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

Irreducible Markov Chains

Definition 1

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

Example 2

$$\begin{bmatrix} .8142 & .1773 & .0042 & .0042 \\ .0042 & .9873 & .0042 & .0042 \\ .0042 & .1773 & .8142 & .0042 \\ .0042 & .1773 & .0042 & .8142 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ .0127 & .9873 & 0 & 0 \\ .0042 & .1773 & .8142 & .0042 \\ .0042 & .1773 & .0042 & .8142 \end{bmatrix}$$

The first stochastic matrix is irreducible, while the second one is not.

Strong Markov Property and Reduced Chains

Theorem 3 (Strong Markov Property [Norris(1998)])

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

Example 4

Given a Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ with state space $\mathscr{X}$ and a non-empty subset A of $\mathscr{X}$, 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, $\mathscr{M}_A = \left\{X^{(T_{A,l})}\right\}_{-\infty}^{\infty}$ is Markov.

Invariant Distributions

Theorem 5

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

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

Problem 6

In Theorem 5, how to compute π given $\mathbf{P}$ efficiently for very large sized $\mathbf{P}$?

Problem 7

In Example 4, if the $\mathscr{M}$ is finite-state and irreducible, will $\mathscr{M}_A$ be irreducible? If yes, what are the transition matrix and the invariant distribution of $\mathscr{M}_A$?

Stochastic Complements

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

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{A,A} & \mathbf{P}_{A,A^c} \\ \mathbf{P}_{A^c,A} & \mathbf{P}_{A^c,A^c} \end{bmatrix}, \text{ where } \mathbf{P}_{A,B} = [p_{i,j}]_{i \in A, j \in B} \text{ for } A, B \subseteq \mathscr{X},$$

is the transition matrix of $\mathscr{M}$, 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 $\mathbf{P}$ ($\mathscr{M}$) is irreducible, then

1. $\mathbf{S}_A$ is stochastic [Meyer(1989), Theorem 2.1];
2. $\mathbf{S}_A$ is irreducible [Meyer(1989), Theorem 2.3];
3. $\mathbf{S}_A$ is the transition matrix of $\mathscr{M}_A$ [Meyer(1989), Section 3];
4. Let $\pi = [p_i]_{i \in \mathscr{X}}$ be the invariant distribution of $\mathbf{P}$ ($\mathscr{M}$). The invariant distribution of $\mathbf{S}_A$ is

$$\left[\frac{p_i}{\sum_{j \in A} p_j} \right]_{i \in A} \quad \text{[Meyer(1989), Theorem 2.2].}$$

Remark 1

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

The Coupling Theorem

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

In Definition 8, if $\mathbf{P}$ $(\mathscr{M})$ is irreducible, let π_A and π_{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 π of $\mathbf{P}$ is given by

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

where

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

is the unique unitary vector such that

$$\overline{\xi}\mathbf{C} = \overline{\xi}, \mathbf{C} = \begin{bmatrix} \pi_A \mathbf{P}_{A,A}\mathbf{e} & \pi_A \mathbf{P}_{A,A^c}\mathbf{e} \\ \pi_{A^c}\mathbf{P}_{A^c,A}\mathbf{e} & \pi_{A^c}\mathbf{P}_{A^c,A^c}\mathbf{e} \end{bmatrix}. \tag{1}$$

$\mathbf{C}$ is named the *coupling matrix*, and ξ_A and ξ_{A^c} are termed the *coupling factors*.

Strongly Markov Typicality

Definition 11 (Strongly Markov Typicality)

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

$$\begin{cases} \left| \dfrac{N(i,j;\mathbf{x})}{N(i;\mathbf{x})} - p_{i,j} \right| < \epsilon; \\ \left| \dfrac{N(i;\mathbf{x})}{n} - p_i \right| < \epsilon, \end{cases} \quad \forall\, i,j \in \mathscr{X}, \tag{2}$$

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

Supremus Typicality

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

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

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

$$00102 - 02210 - 02112 - 01022 - 01102 \in \mathcal{S}_\epsilon(25, \mathbf{P}) \subseteq \mathcal{T}_\epsilon(25, \mathbf{P}),$$

$$\Longrightarrow \begin{cases} 0002 - 0220 - 022 - 0022 - 002 \in \mathcal{T}_\epsilon(18, \mathbf{S}_{\{0,2\}}); \\ 0010 - 010 - 011 - 010 - 0110 \in \mathcal{T}_\epsilon(17, \mathbf{S}_{\{0,1\}}). \end{cases}$$

AEP of Supremus Typical Sequences I

Notation:

1. Given a set $\mathscr{X}$, a *partition* $\coprod_{k \in \mathscr{K}} A_k$ of $\mathscr{X}$ is a disjoint union of $\mathscr{X}$, i.e. $A_{k'} \cap A_{k''} \neq \emptyset \Leftrightarrow k' = k''$, $\bigcup_{k \in \mathscr{K}} A_k = \mathscr{X}$ and A_k's are not empty.
2. For discrete random variables X and Y with sample spaces $\mathscr{X}$ and $\mathscr{Y}$, respectively, $X \sim [p_i]_{i \in \mathscr{X}}$ and $(X, Y) \sim [p_i]_{i \in \mathscr{X}} [p_{i,j}]_{i \in \mathscr{X}, j \in \mathscr{Y}}$ state for
$$\Pr\{X = i\} = p_i \text{ and } \Pr\{X = i, Y = j\} = p_i p_{i,j},$$
for all $i \in \mathscr{X}$ and $j \in \mathscr{Y}$, respectively.
3. Given a unitary vector π and a stochastic matrix $\mathbf{P}$ (not necessary that $\pi \mathbf{P} = \pi$), we write
$$H(\mathbf{P}|\pi) = H(Y|X),$$
for $(X, Y) \sim \pi \mathbf{P}$.

AEP of Supremus Typical Sequences II

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

Let $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ be an irreducible Markov chain with state space $\mathscr{X}$, and $\mathbf{P} = [p_{i,j}]_{i,j\in\mathscr{X}}$ and $\pi = [p_j]_{j\in\mathscr{X}}$ 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 $\forall\ \mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P})$,

1. $\exp_2\left[-n\left(H(\mathbf{P}|\pi) + \eta\right)\right] < \Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] = \mathbf{x}\right\} < \exp_2\left[-n\left(H(\mathbf{P}|\pi) - \eta\right)\right]$;
2. $\Pr\left\{\mathbf{X} \notin \mathcal{S}_\epsilon(n, \mathbf{P})\right\} < \eta$, where $\mathbf{X} = \left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right]$; and
3. $\left|\mathcal{S}_\epsilon(n, \mathbf{P})\right| < \exp_2\left[n\left(H(\mathbf{P}|\pi) + \eta\right)\right]$.

Typicality Lemmata of Supremus Typical Sequences I

Lemma 14 (Lemma D.1 of [Huang and Skoglund(2013)])

Given an irreducible Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ with finite state space $\mathscr{X}$, transition matrix $\mathbf{P}$ and invariant distribution $\pi = [p_j]_{j \in \mathscr{X}}$. Let $\coprod_{k=1}^{m} A_k$ be any partition of $\mathscr{X}$. 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 $\forall\ \mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P})$,

$$|S_\epsilon(\mathbf{x})| < \exp_2 \left\{ n \left[H \left(\operatorname{diag} \left\{ \{\mathbf{S}_k\}_{1 \le k \le m} \right\} \middle| \pi \right) + \eta \right] \right\}, \tag{3}$$

where

$$S_\epsilon(\mathbf{x}) = \left\{ \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P}) \middle| y^{(l)} \in A_k \Leftrightarrow x^{(l)} \in A_k, \right.$$
$$\left. \forall\ 1 \le l \le n, \forall\ 1 \le k \le m \right\},$$

and $\mathbf{S}_k$ is the stochastic complement of $\mathbf{P}_{A_k, A_k}$ in $\mathbf{P}$.

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

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

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

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

$$\mathscr{M} = \left\{ \begin{bmatrix} X^{(n)} \\ Y^{(n)} \end{bmatrix} \right\}_{-\infty}^{\infty}$$

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

Typicality Lemmata of Supremus Typical Sequences III

Remark 2

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

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

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

$$|S_\epsilon(\mathbf{x})| < \exp_2\left[n\left(H\left(X^{(1)}\right) - H\left(Y^{(1)}\right) + \eta\right)\right],$$

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 $\left\{Y^{(n)}\right\}_{-\infty}^{\infty}$.

Typicality Lemmata of Supremus Typical Sequences IV

Remark 3

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

$$\begin{aligned} |S_\epsilon(\mathbf{x})| &< \exp_2 \left\{ n \left[H\left(\mathbf{P}|\pi\right) - \lim_{w\to\infty} H\left(Y^{(w)} \middle| Y^{(w-1)} \right) + \eta \right] \right\} \\ &= \exp_2 \left\{ n \left[H\left(\mathbf{P}|\pi\right) - H\left(\mathbf{P}'|\pi'\right) + \eta \right] \right\}, \end{aligned} \tag{5}$$

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

Typicality Lemmata of Supremus Typical Sequences V

Remark 4

In Lemma 14 and Lemma 15, if

1. $\mathscr{X}$ is a finite ring $\mathfrak{R}$, then all costs $\mathfrak{R}/\mathfrak{I}$ over a left (right) ideal $\mathfrak{I}$ give raise to a partition of $\mathfrak{R} = \mathscr{X}$; and
2. $\mathscr{X}$ is a finite dimension vector space (module) over a finite field (ring) $\mathfrak{R}^n$ ($n \in \mathbb{N}^+$), then all affine subspaces (submodule) "parallel" to a given subspace (submodule) create a partition of $\mathfrak{R}^n = \mathscr{X}$; and
3. $\mathscr{X}$ is a finite group G, then all costs G/H over some subgroup H present a partition of $G = \mathscr{X}$; and etc.

From this, we see that the typicality lemmata can be tailored for specific situations.

Linear Coding over Ring for Irreducible Markov Sources

Theorem 16 ([Huang and Skoglund(2013)])

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

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

provided that

$$k > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{n}{\log|\mathfrak{I}|} \min\left\{H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), \; H\left(\mathbf{P}|\pi\right) - \lim_{w\to\infty}\frac{1}{w}H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(w)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(w-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right)\right\},$$

where $\mathbf{S}_{\mathfrak{R}/\mathfrak{I}} = \mathrm{diag}\left\{\{\mathbf{S}_A\}_{A\in\mathfrak{R}/\mathfrak{I}}\right\}$ with $\mathbf{S}_A$ being the stochastic complement of $\mathbf{P}_{A,A}$ in $\mathbf{P}$ and $Y_{\mathfrak{R}/\mathfrak{I}}^{(i)} = X^{(i)} + \mathfrak{I}$.

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

1. 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);
2. 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.

Generalisations

1. It is possible to consider defining "Supremus typical sequences" on multivariate Markov chains [Fung et al.(2002)] or general jointly ergodic processes [Cover(1975)];
2. Related stochastic properties (AEP, (conditional) typicality lemmata and etc) regarding the above generalisations;
3. 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

♫ ♬

Thanks!

Bibliography I

- J. R. Norris, *Markov Chains*. Cambridge University Press, Jul. 1998.
- L. Breuer and D. Baum, *An Introduction to Queueing Theory: and Matrix-Analytic Methods*, 2005th ed. Springer, Dec. 2005.
- C. D. Meyer, "Stochastic complementation, uncoupling markov chains, and the theory of nearly reducible systems," *SIAM Rev.*, vol. 31, no. 2, pp. 240–272, Jun. 1989. [Online]. Available: http://dx.doi.org/10.1137/1031050
- S. Huang and M. Skoglund, *Coding for Computing Irreducible Markovian Functions of Sources with Memory*, KTH Royal Institute of Technology, May 2013. [Online]. Available: http://www.ee.kth.se/~sheng11
- T. M. Cover, "A proof of the data compression theorem of slepian and wolf for ergodic sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 226–228, Mar. 1975.

Bibliography II

S. Huang and M. Skoglund, "On linear coding over finite rings and applications to computing," *IEEE Transactions on Information Theory*, Submitted. [Online]. Available: http://www.ee.kth.se/~sheng11

J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, Mar. 1979.

T. S. Han and K. Kobayashi, "A dichotomy of functions f(x, y) of correlated sources (x, y) from the viewpoint of the achievable rate region," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 69–76, Jan. 1987.

E. Fung, W. K. Ching, S. Chu, M. Ng, and W. Zang, "Multivariate markov chain models," in *2002 IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, Oct. 2002, p. 5 pp. vol.3.

Bibliography III

- I. Csiszar, "The method of types [information theory]," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.
- L. D. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic markov sources," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 431–438, Jul. 1981.