From the Uncoupling-Coupling Theorems of Markov Chains to Supremus Typicality and some Generalised Typicality Lemmata

Sheng Huang and Mikael Skoglund

Communication Theory
Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden

May 30, 2013
Internal Seminar
Outline

1. Stochastic Complementation and the Uncoupling-Coupling Theorems
   - Irreducible Markov Chains and the Strong Markov Property
   - Stochastic Complements
   - The Uncoupling Theorem
   - The Coupling Theorem

2. Supremus Typicality
   - Strongly Markov Typicality
   - Supremus Typicality
   - AEP of Supremus Typical Sequences
   - Typicality Lemmata of Supremus Typical Sequences

3. Applications and Generalisations
   - Linear Coding over Ring for Irreducible Markov Sources
   - Applications: Coding over Algebraic Structures
   - Generalisations

4. Thanks / References
   - Thanks
   - Bibliography
Irreducible Markov Chains

Definition 1

Given a homogeneous Markov chain $M = \{X^{(n)}\}_{-\infty}^{\infty}$ with a countable state space $\mathcal{X}$, the transition matrix of $M$ is defined to be the stochastic matrix $P = [p_{i,j}]_{i,j \in \mathcal{X}}$, where $p_{i,j} = \Pr\{X^{(2)} = j \mid X^{(1)} = i\}$. Moreover, $M$ is said to be irreducible if and only if $P$ is irreducible, namely, there exists no $\emptyset \neq A \subsetneq \mathcal{X}$ such that $P_{A,A^c} = [p_{i,j}]_{i \in A, j \in A^c} = 0$. $M$ is said to be finite-state if $\mathcal{X}$ is finite.

Example 2

The first stochastic matrix is irreducible, while the second one is not.
Theorem 3 (Strong Markov Property [Norris(1998)])

Given a Markov chain \( \mathcal{M} = \{ X^{(n)} \}_{-\infty}^{\infty} \) and a random stopping time \(-\infty < T < \infty, \Pr \{ X_{T+1} | X_T, X_{T-1}, \ldots, X_0, \ldots \} = \Pr \{ X_{T+1} | X_T \} \).

Example 4

Given a Markov chain \( \mathcal{M} = \{ X^{(n)} \}_{-\infty}^{\infty} \) with state space \( \mathcal{X} \) and a non-empty subset \( A \) of \( \mathcal{X} \), let

\[
T_{A,l} = \begin{cases} 
\inf \left\{ n > 0 | X^{(n)} \in A \right\}; & l = 1, \\
\inf \left\{ n > T_{A,l-1} | X^{(n)} \in A \right\}; & l > 1, \\
\sup \left\{ n < T_{A,l+1} | X^{(n)} \in A \right\}; & l < 1.
\end{cases}
\]

By the strong Markov property, \( \mathcal{M}_A = \{ X^{(T_{A,l})} \}_{-\infty}^{\infty} \) is Markov.
Invariant Distributions

Theorem 5

A finite-state irreducible Markov chain with transition matrix $P$ admits a unique invariant distribution $\pi$, i.e. $\pi P = \pi$.

Proved by [Breuer and Baum(2005), Theorem 2.31] and [Norris(1998), Theorem 1.7.7].

Problem 6

In Theorem 5, how to compute $\pi$ given $P$ efficiently for very large sized $P$?

Problem 7

In Example 4, if the $M$ is finite-state and irreducible, will $M_A$ be irreducible? If yes, what are the transition matrix and the invariant distribution of $M_A$?
Assumption: all Markov chains considered hereafter are homogeneous and finite-state unless specify. They are not necessarily stationary or their initial distributions are unknown.

Definition 8 (Definition 2.1 [Meyer(1989)])

In Example 4, assume that

\[ \mathbf{P} = \begin{bmatrix} \mathbf{P}_{A,A} & \mathbf{P}_{A,A^c} \\ \mathbf{P}_{A^c,A} & \mathbf{P}_{A^c,A^c} \end{bmatrix}, \]

where \( \mathbf{P}_{A,B} = [p_{i,j}]_{i \in A, j \in B} \) for \( A, B \subseteq \mathcal{X} \),

is the transition matrix of \( \mathcal{M} \), the stochastic complement of \( \mathbf{P}_{A,A} \) in \( \mathbf{P} \) is defined to be

\[ \mathbf{S}_A = \mathbf{P}_{A,A} + \mathbf{P}_{A,A^c}(1 - \mathbf{P}_{A^c,A^c})^{-1}\mathbf{P}_{A^c,A}. \]
The Uncoupling Theorem

**Theorem 9**

In Definition 8, if $P$ ($M$) is irreducible, then
- $S_A$ is stochastic [Meyer(1989), Theorem 2.1];
- $S_A$ is irreducible [Meyer(1989), Theorem 2.3];
- $S_A$ is the transition matrix of $M_A$ [Meyer(1989), Section 3];
- Let $\pi = [p_i]_{i \in \mathcal{X}}$ be the invariant distribution of $P$ ($M$). The invariant distribution of $S_A$ is

$$\left[ \frac{p_i}{\sum_{j \in A} p_j} \right]_{i \in A}$$

[Meyer(1989), Theorem 2.2].

**Remark 1**

One is kindly referred to [Norris(1998), Chapter 1] for the case that $M$ is not finite-state and other interested subchains of $M$. 

Sheng Huang and Mikael Skoglund  
Internal Seminar
The Coupling Theorem

Theorem 10 (Theorem 4.1 of [Meyer(1989)])

In Definition 8, if $\mathbf{P}$ ($\mathcal{M}$) is irreducible, let $\pi_A$ and $\pi_{A^c}$ be the invariant distribution of $\mathbf{S}_A$ and $\mathbf{S}_{A^c}$ (guaranteed by Theorem 9), respectively. We have that the invariant distribution $\pi$ of $\mathbf{P}$ is given by

$$
\pi = [\xi_1 \pi_A, \xi_2 \pi_{A^c}],
$$

where

$$
\bar{\xi} = [\xi_1, \xi_2]
$$

is the unique unitary vector such that

$$
\bar{\xi} \mathbf{C} = \bar{\xi}, \quad \mathbf{C} = \begin{bmatrix}
\pi_A \mathbf{P}_{A, A} \mathbf{e} & \pi_A \mathbf{P}_{A, A^c} \mathbf{e} \\
\pi_{A^c} \mathbf{P}_{A^c, A} \mathbf{e} & \pi_{A^c} \mathbf{P}_{A^c, A^c} \mathbf{e}
\end{bmatrix}.
$$ (1)

$\mathbf{C}$ is named the coupling matrix, and $\xi_A$ and $\xi_{A^c}$ are termed the coupling factors.
Strongly Markov Typicality

Definition 11 (Strongly Markov Typicality)

Let \( M = \{X^{(n)}\}_{-\infty}^{\infty} \) be an irreducible Markov chain with state space \( \mathcal{X} \), and \( P = [p_{i,j}]_{i,j \in \mathcal{X}} \) and \( \pi = [p_j]_{j \in \mathcal{X}} \) be its transition matrix and invariant distribution, respectively. For any \( \epsilon > 0 \), a sequence \( x \in \mathcal{X}^n \) of length \( n \) \((\geq 2)\) is said to be strongly Markov \( \epsilon \)-typical with respect to \( P \) if

\[
\begin{align*}
\left| \frac{N(i,j;x)}{N(i;x)} - p_{i,j} \right| < \epsilon; \\
\left| \frac{N(i;x)}{n} - p_i \right| < \epsilon,
\end{align*}
\]

(2)

where \( N(i,j;x) \) is the occurrences of sub-sequence \( [i,j] \) in \( x \) and \( N(i;x) = \sum_{j \in \mathcal{X}} N(i,j;x) \). The set of all strongly Markov \( \epsilon \)-typical sequences with respect to \( P \) in \( \mathcal{X}^n \) is denoted by \( T_\epsilon(n,P) \) or \( T_\epsilon \) for simplicity.
Supremus Typicality

Definition 12 (Supremus Typicality [Huang and Skoglund(2013)])

Follow the notation defined above. Given $\epsilon > 0$ and a sequence $\mathbf{x} = [x^{(1)}, x^{(2)}, \ldots, x^{(n)}] \in \mathcal{X}^n$ of length $n \geq 2 |\mathcal{X}|$, let $\mathbf{x}_A$ be the subsequence of $\mathbf{x}$ formed by all those $x^{(l)}$'s belong to $A \subseteq \mathcal{X}$ in the original ordering. $\mathbf{x}$ is said to be Supremus $\epsilon$-typical with respect to $\mathbf{P}$, if and only if $\mathbf{x}_A$ is strongly Markov $\epsilon$-typical with respect to $\mathbf{S}_A$ for all feasible non-empty subset $A$ of $\mathcal{X}$. The set of all Supremus $\epsilon$-typical sequences with respect to $\mathbf{P}$ in $\mathcal{X}^n$ is designated by $S_\epsilon(n, \mathbf{P})$ or $S_\epsilon$ for simplicity.

Let $\mathcal{X} = \{0, 1, 2\}$.

\[
\begin{align*}
00102 - 02210 - 02112 - 01022 - 01102 \in S_\epsilon(25, \mathbf{P}) \subseteq T_\epsilon(25, \mathbf{P}), \\
\implies \begin{cases} 
0002 - 0220 - 022 - 0022 - 002 \in T_\epsilon(18, S\{0,2\}); \\
0010 - 010 - 011 - 010 - 0110 \in T_\epsilon(17, S\{0,1\}).
\end{cases}
\end{align*}
\]
AEP of Supremus Typical Sequences I

Notation:

- Given a set $\mathcal{X}$, a partition $\bigcup_{k \in \mathcal{K}} A_k$ of $\mathcal{X}$ is a disjoint union of $\mathcal{X}$, i.e.
  \[ A_{k'} \cap A_{k''} \neq \emptyset \iff k' = k'', \quad \bigcup_{k \in \mathcal{K}} A_k = \mathcal{X} \text{ and } A_k \text{'s are not empty.} \]

- For discrete random variables $X$ and $Y$ with sample spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, $X \sim [p_i]_{i \in \mathcal{X}}$ and $(X, Y) \sim [p_i]_{i \in \mathcal{X}} [p_{i,j}]_{i \in \mathcal{X}, j \in \mathcal{Y}}$ state for
  \[ \Pr \{ X = i \} = p_i \text{ and } \Pr \{ X = i, Y = j \} = p_i p_{i,j}, \]
  for all $i \in \mathcal{X}$ and $j \in \mathcal{Y}$, respectively.

- Given a unitary vector $\pi$ and a stochastic matrix $P$ (not necessary that $\pi P = \pi$), we write
  \[ H(P|\pi) = H(Y|X), \]
  for $(X, Y) \sim \pi P$. 


Theorem 13 (AEP of Supremus Typicality [Huang and Skoglund(2013)])

Let $\mathcal{M} = \{X^{(n)}\}_{-\infty}^{\infty}$ be an irreducible Markov chain with state space $\mathcal{X}$, and $P = [p_{i,j}]_{i,j \in \mathcal{X}}$ and $\pi = [p_j]_{j \in \mathcal{X}}$ be its transition matrix and invariant distribution, respectively. For any $\eta > 0$, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall \epsilon_0 > \epsilon > 0$, $\forall n > N_0$ and $\forall x = [x^{(1)}, x^{(2)}, \ldots, x^{(n)}] \in S_\epsilon(n, P)$,

1. $\exp_2 [-n (H(P|\pi) + \eta)] < \Pr \{[X^{(1)}, X^{(2)}, \ldots, X^{(n)}] = x\} < \exp_2 [-n (H(P|\pi) - \eta)];$

2. $\Pr \{X \notin S_\epsilon(n, P)\} < \eta$, where $X = [X^{(1)}, X^{(2)}, \ldots, X^{(n)}]$; and

3. $|S_\epsilon(n, P)| < \exp_2 [n (H(P|\pi) + \eta)].$
Lemma 14 (Lemma D.1 of [Huang and Skoglund(2013)])

Given an irreducible Markov chain $\mathcal{M} = \{X^{(n)}\}_{-\infty}^{\infty}$ with finite state space $\mathcal{X}$, transition matrix $P$ and invariant distribution $\pi = [p_j]_{j \in \mathcal{X}}$. Let $\prod_{k=1}^{m} A_k$ be any partition of $\mathcal{X}$. For any $\eta > 0$, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall \epsilon_0 > \epsilon > 0$, $\forall \, n > N_0$ and $\forall \, x = \left[ x^{(1)}, x^{(2)}, \ldots, x^{(n)} \right] \in S_{\epsilon}(n, P)$,

$$|S_{\epsilon}(x)| < \exp_2 \left\{ n \left[ H \left( \text{diag} \left\{ \{S_k\}_{1 \leq k \leq m} \right\} | \pi \right) + \eta \right] \right\}, \quad (3)$$

where

$$S_{\epsilon}(x) = \left\{ \left[ y^{(1)}, y^{(2)}, \ldots, y^{(n)} \right] \in S_{\epsilon}(n, P) \! \mid \! y^{(l)} \in A_k \iff x^{(l)} \in A_k, \quad \forall \, 1 \leq l \leq n, \forall \, 1 \leq k \leq m \right\},$$

and $S_k$ is the stochastic complement of $P_{A_k, A_k}$ in $P$. 

Sheng Huang and Mikael Skoglund  
Internal Seminar
Typicality Lemmata of Supremus Typical Sequences II

Lemma 15 (Lemma D.2 of [Huang and Skoglund(2013)])

In Lemma 14, define \( \Gamma(x) = l \Leftrightarrow x \in A_l \). We have that

\[
|S_\epsilon(x)| < \exp_2 \left\{ n \left[ H(P|\pi) - \lim_{w \to \infty} \frac{1}{w} H(Y^{(w)}, \ldots, Y^{(1)}) + \eta \right] \right\}, \tag{4}
\]

where \( Y^{(w)} = \Gamma(X^{(w)}) \).

Lemma 15 can be easily generalised to jointly ergodic processes [Cover(1975)]

\[
\mathcal{M} = \left\{ \begin{bmatrix} X^{(n)} \\ Y^{(n)} \end{bmatrix} \right\}_{-\infty}^{\infty}
\]

to obtained related conditional typicality lemma. Please refer to [Huang and Skoglund(Submitted), Lemma III.5] for a special case that \( \mathcal{M} \) is i.i.d.
Remark 2

If in Lemma 14 $m = 1$, then both (3) and (4) are equivalent to

$$|S_\varepsilon(x)| < \exp_2 \left[ n \left( H(P|\pi) + \eta \right) \right].$$

Or, if $\mathcal{M}$ in Lemma 14 is i.i.d., then both (3) and (4) are equivalent to

$$|S_\varepsilon(x)| < \exp_2 \left[ n \left( H(X(1)) - H(Y(1)) + \eta \right) \right],$$

which is a special case of the generalized conditional typicality lemma [Huang and Skoglund(Submitted), Lemma III.5]. However, it is hard to determine which bound of these two is tighter in general. Nevertheless, (3) is seemingly easier to analyze, while (4) is more complicated for associating with the entropy rate of the ergodic process $\{Y^{(n)}\}_{-\infty}^{\infty}$. 
Remark 3

In Lemma 15, if $P = c_1 U + (1 - c_1) I$ with all rows of $U$ being identical, $I$ is an identity matrix and $0 \leq c_1 \leq 1$, then $M = \{Y^{(n)}\}_{-\infty}^{\infty}$ is Markovian by [Huang and Skoglund(2013), Lemma C.1]. As a conclusion,

$$|S_\epsilon(x)| < \exp_2 \left\{ n \left[ H(P|\pi) - \lim_{w \to \infty} H \left( Y^{(w)} \Big| Y^{(w-1)} \right) + \eta \right] \right\}$$

$$= \exp_2 \left\{ n \left[ H(P|\pi) - H(P'|\pi') + \eta \right] \right\},$$

where $P'$ and $\pi'$ are the transition matrix and the invariant distribution of $M'$ that can be easily calculated from $P$. However, in general $M'$ is ergodic, but not Markovian. Its entropy rate is difficult to obtain.
Remark 4

In Lemma 14 and Lemma 15, if

- $\mathcal{X}$ is a finite ring $\mathcal{R}$, then all costs $\mathcal{R}/I$ over a left (right) ideal $I$ give raise to a partition of $\mathcal{R} = \mathcal{X}$; and

- $\mathcal{X}$ is a finite dimension vector space (module) over a finite field (ring) $\mathcal{R}^n$ ($n \in \mathbb{N}^+$), then all affine subspaces (submodule) “parallel” to a given subspace (submodule) create a partition of $\mathcal{R}^n = \mathcal{X}$; and

- $\mathcal{X}$ is a finite group $G$, then all costs $G/H$ over some subgroup $H$ present a partition of $G = \mathcal{X}$; and etc.

From this, we see that the typicality lemmata can be tailored for specific situations.
Theorem 16 ([Huang and Skoglund(2013)])

Let $\mathcal{R}$ be a finite ring, and $\{X^{(n)}\}_{-\infty}^{\infty}$ be an irreducible Markov source with state space $\mathcal{R}$, transition matrix $P$ and invariant distribution $\pi$. For any $\epsilon > 0$, there exists $N_0 \in \mathbb{N}^+$, such that there exist a linear encoder $\phi : \mathcal{R}^n \rightarrow \mathcal{R}^k$ and a decoder $\psi : \mathcal{R}^k \rightarrow \mathcal{R}^n$ for all $n > N_0$ with

$$\Pr \{\psi (\phi (X^n)) \neq X^n\} < \epsilon,$$

provided that

$$k > \max_{0 \neq \mathcal{I} \subseteq \mathcal{R}} \frac{n}{\log |\mathcal{I}|} \min \left\{ H(S_{\mathcal{R}/\mathcal{I}} | \pi), \right.$$

$$H (P | \pi) - \lim_{w \rightarrow \infty} \frac{1}{w} H \left( Y^{(w)}_{\mathcal{R}/\mathcal{I}}, Y^{(w-1)}_{\mathcal{R}/\mathcal{I}}, \ldots, Y^{(1)}_{\mathcal{R}/\mathcal{I}} \right) \left\}, \right.$$ 

where $S_{\mathcal{R}/\mathcal{I}} = \text{diag} \left\{ \{S_A\}_{A \in \mathcal{R}/\mathcal{I}} \right\}$ with $S_A$ being the stochastic complement of $P_{A,A}$ in $P$ and $Y^{(i)}_{\mathcal{R}/\mathcal{I}} = X^{(i)} + \mathcal{I}$.
Applications: Coding over Algebraic Structures

The (conditional) typicality lemmata can be tailored to specific versions regarding certain algebraic structures (group, field, ring, rng, vector space, module and algebra). Based the these, related algebraic coding schemes can be developed. For instance, ring specials are found in [Huang and Skoglund(Submitted), Lemma III.5] and [Huang and Skoglund(2013), Lemma III.3 and Lemma III.4], and linear source coding techniques have been introduced in the previous works. These algebraic coding techniques are useful because

- It proves that linear coding over field is not optimal in a generalised problem ([Huang and Skoglund(2013), Problem 2]) of Körner–Marton [Körner and Marton(1979)]. The ring version can strictly outperform its field counterpart (in case of sources with/without memory);

- They help in proving that the Han–Kobayashi conjecture [Han and Kobayashi(1987), the converse of Theorem 2] is not valid when the number of sources is larger than or equal to 3.
It is possible to consider defining “Supremus typical sequences” on multivariate Markov chains [Fung et al.(2002)] or general jointly ergodic processes [Cover(1975)];

Related stochastic properties (AEP, (conditional) typicality lemmata and etc) regarding the above generalisations;

Define *Supremus Type* resembling other classic types [Csiszar(1998)], e.g. Markov type [Davisson et al.(1981)], in order to analyse the error exponents of related coding schemes (e.g. linear coding scheme over ring);
Thanks!
Bibliography I


