectation operator, independent from $\overline{E}_0$.
Moreover, $\overline{T}$ being weakly product scrambling is equivalent to unique convergence.

Example

Let $\mathcal{X} = \{x_1, x_2, x_3\}$. Denote all non-trivial subsets of $\mathcal{X}$ in lexicographical order:

$$A_1 = \{x_1\} \qquad A_4 = \{x_3\}$$
$$A_2 = \{x_2\} \qquad A_5 = \{x_1, x_3\}$$
$$A_3 = \{x_1, x_2\} \qquad A_6 = \{x_2, x_3\}$$

An upper probability is then a vector of values, where ith component contains the upper probability $\overline{P}(A_i)$.

Similarly, an upper transition operator can be represented with a matrix with n rows and $2^n - 2$ columns.

Let

$$\overline{P} = \begin{bmatrix} 0.5 & 0.5 & 1 & 0.2 & 0.6 & 0.7 \\ 0.3 & 0.7 & 0.9 & 0.2 & 0.4 & 0.8 \\ 0.1 & 0.4 & 0.5 & 0.6 & 0.7 & 1 \end{bmatrix}$$

be an upper transition operator. The distances between the upper expectation operators corresponding to its rows are the maximal absolute distances between elements in the rows:

$$d(\overline{P}_1, \overline{P}_2) = 0.2$$
$$d(\overline{P}_1, \overline{P}_3) = 0.5$$
$$d(\overline{P}_2, \overline{P}_3) = 0.4$$

The weak coefficient of ergodicity is therefore equal to $\rho(\overline{P}) = 0.5$.

Convergence