Distances between probability measures and coefficients of ergodicity for imprecise Markov chains

Damjan Škulj

University of Ljubljana, Faculty of Social Sciences, Slovenia

Second Workshop on Principles and Methods of Statistical Inference with Interval Probability
Munich, September 2009

Outline

Markov chains

- A (discrete time) Markov chain is a random process with the Markov property...
- ...which means that future states only depend on the present state and not on the past states.
- This dependence is expressed through transition probabilities:

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

  for every $n \in \mathbb{N}$.
- The knowledge about the first state is given by initial probability

$$P(X_0 = j) = q_j.$$

Imprecise Markov chain

- A Markov chain has many parameters which may not be known precisely.
- An imprecise Markov chain is a Markov chain where the imprecise knowledge of parameters is built into the model...
- ...and reflected in the results:
  - probabilities of states at future steps;
  - long term distributions.

Representation with sets of probabilities

- The uncertainty of parameters can be expressed through sets of probabilities:
  - instead of single precisely known initial and transition probabilities we take sets of possible candidates.
- Let $\mathcal{M}_n$ be the set of possible distributions at step n and $\mathcal{P}$ the set of possible transition matrices.
- The following relation must hold:

$$\mathcal{M}_{n+1} = \mathcal{M}_n \cdot \mathcal{P}.$$

- Another important question is whether the sets converge to some limit set $\mathcal{M}_\infty$ and what can we say about this convergence.

Convexity of the sets $\mathcal{M}_n$

- The sets $\mathcal{M}_0$ and $\mathcal{P}$ are usually assumed to be convex.
- $\mathcal{M}_n$, for $n > 0$, are not necessarily convex any more.
- What are sufficient conditions for convexity?
- Answer: Rows must be separately specified.

Separately specified rows

- Let $\mathcal{P}$ be a convex set of transition matrices.
- $\mathcal{P}_i$ the set of all possible i-th rows. It is a convex set.

Definition

$\mathcal{P}$ has separately specified rows if the choice of i-th row is independent of the choice of other rows.

Theorem

If $\mathcal{M}_0$ is convex and $\mathcal{P}$ is convex with separately specified rows then all $\mathcal{M}_n$ are convex.

Representation with expectation operators

- Convex sets of probabilities can equivalently be expressed through (lower) expectation operators

$$\underline{P}_n[f] = \min_{p \in \mathcal{M}_n} E_p[f],$$

where f is a real valued map on the set of states.

- Convex sets of transition matrices can be represented through (lower) transition operators:

$$\underline{T}[f] = \begin{pmatrix} \underline{T}_1[f] \\ \vdots \\ \underline{T}_m[f] \end{pmatrix}.$$

Coefficients of ergodicity

- Coefficients of ergodicity or contraction coefficients measure the rate of convergence of Markov chains.
- Let p be a stochastic matrix without zero columns.
- The value $\tau(p)$ of a coefficient of ergodicity satisfies:
    - $0 \le \tau(p) \le 1$;
    - $\tau(p_1 p_2) \le \tau(p_1)\tau(p_2)$;
    - $\tau(p) = 0$ iff p has rank 1: $p = \mathbf{1}\boldsymbol{v}$ for some vector $\boldsymbol{v}$.
- Clearly: $\tau(p) < 1$ implies that powers p^n converge to a matrix with rank 1, which is equivalent to unique convergence of the corresponding Markov chain.

Calculation and generalisation

- A coefficient of ergodicity can be calculated as

$$\tau(p) = \max_{i,j} d(p_i, p_j),$$

where p_i and p_j are the i-th and the j-th row of p; and

$$d(p_i, p_j) = \max_{A \subseteq \Omega} |p_i(A) - p_j(A)|.$$

- We can generalise this to imprecise Markov chains if the distance function d is generalised to imprecise probabilities.
- This can be done in different ways with different implications to imprecise Markov chains.

Hausdorff metric

- The Hausdorff metric makes the set of compact non-empty subsets of a metric space a metric space.
- It is defined by:

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\}.$$

- It can be used to measure distances between closed sets of probability distributions on finite measurable sets.
- The Hausdorff distance is equal to 0 iff the sets are equal.

Distances between expectation operators

Using lower expectation operators there is another way to measure the distance between imprecise probabilities:

$$d(\underline{P}, \underline{P}') = \max_{0 \le f \le 1} |\underline{P}(f) - \underline{P}'(f)|.$$

Theorem (Škulj, Hable (2009))

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be closed convex sets of probabilities and let $\underline{P}_1$ and $\underline{P}_2$ be their lower expectation operators. Then:

$$d(\underline{P}_1, \underline{P}_2) = d_H(\mathcal{M}_1, \mathcal{M}_2).$$

Maximal distance between imprecise probabilities

Sometimes we need the maximal distance between sets of probabilities:

$$\max_{\substack{p_1 \in \mathcal{M}_1 \\ p_2 \in \mathcal{M}_2}} d(p_1, p_2)$$

Theorem

$$\max_{A \subset \Omega} \max\{\overline{P}_1(1_A) - \underline{P}_2(1_A), \overline{P}_2(1_A) - \underline{P}_1(1_A)\}.$$

Uniform coefficient of ergodicity

- Let $\mathcal{P}$ be a set of transition matrices.
- Let $\mathcal{P}_i$ denote it's i-th row and $\overline{T}_i$ and $\underline{T}_i$ be its upper and lower expectation operators.
- The uniform coefficient of ergodicity is defined as

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

by Hartfiel (1998).

- Using a previous result we can see that

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

where $\underline{T}_i$ and $\overline{T}_j$ are lower and upper expectation operators corresponding to $\mathcal{P}_i$ and $\mathcal{P}_j$ respectively.

Convergence

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

Theorem (Hartfiel (1998))

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

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

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

Weak coefficient of ergodicity

- The previous requirements are clearly sufficient for convergence, but too strong (this follows from the results of de Cooman, Hermans, Quaeghebeur (2009)).
- We need another coefficient of ergodicity to describe this type of convergence.
- Instead of taking maximal possible distances between rows of imprecise transition matrices, we take a distance that reflects only the difference between the rows.
- Hausdorff distance seems a good candidate.

Definition of the weak coefficient of ergodicity

- We define the weak coefficient of ergodicity by means of lower expectation operators.
- Let $\underline{T}$ be a lower transition operator with rows $\underline{T}_i$.
- Then we define the weak coefficient of ergodicity as

$$\max_{i,j} d(\underline{T}_i, \underline{T}_j),$$

which is equal to the Hausdorff distances between the corresponding sets.

Convergence

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

Theorem (Škulj, Hable (2009))

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

$$d(\underline{P}_0\underline{T}^h, \underline{P}_\infty) \leq K\beta^h$$

for some constants K and β; and $\underline{P}_\infty$ is a unique lower expectation operator, independent from $\underline{P}_0$.
Moreover, $\underline{T}$ being weakly product scrambling is equivalent to unique convergence.

Markov chains with absorbing states

- A Markov chain with the transition matrix of the form

$$\begin{bmatrix} 1 & \mathbf{0} \\ \boldsymbol{p} & Q \end{bmatrix},$$

  with $\boldsymbol{p} \neq \mathbf{0}$ and Q satisfies a regularity assumption is considered (Darroch and Seneta (1965)).
- Crossman, Coolen-Schrijner and Coolen (2009) study this problem with imprecise probabilities.
- The unique limit distribution is equal to $(1, \mathbf{0})$.
- If conditioned on non-absorption, the conditional distributions converge to a unique limit distribution.
- Crossman and Škulj (2009) generalise this convegence result to imprecise probabilities.

Another distance function

- A new distance measure between probability vectors was used in the proof.
- Let $(u_{-1}, \boldsymbol{u})$ and $(v_{-1}, \boldsymbol{v})$ be probability vectors with $\boldsymbol{u}, \boldsymbol{v} > 0$.
- We define:

$$d_1(\boldsymbol{u}, \boldsymbol{v}) = \frac{\overline{\alpha}_{\boldsymbol{u},\boldsymbol{v}} - \underline{\alpha}_{\boldsymbol{u},\boldsymbol{v}}}{\underline{\alpha}_{\boldsymbol{u},\boldsymbol{v}}}$$

  where

$$\underline{\alpha}_{\boldsymbol{u},\boldsymbol{v}} = \min_{i \geq 0} \frac{u_i}{v_i} \quad \text{and} \quad \overline{\alpha}_{\boldsymbol{u},\boldsymbol{v}} = \max_{i \geq 0} \frac{u_i}{v_i}.$$

- Clearly $\frac{1}{1-u_{-1}}\boldsymbol{u} = \frac{1}{1-v_{-1}}\boldsymbol{v}$ iff $d_1(\boldsymbol{u}, \boldsymbol{v}) = 0$.
- A similar distance function is used in Seneta's book (2006):

$$d_2(\boldsymbol{u}, \boldsymbol{v}) = \ln \frac{\overline{\alpha}_{\boldsymbol{u},\boldsymbol{v}}}{\underline{\alpha}_{\boldsymbol{u},\boldsymbol{v}}}.$$

Birkhoff's coefficient of ergodicity

Birkhoff's coefficient of ergodicity is defined as:

$$\frac{1-\sqrt{\phi(T)}}{1+\sqrt{\phi(T)}}$$

where

$$\phi(T)=\min_{i,j}\underline{\alpha}_{T_i,T_j}\underline{\alpha}_{T_j,T_i}.$$

and measures the rate of convergence with respect to the distance d_2.

Generalisation to imprecise probabilities

- The distance functions d_1 or d_2 can be generalised to sets of probabilities by a construction similar to the definition of the Hausdorff distance.
- Questions: What would be the corresponding distance function between lower or upper expectation operators that would correspond to these distances between sets of probabilities?
- How could the corresponding coefficients of ergodicity for the imprecise case be defined?