Supremus Typicality

Sheng Huang and Mikael Skoglund

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

July 3, 2014
Honolulu, USA
Outline

1 What is Supremus Typicality
   - An Observation from the Classic Typicality
   - The Technical Issues
   - Supremus Typicality in the Strong Sense
   - Supremus Typicality in the Weak Sense

2 Where Is It Used

3 What Makes the Difference

4 Conclusion

5 Thanks / References
   - Thanks
   - Bibliography
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 $\mathcal{X}$, say $\{0, 1, 2\}$) sequence $X^n$, say

$$00102 - 02210 - 02112 - 01022 - 01102,$$

in probability $\epsilon$ close to 1 that $X^n$ is classic $\epsilon$-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 \( \mathcal{X} \), say \( \{0, 1, 2\} \)) sequence \( X^n \), say

\[
00102 - 02210 - 02112 - 01022 - 01102,
\]

in probability \( \epsilon \) close to 1 that \( X^n \) is classic \( \epsilon \)-typical (in the strong or weak sense) for large enough \( n \). Let \( \emptyset \neq A \subseteq \mathcal{X} \). Define \( Y_A^{(l)} = X(T_{A,l}) \) where

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

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

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

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

\[
P = \begin{bmatrix}
1/3 & 1/3 & 1/3 \\
1/3 & 1/3 & 1/3 \\
1/3 & 1/3 & 1/3 \\
\end{bmatrix}.
\] (1)

For

\[
x = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma),
\] (2)

it is easy to verify that \( x \) 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

\[
P = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{bmatrix}.
\] (1)

For

\[
x = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma),
\] (2)

it is easy to verify that \( x \) is a strongly Markov 5/12-typical sequence.

However, the subsequence

\[
x_{\{\alpha, \gamma\}} = (\alpha, \gamma, \alpha, \gamma, \alpha, \gamma)
\] (3)

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

\[
\text{the number of subsequence} \ (\alpha, \alpha)'s \text{ in } x_{\{\alpha, \gamma\}} - 0.5 = \left| 0 - 0.5 \right| > \frac{5}{12}. \] (4)
The Technical Issues

Is \( \{ Y_A^{(l)} \} \) still a 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 a 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 \( \{ Y_A^{(l)} \} \)?
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

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

- If \( \{X^{(j)}\} \) is irreducible Markov with transition matrix \( P \), then \( \{Y_A^{(l)}\} \) is also irreducible Markov. Moreover, the stochastic complement

\[
S_A = P_{A,A} + P_{A,A^c} \left( I - P_{A^c,A^c}\right)^{-1} P_{A^c,A}
\]

of \( P = \begin{bmatrix} P_{A,A} & P_{A,A^c} \\ P_{A^c,A} & P_{A^c,A^c} \end{bmatrix} \) is the transition matrix of \( \{Y_A^{(l)}\} \) [Mey89].
Supremus Typicality in the Strong Sense

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

2. If \( \{X^{(j)}\} \) is irreducible Markov with transition matrix \( P \), then \( \{Y^{(l)}_A\} \) is also irreducible Markov. Moreover, the stochastic complement

   \[ S_A = P_{A,A} + P_{A,A^c} (1 - P_{A^c,A^c})^{-1} P_{A^c,A} \]

   of \( P = \begin{bmatrix} P_{A,A} & P_{A,A^c} \\ P_{A^c,A} & P_{A^c,A^c} \end{bmatrix} \) is the transition matrix of \( \{Y^{(l)}_A\} \) [Mey89].

3. Ergodic theorem of irreducible Markov chain [Nor98, Theorem 1.10.2].
Supremus Typicality in the Strong Sense

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

- If \( \{X^{(j)}\} \) is irreducible Markov with transition matrix \( P \), then \( \{Y_A^{(l)}\} \) is also irreducible Markov. Moreover, the stochastic complement \( S_A = P_{A,A} + P_{A,A^c} (1 - P_{A^c,A^c})^{-1} P_{A^c,A} \) of \( P = \begin{bmatrix} P_{A,A} & P_{A,A^c} \\ P_{A^c,A} & P_{A^c,A^c} \end{bmatrix} \) is the transition matrix of \( \{Y_A^{(l)}\} \) [Mey89].

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

**AEP of Supremus Typical in the Strong Sense:** In probability \( \epsilon \) close to 1, all reduced sequences of a randomly generated sequence of an irreducible Markov chain is Supremus \( \epsilon \)-typical for large enough \( n \).
Supremus Typicality in the Weak Sense: Backgrounds

1. Probability space $\left( \Omega, \mathcal{F}, \mu \right)$.
2. Measurable transformation $T : \Omega \rightarrow \Omega$ (not necessarily probability preserving).
3. Dynamical system $\left( \Omega, \mathcal{F}, \mu, T \right)$.
4. Let $X : \Omega \rightarrow \mathcal{X}$ ($\mathcal{X}$ is always assumed to be finite from now on) be a measurable function. $\{X(i)\} = \{X(T^i)\}$ defines a random process with state space $\mathcal{X}$.
5. $(\Omega, \mathcal{F}, \mu, T)$ is said to be asymptotically mean stationary (a.m.s.) if there exists a measure $\overline{\mu}$ on $(\Omega, \mathcal{F})$ satisfying

$$\overline{\mu}(B) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=0}^{m} \mu(T^{-i} B), \forall B \in \mathcal{F}.$$  

6. $(\Omega, \mathcal{F}, \mu, T)$ is said to be ergodic if $T^{-1} B = B \implies \mu(B) = 0$ or $\mu(B) = 1$.

7. The random process $\{X(i)\} = \{X(T^i)\}$ is said to be a.m.s. (ergodic) if $(\Omega, \mathcal{F}, \mu, T)$ 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 \((\Omega, \mathcal{F}, \mu, T)\) and \(A \in \mathcal{F} \ (\mu(A) > 0)\), one can define the induced transformation \([\text{Kak43}]\) \(T_A\) on \((A_0, \mathcal{A}, \mu|_\mathcal{A})\), where

\[
A_0 = A \cap \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j}A \quad \text{and} \quad \mathcal{A} = \{A_0 \cap B | B \in \mathcal{F}\},
\]

such that

\[
T_A(x) = T^{\psi_A^{(1)}(x)}(x), \quad \forall \ x \in A_0,
\]

where

\[
\psi_A^{(1)}(x) = \min \left\{ i \in \mathbb{N}^+ | T^i(x) \in A_0 \right\}
\]

is the first return time function.
Supremus Typicality in the Weak Sense: Induced Transformations

Given a recurrent system \((\Omega, \mathcal{F}, \mu, T)\) and \(A \in \mathcal{F}\) \((\mu(A) > 0)\), one can define the induced transformation \([\text{Kak43}]\) \(T_A\) on \((A_0, \mathcal{A}, \mu|_\mathcal{A})\), where

\[
A_0 = A \cap \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j} A \quad \text{and} \quad \mathcal{A} = \{A_0 \cap B | B \in \mathcal{F}\},
\]

such that

\[
T_A(x) = T^{\psi_A^{(1)}(x)}(x), \quad \forall \ x \in A_0,
\]

where

\[
\psi_A^{(1)}(x) = \min \{i \in \mathbb{N}^+ | T^i(x) \in A_0\}
\]

is the first return time function.

One can verify that \((A_0, \mathcal{A}, \frac{1}{\mu(A_0)}\mu|_\mathcal{A}, T_A)\) forms a new dynamical system. Moreover ...
Supremus Typicality in the Weak Sense

If $\{X^{(j)}\} = \{X(T^j)\}$ is recurrent, then $\{Y^{(l)}_A\} = \{X(T^l_A)\}$ with $A = X^{-1}(A)$, and $\{Y^{(l)}_A\}$ is recurrent.
Supremus Typicality in the Weak Sense

1. If \( \{X(j)\} = \{X(T^j)\} \) is recurrent, then \( \{Y_{\mathcal{A}}(l)\} = \{X(T^l)\} \) with \( \mathcal{A} = X^{-1}(A) \), and \( \{Y_{\mathcal{A}}(l)\} \) is recurrent.

2. If \( \{X(j)\} = \{X(T^j)\} \) is ergodic, so is \( \{Y_{\mathcal{A}}(l)\} = \{X(T^l)\} \) [Aar97, Proposition 1.5.2].
Supremus Typicality in the Weak Sense

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

If \( \{X(j)\} = \{X(T^j)\} \) is ergodic, so is \( \{Y_A(l)\} = \{X(T_A^l)\} \) [Aar97, Proposition 1.5.2].

If \( \{X(j)\} = \{X(T^j)\} \) is a.m.s., so is \( \{Y_A(l)\} = \{X(T_A^l)\} \) [HS14].

If \( \{X(j)\} = \{X(T^j)\} \) is a.m.s. and ergodic, then the Shannon–McMillan–Breiman (SMB) Theorem holds. In exact terms,

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

where \( H \) is the entropy rate of \( \{X(l)\} \) [GK80].
Supremus Typicality in the Weak Sense

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

\[
\frac{1}{n} \log p(X^{(0)}, X^{(1)}, \ldots, X^{(n-1)}) \to H \quad \text{with probability 1},
\]

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

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

For i.i.d. sources,

<table>
<thead>
<tr>
<th>SW scheme [SW73]</th>
<th>field linear scheme [Eli55, Csi82]</th>
<th>non-field ring linear scheme [HS12]</th>
</tr>
</thead>
<tbody>
<tr>
<td>dominate SW for encoding binary sum $x \oplus y$ over $\mathbb{Z}_2$ [KM79]</td>
<td>dominate field linear for encoding some functions e.g. $x + 2y + 3z$ over $\mathbb{Z}_4$ [HS12]</td>
<td></td>
</tr>
</tbody>
</table>
## Where is it used

For i.i.d. sources,

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

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.

<table>
<thead>
<tr>
<th>non-field ring linear scheme for i.i.d. sources [HS12]</th>
<th>non-field ring linear scheme for irreducible Markov sources [HS13]</th>
</tr>
</thead>
<tbody>
<tr>
<td>classic typicality</td>
<td>Supremus typicality</td>
</tr>
</tbody>
</table>
What Makes the Difference

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

\[
P = \begin{bmatrix}
  \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
  \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
  \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{bmatrix}.
\] (5)

For

\[
x = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma),
\] (6)

it is easy to verify that \( x \) is a strongly Markov 5/12-typical sequence.
Let \( \{\alpha, \beta, \gamma\} \) be the state space of the Markov (i.i.d.) process with transition matrix

\[
P = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{bmatrix}.
\tag{5}
\]

For

\[
x = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma),
\tag{6}
\]

it is easy to verify that \( x \) is a strongly Markov 5/12-typical sequence.

However, the subsequence

\[
x_{\{\alpha, \gamma\}} = (\alpha, \gamma, \alpha, \gamma, \alpha, \gamma)
\tag{7}
\]

is no longer a strongly Markov 5/12-typical sequence, because the stochastic complement [Mey89] \( S_{\{\alpha, \gamma\}} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \)

and

\[
\left| \frac{\text{the number of subsequence } (\alpha, \alpha)'s \text{ in } x_{\{\alpha, \gamma\}}}{6} - 0.5 \right| = |0 - 0.5| > \frac{5}{12}. \tag{8}
\]
Supremus typicality refines Shannon’s idea on typicality [SW49];
Supremus typicality refines Shannon’s idea on typicality [SW49];

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

Apart from the ergodic theorem, it takes the self-iterating properties of the random process into account;

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!
Jon Aaronson.
*An Introduction to Infinite Ergodic Theory.*

Imre Csiszár.
Linear codes for sources and source networks: Error exponents, universal coding.

P. Elias.
Coding for noisy channels.

Robert M. Gray and J. C. Kieffer.
Asymptotically mean stationary measures.
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. 
János Körner and Katalin Marton.
How to encode the modulo-two sum of binary sources.

Carl D. Meyer.
Stochastic complementation, uncoupling markov chains, and the theory of nearly reducible systems.

James R. Norris.
*Markov Chains*.

Claude Elwood Shannon and Warren Weaver.
*The mathematical theory of communication*.
David Slepian and Jack K. Wolf.
Noiseless coding of correlated information sources.