The Asymptotical Equipartition Property of Supremus Typicality in the Weak Sense

Sheng Huang and Mikael Skoglund

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

March 13, 2014
Internal Seminar

Outline

A Starting Example

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.

A Starting Example

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. However, the subsequence

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

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

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

A.M.S. Random Processes (I)

Given a probability space $(\Omega, \mathscr{F}, \mu)$ and a measurable transformation $T : \Omega \to \Omega$ (not necessarily probability preserving), the tuple $(\Omega, \mathscr{F}, \mu, T)$ is called a *dynamical system* (or *ergodic system*).

A.M.S. Random Processes (I)

Given a probability space $(\Omega, \mathscr{F}, \mu)$ and a measurable transformation $T : \Omega \to \Omega$ (not necessarily probability preserving), the tuple $(\Omega, \mathscr{F}, \mu, T)$ is called a *dynamical system* (or *ergodic system*).
Let $X : \Omega \to \mathscr{X}$ (e.g. $\mathscr{X}$ is a finite set) be a measurable function. Then

$$\{X^{(n)}\} = \{X(T^n)\} \tag{5}$$

defines a *random process* with state space $\mathscr{X}$ and pdf/pmf

$$p\left(x^{(0)}, x^{(1)}, \cdots, x^{(n-1)}\right) = \mu\left(\bigcap_{i=0}^{n-1} T^{-i}\left(X^{-1}\left(x^{(i)}\right)\right)\right). \tag{6}$$

A.M.S. Random Processes (I)

Given a probability space $(\Omega, \mathscr{F}, \mu)$ and a measurable transformation $T : \Omega \to \Omega$ (not necessarily probability preserving), the tuple $(\Omega, \mathscr{F}, \mu, T)$ is called a *dynamical system* (or *ergodic system*).
Let $X : \Omega \to \mathscr{X}$ (e.g. $\mathscr{X}$ is a finite set) be a measurable function. Then

$$\{X^{(n)}\} = \{X(T^n)\} \tag{5}$$

defines a *random process* with state space $\mathscr{X}$ and pdf/pmf

$$p\left(x^{(0)}, x^{(1)}, \cdots, x^{(n-1)}\right) = \mu\left(\bigcap_{i=0}^{n-1} T^{-i}\left(X^{-1}\left(x^{(i)}\right)\right)\right). \tag{6}$$

Example 1

Given a random process $\{X^{(n)}\}$ with sample space. Let $\Omega = \prod_{i=-\infty}^{\infty} \mathscr{X}$, T be a time shift and X be the coordinate function

$$X : (\cdots, x^{(-1)}, x^{(0)}, x^{(1)}, \cdots) \mapsto x^{(0)}. \tag{7}$$

Define μ satisfying $\mu\left(\bigcap_{i=0}^{n-1} T^{-i}\left(X^{-1}\left(x^{(i)}\right)\right)\right) = p\left(x^{(0)}, x^{(1)}, \cdots, x^{(n-1)}\right)$.
By the Kolmogorov Extension Theorem, $\{X^{(n)}\} = \{X(T^n)\}$.

A.M.S. Random Processes (II)

$(\Omega, \mathscr{F}, \mu, T)$ is said to be *asymptotically mean stationary* (*a.m.s.*) [1] [GK80] if there exists a measure $\overline{\mu}$ on $(\Omega, \mathscr{F})$ satisfying

$$\overline{\mu}(B) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m} \mu(T^{-i}B), \forall\, B \in \mathscr{F}. \tag{8}$$

[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].

A.M.S. Random Processes (II)

$(\Omega, \mathscr{F}, \mu, T)$ is said to be *asymptotically mean stationary* (*a.m.s.*) [1] [GK80] if there exists a measure $\overline{\mu}$ on $(\Omega, \mathscr{F})$ satisfying

$$\overline{\mu}(B) = \lim_{m \to \infty} \frac{1}{m} \sum_{i=0}^{m} \mu(T^{-i}B), \forall\, B \in \mathscr{F}. \tag{8}$$

Obviously, if $(\Omega, \mathscr{F}, \mu, T)$ is stationary, i.e. $\mu(B) = \mu(T^{-1}B)$, then it is a.m.s.. In addition, $(\Omega, \mathscr{F}, \mu, T)$ is said to be ergodic if

$$T^{-1}B = B \implies \mu(B) = 0 \text{ or } \mu(B) = 1. \tag{9}$$

[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].

A.M.S. Random Processes (II)

$(\Omega, \mathscr{F}, \mu, T)$ is said to be *asymptotically mean stationary* (*a.m.s.*) [1] [GK80] if there exists a measure $\overline{\mu}$ on $(\Omega, \mathscr{F})$ satisfying

$$\overline{\mu}(B) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m} \mu(T^{-i}B), \forall\, B \in \mathscr{F}. \quad (8)$$

Obviously, if $(\Omega, \mathscr{F}, \mu, T)$ is stationary, i.e. $\mu(B) = \mu(T^{-1}B)$, then it is a.m.s.. In addition, $(\Omega, \mathscr{F}, \mu, T)$ is said to be ergodic if

$$T^{-1}B = B \implies \mu(B) = 0 \text{ or } \mu(B) = 1. \quad (9)$$

The random process $\{X^{(n)}\} = \{X(T^n)\}$ is said to be *a.m.s.* (stationary/ergodic) if $(\Omega, \mathscr{F}, \mu, T)$ is a.m.s. (stationary/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].

Induced Transformations and Reduced Processes (I)

Definition 2

A dynamical system $(\Omega, \mathscr{F}, \mu, T)$ is said to be *recurrent* (*conservative*) if $\mu\left(B - \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j}B\right) = 0, \forall\, B \in \mathscr{F}$.

Induced Transformations and Reduced Processes (I)

Definition 2

A dynamical system $(\Omega, \mathscr{F}, \mu, T)$ is said to be *recurrent* (*conservative*) if $\mu\left(B - \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j}B\right) = 0, \forall\, B \in \mathscr{F}$.

Given a recurrent system $(\Omega, \mathscr{F}, \mu, T)$ and $A \in \mathscr{F}$ ($\mu(A) > 0$), one can define a new transformation T_A on $(A_0, \mathscr{A}, \mu|_{\mathscr{A}})$, where $A_0 = A \cap \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j}A$ and $\mathscr{A} = \{A_0 \cap B | B \in \mathscr{F}\}$, such that

$$T_A(x) = T^{\psi_A^{(1)}(x)}(x), \forall\, x \in A_0, \tag{10}$$

where

$$\psi_A^{(1)}(x) = \min\left\{i \in \mathbb{N}^+ | T^i(x) \in A_0\right\} \tag{11}$$

is the first return time function.

1. $(A_0, \mathscr{A}, \mu|_{\mathscr{A}}, T_A)$ forms a new dynamical system;
2. T_A is called an *induced transformation* of $(\Omega, \mathscr{F}, \mu, T)$ with respect to A [Kak43].

Induced Transformations and Reduced Processes (II)

Let $\{X^{(n)}\}$ be a random process with state space $\mathscr{X}$. A reduced process $\left\{X_{\mathscr{Y}}^{(k)}\right\}$ of $\{X^{(n)}\}$ with sub-state space $\mathscr{Y} \subseteq \mathscr{X}$ is defined to be $\left\{X_{\mathscr{Y}}^{(k)}\right\} = \left\{X^{(n_k)}\right\}$, where

$$n_k = \begin{cases} \min\{n \geq 0 | X^{(n)} \in \mathscr{Y}\}; & k = 0, \\ \min\{n > n_{k-1} | X^{(n)} \in \mathscr{Y}\}; & k > 0. \end{cases}$$

Induced Transformations and Reduced Processes (II)

Let $\{X^{(n)}\}$ be a random process with state space $\mathscr{X}$. A reduced process $\left\{X_{\mathscr{Y}}^{(k)}\right\}$ of $\{X^{(n)}\}$ with sub-state space $\mathscr{Y} \subseteq \mathscr{X}$ is defined to be $\left\{X_{\mathscr{Y}}^{(k)}\right\} = \left\{X^{(n_k)}\right\}$, where

$$n_k = \begin{cases} \min\{n \geq 0 | X^{(n)} \in \mathscr{Y}\}; & k = 0, \\ \min\{n > n_{k-1} | X^{(n)} \in \mathscr{Y}\}; & k > 0. \end{cases}$$

Assume that $\{X^{(n)}\} = \{X(T^n)\}$ defined by $(\Omega, \mathscr{F}, \mu, T)$ and the measurable function $X : \Omega \to \mathscr{X}$, and let $A = X^{-1}(\mathscr{Y})$. It is easily seen that $\left\{X_{\mathscr{Y}}^{(k)}\right\}$ is essentially the random process $\left\{X\left(T_A^k\right)\right\}$ defined by the system

$$\left(A_0, A_0 \cap \mathscr{F}, \frac{1}{\mu(A)} \mu|_{A_0 \cap \mathscr{F}}, T_A\right) \tag{12}$$

and X.

Supremus Typicality in the Weak Sense

Let $\mathbf{x}_{\mathscr{Y}}$ be the subsequence of $\mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathscr{X}^n$ formed by all those $x^{(l)}$'s that belong to $\mathscr{Y} \subseteq \mathscr{X}$ in the original ordering. $\mathbf{x}_{\mathscr{Y}}$ is called a *reduced subsequence* of $\mathbf{x}$ with respect to $\mathscr{Y}$.

Supremus Typicality in the Weak Sense

Let $\mathbf{x}_{\mathscr{Y}}$ be the subsequence of $\mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathscr{X}^n$ formed by all those $x^{(l)}$'s that belong to $\mathscr{Y} \subseteq \mathscr{X}$ in the original ordering. $\mathbf{x}_{\mathscr{Y}}$ is called a *reduced subsequence* of $\mathbf{x}$ with respect to $\mathscr{Y}$.

Definition 3 (Supremus Typicality in the Weak Sense [HS14])

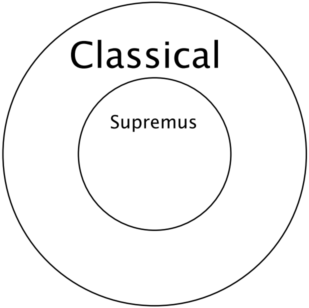
Let $\{X^{(n)}\}$ be a recurrent a.m.s. ergodic process with state space $\mathscr{X}$. A sequence $\mathbf{x} \in \mathscr{X}^n$ is said to be *Supremus ϵ-typical* with respect to $\{X^{(n)}\}$ for some $\epsilon > 0$, if

$$|\mathbf{x}_{\mathscr{Y}}| (H_{\mathscr{Y}} - \epsilon) < -\log p_{\mathscr{Y}}(\mathbf{x}_{\mathscr{Y}}) < |\mathbf{x}_{\mathscr{Y}}| (H_{\mathscr{Y}} + \epsilon), \forall\, \emptyset \neq \mathscr{Y} \subseteq \mathscr{X}, \quad (13)$$

where $p_{\mathscr{Y}}$ and $H_{\mathscr{Y}}$ are the joint distribution and entropy rate of the reduced process $\left\{X_{\mathscr{Y}}^{(k)}\right\}$ of $\{X^{(n)}\}$ with sub-state space $\mathscr{Y}$, respectively.

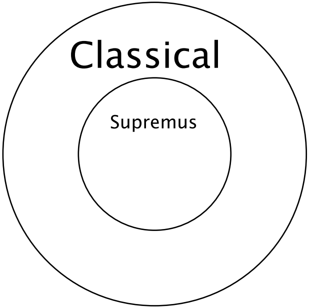
Asymptotical Equipartition Property

Designate $\mathcal{S}_\epsilon(n, \{X^{(n)}\})$ as the set of all Supremus ϵ-typical sequences with respect to $\{X^{(n)}\}$ in $\mathscr{X}^n$. Obviously, $\mathcal{S}_\epsilon(n, \{X^{(n)}\})$ is a subset of all classical ϵ-typical sequences [SW49].



Asymptotical Equipartition Property

Designate $\mathcal{S}_\epsilon(n, \{X^{(n)}\})$ as the set of all Supremus ϵ-typical sequences with respect to $\{X^{(n)}\}$ in $\mathscr{X}^n$. Obviously, $\mathcal{S}_\epsilon(n, \{X^{(n)}\})$ is a subset of all classical ϵ-typical sequences [SW49].



Theorem 4 (AEP of Weak Supremus Typicality [HS14])

In Definition 3, $\forall\ \eta > 0$, there exists some positive integer N_0, such that

$$\Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] \notin \mathcal{S}_\epsilon(n, \{X^{(n)}\})\right\} < \eta,$$

for all $n > N_0$.

Supporting Results

Theorem 5 ([HS13b])

If $(\Omega, \mathscr{F}, \mu, T)$ is recurrent a.m.s., then $(A_0, \mathscr{A}, \mu|_{\mathscr{A}}, T_A)$ is a.m.s. for all $A \in \mathscr{F}$ $(\mu(A) > 0)$.

Supporting Results

Theorem 5 ([HS13b])

If $(\Omega, \mathscr{F}, \mu, T)$ is recurrent a.m.s., then $(A_0, \mathscr{A}, \mu|_{\mathscr{A}}, T_A)$ is a.m.s. for all $A \in \mathscr{F}$ $(\mu(A) > 0)$.

Theorem 6 (Shannon–McMillan–Breiman–Gray Theorem [GK80])

If

$$\{X^{(n)}\} = \{X(T^n)\}$$

is a.m.s. and ergodic, then the Shannon–McMillan–Breiman 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 $\{X^{(n)}\}$.

The Proof (I)

Let $\mathbf{X} = \left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right]$. Then

$$\left\{\mathbf{X} \notin \mathcal{S}_{\epsilon}(n, \{X^{(n)}\})\right\} = \bigcup_{\emptyset \neq \mathscr{Y} \subseteq \mathscr{X}} \left\{\mathbf{X}_{\mathscr{Y}} \notin \mathcal{T}_{\epsilon}(n, \{X_{\mathscr{Y}}^{(k)}\})\right\}. \tag{14}$$

Assume that $(\Omega, \mathscr{F}, \mu, T)$ and X are the recurrent a.m.s. ergodic system and the measurable function define $\{X^{(n)}\}$, i.e. $\{X^{(n)}\} = \{X(T^n)\}$. For any non-empty $\mathscr{Y} \subseteq \mathscr{X}$, we have that $\left\{X_{\mathscr{Y}}^{(k)}\right\} = \left\{X\left(T_A^k\right)\right\}$, where $A = X^{-1}(\mathscr{Y})$ and T_A is an induced transformation of $(\Omega, \mathscr{F}, \mu, T)$ with respect to A. Furthermore, Theorem 5 and [Aar97, Proposition 1.5.2] guarantee that

$$\left(A_0, A_0 \cap \mathscr{F}, \frac{1}{\mu(A)} \mu|_{A_0 \cap \mathscr{F}}, T_A\right), \tag{15}$$

is a.m.s. ergodic.

The Proof (II)

Consequently, the Shannon–McMillan–Breiman–Gray Theorem (Theorem 6) says

$$-\frac{1}{n}\log p_{\mathscr{Y}}\left(X_{\mathscr{Y}}^{(0)}, X_{\mathscr{Y}}^{(1)}, \cdots, X_{\mathscr{Y}}^{(n-1)}\right) \to H_{\mathscr{Y}}, \text{ with probability 1.}$$

This implies that there exists a positive integer $N_{\mathscr{Y}}$ such that

$$\Pr\left\{\mathbf{X}_{\mathscr{Y}} \notin \mathcal{T}_{\epsilon}(n, \{X_{\mathscr{Y}}^{(k)}\})\right\} < \frac{\eta}{2^{|\mathscr{X}|} - 1}, \forall\ n > N_{\mathscr{Y}}.$$

Let $N_0 = \max_{\emptyset \neq \mathscr{Y} \subseteq \mathscr{X}} N_{\mathscr{Y}}$. One easily concludes that

$$\Pr\left\{\mathbf{X} \notin \mathcal{S}_{\epsilon}(n, \{X^{(n)}\})\right\} < \eta, \forall\ n > N_0.$$

The statement is proved.

Thanks

♫ ♬

Thanks!

Bibliography I

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

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

Sheng Huang and Mikael Skoglund.
Encoding Irreducible Markovian Functions of Sources: An Application of Supremus Typicality.
KTH Royal Institute of Technology, May 2013.

Sheng Huang and Mikael Skoglund.
Induced Transformations of Recurrent A.M.S. Dynamical Systems.
KTH Royal Institute of Technology, October 2013.

Bibliography II

Sheng Huang and Mikael Skoglund.
Supremus Typicality.
KTH Royal Institute of Technology, January 2014.

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

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

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