Supremus Typicality

Sheng Huang and Mikael Skoglund

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

July 3, 2014
Honolulu, USA

Outline

An Observation from the Classic Typicality (I)

Classic Asymptotically Equipartition Property (AEP): Given a randomly generated (w.r.t. some stationary ergodic process of sample space $\mathscr{X}$, say $\{0, 1, 2\}$) sequence X^n, say

00102 - 02210 - 02112 - 01022 - 01102,

in probability ϵ close to 1 that X^n is classic ϵ-typical (in the strong or weak sense) for large enough n.

An Observation from the Classic Typicality (I)

Classic Asymptotically Equipartition Property (AEP): Given a randomly generated (w.r.t. some stationary ergodic process of sample space $\mathscr{X}$, say $\{0, 1, 2\}$) sequence X^n, say

$$00102 \text{ - } 02210 \text{ - } 02112 \text{ - } 01022 \text{ - } 01102,$$

in probability ϵ close to 1 that X^n is classic ϵ-typical (in the strong or weak sense) for large enough n.

Let $\emptyset \neq A \subseteq \mathscr{X}$. Define $Y_A^{(l)} = X^{(T_{A,l})}$ where

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

Let $k = \max\{j | X^{(j)} \in A\}$. The property that the reduced sequence Y_A^k of X^n, say

$$0010 \text{ - } 010 \text{ - } 011 \text{ - } 010 \text{ - } 0110 \text{ when } A = \{0, 1\},$$

is also typical in probability close to 1 is important. Can one make such a claim in general for all non-empty subset A of $\mathscr{X}$?

An Observation from the Classic Typicality (II)

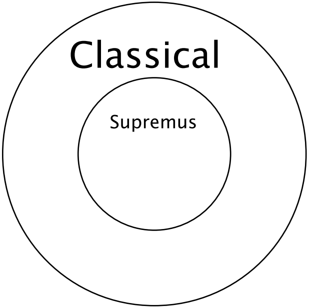
Let $\{\alpha, \beta, \gamma\}$ be the state space of the Markov (i.i.d.) process with transition matrix

$$\mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}. \tag{1}$$

For

$$\mathbf{x} = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma), \tag{2}$$

it is easy to verify that $\mathbf{x}$ is a strongly Markov 5/12-typical sequence.



An Observation from the Classic Typicality (II)

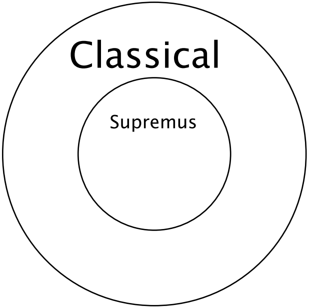
Let $\{\alpha, \beta, \gamma\}$ be the state space of the Markov (i.i.d.) process with transition matrix

$$\mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}. \qquad (1)$$

For

$$\mathbf{x} = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma), \qquad (2)$$

it is easy to verify that $\mathbf{x}$ is a strongly Markov 5/12-typical sequence.



However, the subsequence

$$\mathbf{x}_{\{\alpha,\gamma\}} = (\alpha, \gamma, \alpha, \gamma, \alpha, \gamma) \qquad (3)$$

is no long a strongly Markov 5/12-typical sequence, because the stochastic complement [Mey89] $\mathbf{S}_{\{\alpha,\gamma\}} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ and

$$\left| \frac{\text{the number of subsequence } (\alpha, \alpha)\text{'s in } \mathbf{x}_{\{\alpha,\gamma\}}}{6} - 0.5 \right| = |0 - 0.5| > \frac{5}{12}. \qquad (4)$$

The Technical Issues

1. Is $\{Y_A^{(l)}\}$ still an random process in general?

The Technical Issues

1. Is $\{Y_A^{(l)}\}$ still an random process in general?
2. How is (the stochastic properties of) $\{Y_A^{(l)}\}$ described mathematically that allows the analysis to be carried out?

The Technical Issues

1. Is $\{Y_A^{(l)}\}$ still an random process in general?
2. How is (the stochastic properties of) $\{Y_A^{(l)}\}$ described mathematically that allows the analysis to be carried out?
3. Is there an ergodic theorem associated with $\{Y_A^{(l)}\}$?

Supremus Typicality in the Strong Sense

1. If $\{X^{(j)}\}$ is Markov, then $\{Y_A^{(l)}\}$ is also Markov by the strong Markov property [Nor98, Theorem 1.4.2].

Supremus Typicality in the Strong Sense

1. If $\{X^{(j)}\}$ is Markov, then $\{Y_A^{(l)}\}$ is also Markov by the strong Markov property [Nor98, Theorem 1.4.2].
2. If $\{X^{(j)}\}$ is irreducible Markov with transition matrix $\mathbf{P}$, then $\{Y_A^{(l)}\}$ is also irreducible Markov. Moreover, the stochastic complement

$$\mathbf{S}_A = \mathbf{P}_{A,A} + \mathbf{P}_{A,A^c}\left(\mathbf{1} - \mathbf{P}_{A^c,A^c}\right)^{-1}\mathbf{P}_{A^c,A}$$

of $\mathbf{P} = \begin{bmatrix} \mathbf{P}_{A,A} & \mathbf{P}_{A,A^c} \\ \mathbf{P}_{A^c,A} & \mathbf{P}_{A^c,A^c} \end{bmatrix}$ is the transition matrix of $\{Y_A^{(l)}\}$ [Mey89].

Supremus Typicality in the Strong Sense

1. If $\{X^{(j)}\}$ is Markov, then $\{Y_A^{(l)}\}$ is also Markov by the strong Markov property [Nor98, Theorem 1.4.2].
2. If $\{X^{(j)}\}$ is irreducible Markov with transition matrix $\mathbf{P}$, then $\{Y_A^{(l)}\}$ is also irreducible Markov. Moreover, the stochastic complement

$$\mathbf{S}_A = \mathbf{P}_{A,A} + \mathbf{P}_{A,A^c}\left(\mathbf{1} - \mathbf{P}_{A^c,A^c}\right)^{-1}\mathbf{P}_{A^c,A}$$

of $\mathbf{P} = \begin{bmatrix} \mathbf{P}_{A,A} & \mathbf{P}_{A,A^c} \\ \mathbf{P}_{A^c,A} & \mathbf{P}_{A^c,A^c} \end{bmatrix}$ is the transition matrix of $\{Y_A^{(l)}\}$ [Mey89].

3. Ergodic theorem of irreducible Markov chain [Nor98, Theorem 1.10.2].

Supremus Typicality in the Strong Sense

1. If $\{X^{(j)}\}$ is Markov, then $\{Y_A^{(l)}\}$ is also Markov by the strong Markov property [Nor98, Theorem 1.4.2].
2. If $\{X^{(j)}\}$ is irreducible Markov with transition matrix $\mathbf{P}$, then $\{Y_A^{(l)}\}$ is also irreducible Markov. Moreover, the stochastic complement

$$\mathbf{S}_A = \mathbf{P}_{A,A} + \mathbf{P}_{A,A^c}\left(\mathbf{1} - \mathbf{P}_{A^c,A^c}\right)^{-1}\mathbf{P}_{A^c,A}$$

of $\mathbf{P} = \begin{bmatrix} \mathbf{P}_{A,A} & \mathbf{P}_{A,A^c} \\ \mathbf{P}_{A^c,A} & \mathbf{P}_{A^c,A^c} \end{bmatrix}$ is the transition matrix of $\{Y_A^{(l)}\}$ [Mey89].

3. Ergodic theorem of irreducible Markov chain [Nor98, Theorem 1.10.2].

AEP of Supremus Typical in the Strong Sense: In probability ϵ close to 1, all reduced sequences of a randomly generated sequence of an irreducible Markov chain is Supremus ϵ-typical for large enough n.

Supremus Typicality in the Weak Sense: Backgrounds

1. probability space $(\Omega, \mathscr{F}, \mu)$.
2. measurable transformation $T : \Omega \to \Omega$ (not necessarily probability preserving).
3. dynamical system $(\Omega, \mathscr{F}, \mu, T)$.
4. Let $X : \Omega \to \mathscr{X}$ ($\mathscr{X}$ is always assumed to be finite from now on) be a measurable function. $\{X^{(j)}\} = \{X(T^j)\}$ defines a random process with state space $\mathscr{X}$.
5. $(\Omega, \mathscr{F}, \mu, T)$ is said to be asymptotically mean stationary (a.m.s.) [1] [GK80] if there exists a measure $\overline{\mu}$ on $(\Omega, \mathscr{F})$ satisfying
$$\overline{\mu}(B) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m} \mu(T^{-i}B), \forall\, B \in \mathscr{F}.$$
6. $(\Omega, \mathscr{F}, \mu, T)$ is said to be ergodic if $T^{-1}B = B \implies \mu(B) = 0$ or $\mu(B) = 1$.
7. The random process $\{X^{(j)}\} = \{X(T^j)\}$ is said to be a.m.s. (ergodic) if $(\Omega, \mathscr{F}, \mu, T)$ is a.m.s. (ergodic).

[1] The a.m.s. condition is interesting because it is a sufficient and necessary condition for the Point-wise Ergodic Theorem to hold [GK80, Theorem 1].

Supremus Typicality in the Weak Sense: Induced Transformations

Given a recurrent system $(\Omega, \mathscr{F}, \mu, T)$ and $\mathcal{A} \in \mathscr{F}$ $(\mu(\mathcal{A}) > 0)$, one can define the induced transformation [Kak43] $T_{\mathcal{A}}$ on $(\mathcal{A}_0, \mathscr{A}, \mu|_{\mathscr{A}})$, where $\mathcal{A}_0 = \mathcal{A} \cap \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j}\mathcal{A}$ and $\mathscr{A} = \{\mathcal{A}_0 \cap \mathcal{B} | \mathcal{B} \in \mathscr{F}\}$, such that

$$T_{\mathcal{A}}(x) = T^{\psi_{\mathcal{A}}^{(1)}(x)}(x), \forall\, x \in \mathcal{A}_0,$$

where

$$\psi_{\mathcal{A}}^{(1)}(x) = \min\left\{ i \in \mathbb{N}^+ \middle| T^i(x) \in \mathcal{A}_0 \right\}$$

is the first return time function.

Supremus Typicality in the Weak Sense: Induced Transformations

Given a recurrent system $(\Omega, \mathscr{F}, \mu, T)$ and $\mathcal{A} \in \mathscr{F}$ $(\mu(\mathcal{A}) > 0)$, one can define the induced transformation [Kak43] $T_{\mathcal{A}}$ on $(\mathcal{A}_0, \mathscr{A}, \mu|_{\mathscr{A}})$, where $\mathcal{A}_0 = \mathcal{A} \cap \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j}\mathcal{A}$ and $\mathscr{A} = \{\mathcal{A}_0 \cap \mathcal{B} | \mathcal{B} \in \mathscr{F}\}$, such that

$$T_{\mathcal{A}}(x) = T^{\psi_{\mathcal{A}}^{(1)}(x)}(x), \forall\, x \in \mathcal{A}_0,$$

where

$$\psi_{\mathcal{A}}^{(1)}(x) = \min\left\{i \in \mathbb{N}^+ \,|\, T^i(x) \in \mathcal{A}_0\right\}$$

is the first return time function.
One can verify that $(\mathcal{A}_0, \mathscr{A}, \frac{1}{\mu(\mathcal{A}_0)}\mu|_{\mathscr{A}}, T_{\mathcal{A}})$ forms a new dynamical system. Moreover ...

Supremus Typicality in the Weak Sense

1. If $\{X^{(j)}\} = \{X(T^j)\}$ is recurrent, then $\{Y_A^{(l)}\} = \left\{X\left(T_{\mathcal{A}}^l\right)\right\}$ with $\mathcal{A} = X^{-1}(A)$, and $\{Y_A^{(l)}\}$ is recurrent.

Supremus Typicality in the Weak Sense

1. If $\{X^{(j)}\} = \{X(T^j)\}$ is recurrent, then $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ with $\mathcal{A} = X^{-1}(A)$, and $\{Y_A^{(l)}\}$ is recurrent.
2. If $\{X^{(j)}\} = \{X(T^j)\}$ is ergodic, so is $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ [Aar97, Proposition 1.5.2].

Supremus Typicality in the Weak Sense

1. If $\{X^{(j)}\} = \{X(T^j)\}$ is recurrent, then $\{Y_A^{(l)}\} = \{X\left(T_{\mathcal{A}}^l\right)\}$ with $\mathcal{A} = X^{-1}(A)$, and $\{Y_A^{(l)}\}$ is recurrent.
2. If $\{X^{(j)}\} = \{X(T^j)\}$ is ergodic, so is $\{Y_A^{(l)}\} = \{X\left(T_{\mathcal{A}}^l\right)\}$ [Aar97, Proposition 1.5.2].
3. If $\{X^{(j)}\} = \{X(T^j)\}$ is a.m.s., so is $\{Y_A^{(l)}\} = \{X\left(T_{\mathcal{A}}^l\right)\}$ [HS14].

Supremus Typicality in the Weak Sense

1. If $\{X^{(j)}\} = \{X(T^j)\}$ is recurrent, then $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ with $\mathcal{A} = X^{-1}(A)$, and $\{Y_A^{(l)}\}$ is recurrent.
2. If $\{X^{(j)}\} = \{X(T^j)\}$ is ergodic, so is $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ [Aar97, Proposition 1.5.2].
3. If $\{X^{(j)}\} = \{X(T^j)\}$ is a.m.s., so is $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ [HS14].
4. If $\{X^{(j)}\} = \{X(T^j)\}$ is a.m.s. and ergodic, then the Shannon–McMillan–Breiman (SMB) Theorem holds. In exact terms,

$$-\frac{1}{n}\log p(X^{(0)}, X^{(1)}, \cdots, X^{(n-1)}) \to H \text{ with probability } 1,$$

where H is the entropy rate of $\{X^{(l)}\}$ [GK80].

Supremus Typicality in the Weak Sense

1. If $\{X^{(j)}\} = \{X(T^j)\}$ is recurrent, then $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ with $\mathcal{A} = X^{-1}(A)$, and $\{Y_A^{(l)}\}$ is recurrent.
2. If $\{X^{(j)}\} = \{X(T^j)\}$ is ergodic, so is $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ [Aar97, Proposition 1.5.2].
3. If $\{X^{(j)}\} = \{X(T^j)\}$ is a.m.s., so is $\{Y_A^{(l)}\} = \{X(T_{\mathcal{A}}^l)\}$ [HS14].
4. If $\{X^{(j)}\} = \{X(T^j)\}$ is a.m.s. and ergodic, then the Shannon–McMillan–Breiman (SMB) Theorem holds. In exact terms,

$$-\frac{1}{n}\log p(X^{(0)}, X^{(1)}, \cdots, X^{(n-1)}) \to H \text{ with probability } 1,$$

where H is the entropy rate of $\{X^{(l)}\}$ [GK80].

AEP of Supremus Typical in the Weak Sense: In probability ϵ close to 1, all reduced sequences of a randomly generated sequence of a recurrent a.m.s. ergodic process is Supremus ϵ-typical for large enough n.

Where is it used

For i.i.d. sources,

SW scheme [SW73]	field linear scheme [Eli55, Csi82]	non-field ring linear scheme [HS12]
	dominate SW for encoding binary sum $x \oplus y$ over $\mathbb{Z}_2$ [KM79]	dominate field linear for encoding some functions e.g. $x + 2y + 3z$ over $\mathbb{Z}_4$ [HS12]

Where is it used

For i.i.d. sources,

SW scheme [SW73]	field linear scheme [Eli55, Csi82]	non-field ring linear scheme [HS12]
	dominate SW for encoding binary sum $x \oplus y$ over $\mathbb{Z}_2$ [KM79]	dominate field linear for encoding some functions e.g. $x + 2y + 3z$ over $\mathbb{Z}_4$ [HS12]

To generalized results in [HS12] to the non-i.i.d. scenarios, the argument based on classic typicality does not lead to a conclusion that is accessible and easy to analyse.

non-field ring linear scheme for i.i.d. sources [HS12]	non-field ring linear scheme for irreducible Markov sources [HS13]
classic typicality	Supremus typicality

What Makes the Difference

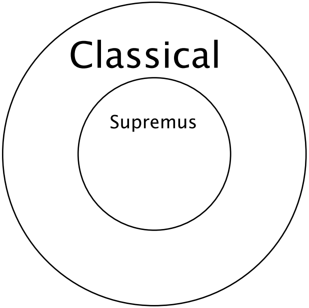
Let $\{\alpha, \beta, \gamma\}$ be the state space of the Markov (i.i.d.) process with transition matrix

$$\mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}. \tag{5}$$

For

$$\mathbf{x} = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma), \tag{6}$$

it is easy to verify that $\mathbf{x}$ is a strongly Markov 5/12-typical sequence.



What Makes the Difference

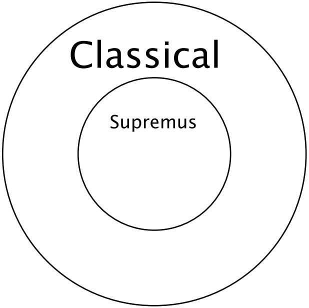
Let $\{\alpha, \beta, \gamma\}$ be the state space of the Markov (i.i.d.) process with transition matrix

$$\mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}. \quad (5)$$

For

$$\mathbf{x} = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma), \quad (6)$$

it is easy to verify that $\mathbf{x}$ is a strongly Markov 5/12-typical sequence.



However, the subsequence

$$\mathbf{x}_{\{\alpha,\gamma\}} = (\alpha, \gamma, \alpha, \gamma, \alpha, \gamma) \quad (7)$$

is no long a strongly Markov 5/12-typical sequence, because the stochastic complement [Mey89] $\mathbf{S}_{\{\alpha,\gamma\}} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ and

$$\left| \frac{\text{the number of subsequence } (\alpha, \alpha)\text{'s in } \mathbf{x}_{\{\alpha,\gamma\}}}{6} - 0.5 \right| = |0 - 0.5| > \frac{5}{12}. \quad (8)$$

Conclusion

1. Supremus typicality refines Shannon's idea on typicality [SW49];

Conclusion

1. Supremus typicality refines Shannon's idea on typicality [SW49];
2. Apart from the ergodic theorem, it takes the self-iterating properties of the random process into account;

Conclusion

1. Supremus typicality refines Shannon's idea on typicality [SW49];
2. Apart from the ergodic theorem, it takes the self-iterating properties of the random process into account;
3. Self-iterating properties:
    1. reduced chains of an irreducible Markov chain are irreducible Markov [Mey89];
    2. reduced processes of a recurrent (a.m.s. and ergodic, resp.) random process are also recurrent (a.m.s. and ergodic, resp.) [HS14, Aar97].

Thanks

♫ ♬

Thanks!

Bibliography I

Jon Aaronson.
An Introduction to Infinite Ergodic Theory.
American Mathematical Society, Providence, R.I., 1997.

Imre Csiszár.
Linear codes for sources and source networks: Error exponents, universal coding.
IEEE Transactions on Information Theory, 28(4):585–592, July 1982.

P. Elias.
Coding for noisy channels.
IRE Convention Record, 3:37–46, March 1955.

Robert M. Gray and J. C. Kieffer.
Asymptotically mean stationary measures.
The Annals of Probability, 8(5):962–973, October 1980.

Bibliography II

Sheng Huang and Mikael Skoglund.
On Linear Coding over Finite Rings and Applications to Computing.
KTH Royal Institute of Technology, October 2012.
http://people.kth.se/~sheng11

Sheng Huang and Mikael Skoglund.
Encoding Irreducible Markovian Functions of Sources: An Application of Supremus Typicality.
KTH Royal Institute of Technology, May 2013.
http://people.kth.se/~sheng11

Sheng Huang and Mikael Skoglund.
Induced transformations of recurrent a.m.s. dynamical systems.
Stochastics and Dynamics, 2014.

Shizuo Kakutani.
Induced measure preserving transformations.
Proceedings of the Imperial Academy, 19(10):635–641, 1943.

Bibliography III

János Körner and Katalin Marton.
How to encode the modulo-two sum of binary sources.
IEEE Transactions on Information Theory, 25(2):219–221, March 1979.

Carl D. Meyer.
Stochastic complementation, uncoupling markov chains, and the theory of nearly reducible systems.
SIAM Rev., 31(2):240–272, June 1989.

James R. Norris.
Markov Chains.
Cambridge University Press, July 1998.

Claude Elwood Shannon and Warren Weaver.
The mathematical theory of communication.
University of Illinois Press, Urbana, 1949.

Bibliography IV

David Slepian and Jack K. Wolf.
Noiseless coding of correlated information sources.
IEEE Transactions on Information Theory, 19(4):471–480, July 1973.