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

1 Asymptotically Mean Stationary (A.M.S.)
   - A Starting Example
   - A.M.S. Dynamical Systems and A.M.S. Random Processes
   - Induced Transformations and Reduced Processes

2 Supremus Typicality in the Weak Sense
   - Supremus Typicality in the Weak Sense
   - Asymptotical Equipartition Property

3 Proof of the AEP
   - Supporting Results
   - The Proof

4 Thanks / References
   - Thanks
   - Bibliography
A Starting Example

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.
A Starting Example

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. 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

\[
\left| \frac{\text{the number of subsequence } (\alpha, \alpha)'s \text{ in } x_{\{\alpha, \gamma\}}}{6} - 0.5 \right| = |0 - 0.5| > \frac{5}{12}.
\]

(4)
A.M.S. Random Processes (I)

Given a probability space \((\Omega, \mathcal{F}, \mu)\) and a measurable transformation \(T : \Omega \to \Omega\) (not necessarily probability preserving), the tuple \((\Omega, \mathcal{F}, \mu, T)\) is called a \textit{dynamical system} (or \textit{ergodic system}).
A.M.S. Random Processes (I)

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

Let \(X : \Omega \to X\) (e.g. \(X\) is a finite set) be a measurable function. Then

\[
\{X^{(n)}\} = \{X(T^n)\}
\]

defines a random process with state space \(X\) and pdf/pmf

\[
p(x^{(0)}, x^{(1)}, \ldots, x^{(n-1)}) = \mu \left( \bigcap_{i=0}^{n-1} T^{-i} (X^{-1}(x^{(i)})) \right).
\]
Given a probability space \((\Omega, \mathcal{F}, \mu)\) and a measurable transformation \(T : \Omega \to \Omega\) (not necessarily probability preserving), the tuple \((\Omega, \mathcal{F}, \mu, T)\) is called a dynamical system (or ergodic system).

Let \(X : \Omega \to \mathcal{X}\) (e.g. \(\mathcal{X}\) is a finite set) be a measurable function. Then

\[
\{X^{(n)}\} = \{X(T^n)\}
\]

(5)
defines a random process with state space \(\mathcal{X}\) and pdf/pmf

\[
p(x^{(0)}, x^{(1)}, \ldots, x^{(n-1)}) = \mu \left( \bigcap_{i=0}^{n-1} T^{-i} \left( X^{-1}(x^{(i)}) \right) \right).
\]

(6)

Example 1

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

\[
X : (\cdots, x^{(-1)}, x^{(0)}, x^{(1)}, \cdots) \mapsto x^{(0)}.
\]

(7)

Define \(\mu\) satisfying

\[
\mu \left( \bigcap_{i=0}^{n-1} T^{-i} \left( X^{-1}(x^{(i)}) \right) \right) = p(x^{(0)}, x^{(1)}, \ldots, x^{(n-1)}).
\]

By the Kolmogorov Extension Theorem, \(\{X^{(n)}\} = \{X(T^n)\}\).
(Ω, ℱ, μ, T) is said to be asymptotically mean stationary (a.m.s.) \(^1\) [GK80] if there exists a measure \(\overline{\mu}\) on \((Ω, ℱ)\) satisfying

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

\(^1\)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].
(Ω, ℱ, μ, T) is said to be *asymptotically mean stationary* (a.m.s.) \(^1\) [GK80] if there exists a measure \(\tilde{\mu}\) on \((Ω, ℱ)\) satisfying

\[
\tilde{\mu}(B) = \lim_{m \to \infty} \frac{1}{m} \sum_{i=0}^{m} \mu(T^{-i}B), \forall B \in ℱ.
\]  

(8)

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

\[
T^{-1}B = B \implies μ(B) = 0 \text{ or } μ(B) = 1.
\]  

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

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

(8)

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

\[
T^{-1}B = B \implies µ(B) = 0 \text{ or } µ(B) = 1.
\]  

(9)

The random process \(\{X^{(n)}\} = \{X(T^n)\}\) is said to be a.m.s. (stationary/ergodic) if \((Ω, ℱ, µ, T)\) 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].
Definition 2

A dynamical system \((\Omega, \mathcal{F}, \mu, T)\) is said to be **recurrent** (**conservative**) if

\[
\mu \left( B - \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j} B \right) = 0, \forall B \in \mathcal{F}.
\]
Definition 2

A dynamical system \((\Omega, \mathcal{F}, \mu, T)\) is said to be recurrent (conservative) if

\[
\mu \left( B - \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} T^{-j} B \right) = 0, \forall B \in \mathcal{F}.
\]

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

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

where

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

is the first return time function.

- \((A_0, \mathcal{A}, \mu|_A, T_A)\) forms a new dynamical system;
- \(T_A\) is called an induced transformation of \((\Omega, \mathcal{F}, \mu, T)\) with respect to \(A\) [Kak43].
Let \( \{X^{(n)}\} \) be a random process with state space \( \mathcal{X} \). A reduced process \( \{X_{\mathcal{Y}}^{(k)}\} \) of \( \{X^{(n)}\} \) with sub-state space \( \mathcal{Y} \subseteq \mathcal{X} \) is defined to be

\[
\{X_{\mathcal{Y}}^{(k)}\} = \{X^{(n_k)}\},
\]

where

\[
n_k = \begin{cases} 
\min\{n \geq 0|X^{(n)} \in \mathcal{Y}\}; & k = 0, \\
\min\{n > n_{k-1}|X^{(n)} \in \mathcal{Y}\}; & k > 0.
\end{cases}
\]
Induced Transformations and Reduced Processes (II)

Let \( \{X^{(n)}\} \) be a random process with state space \( \mathcal{X} \). A reduced process \( \{X_{\mathcal{Y}}^{(k)}\} \) of \( \{X^{(n)}\} \) with sub-state space \( \mathcal{Y} \subseteq \mathcal{X} \) is defined to be
\[
\{X_{\mathcal{Y}}^{(k)}\} = \{X^{(n_k)}\},
\]
where
\[
n_k = \begin{cases} 
\min\{n \geq 0|X^{(n)} \in \mathcal{Y}\}; & k = 0, \\
\min\{n > n_{k-1}|X^{(n)} \in \mathcal{Y}\}; & k > 0. 
\end{cases}
\]

Assume that \( \{X^{(n)}\} = \{X(T^n)\} \) defined by \( (\Omega, \mathcal{F}, \mu, T) \) and the measurable function \( X : \Omega \to \mathcal{X} \), and let \( A = X^{-1}(\mathcal{Y}) \). It is easily seen that \( \{X_{\mathcal{Y}}^{(k)}\} \) is essentially the random process \( \{X(T_A^k)\} \) defined by the system
\[
\left( A_0, A_0 \cap \mathcal{F}, \frac{1}{\mu(A)} \mu|_{A_0 \cap \mathcal{F}}, T_A \right) \quad (12)
\]
and \( X \).
Let $\mathbf{x}_\mathcal{Y}$ be the subsequence of $\mathbf{x} = \left[ x^{(1)}, x^{(2)}, \cdots, x^{(n)} \right] \in \mathcal{X}^n$ formed by all those $x^{(l)}$’s that belong to $Y \subseteq \mathcal{X}$ in the original ordering. $\mathbf{x}_\mathcal{Y}$ is called a *reduced subsequence* of $\mathbf{x}$ with respect to $\mathcal{Y}$. 
Supremus Typicality in the Weak Sense

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

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

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

$$|\mathbf{x}_Y| (H_Y - \epsilon) < -\log p_Y (\mathbf{x}_Y) < |\mathbf{x}_Y| (H_Y + \epsilon), \forall \emptyset \neq Y \subseteq \mathcal{X}, \quad (13)$$

where $p_Y$ and $H_Y$ are the joint distribution and entropy rate of the reduced process $\{X^{(k)}_Y\}$ of $\{X^{(n)}\}$ with sub-state space $Y$, respectively.
Designate $S_\epsilon(n, \{X^{(n)}\})$ as the set of all Supremus $\epsilon$-typical sequences with respect to $\{X^{(n)}\}$ in $\mathcal{X}^n$. Obviously, $S_\epsilon(n, \{X^{(n)}\})$ is a subset of all classical $\epsilon$-typical sequences [SW49].
Asymptotical Equipartition Property

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

$$\Pr \left\{ \left[ X^{(1)}, X^{(2)}, \ldots, X^{(n)} \right] \notin S_\epsilon(n, \{X^{(n)}\}) \right\} < \eta,$$

for all $n > N_0$. 

Sheng Huang and Mikael Skoglund  
Internal Seminar
Theorem 5 ([HS13b])

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

Theorem 5 ([HS13b])

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

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,

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

where \(H\) is the entropy rate of \(\{X^{(n)}\}\).
The Proof (I)

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

$$\left\{ X \notin S_\epsilon(n, \{X^{(n)}\}) \right\} = \bigcup_{\emptyset \neq \mathcal{Y} \subseteq X} \left\{ X_\mathcal{Y} \notin T_\epsilon(n, \{X^{(k)}_\mathcal{Y}\}) \right\}. \quad (14)$$

Assume that $(\Omega, \mathcal{F}, \mu, T)$ 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 $\mathcal{Y} \subseteq X$, we have that $\{X^{(k)}_\mathcal{Y}\} = \{X(T^k_A)\}$, where $A = X^{-1}(\mathcal{Y})$ and $T_A$ is an induced transformation of $(\Omega, \mathcal{F}, \mu, T)$ with respect to $A$. Furthermore, Theorem 5 and [Aar97, Proposition 1.5.2] guarantee that

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

is a.m.s. ergodic.
Consequently, the Shannon–McMillan–Breiman–Gray Theorem (Theorem 6) says

\[-\frac{1}{n} \log p_{\mathcal{Y}} \left( X^{(0)}_{\mathcal{Y}}, X^{(1)}_{\mathcal{Y}}, \ldots, X^{(n-1)}_{\mathcal{Y}} \right) \to H_{\mathcal{Y}}, \text{ with probability 1.} \]

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

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

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

\[\Pr \left\{ X \notin S_{\epsilon}(n, \{X^{(n)}\}) \right\} < \eta, \quad \forall \ n > N_0. \]

The statement is proved.
Thanks

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

Robert M. Gray and J. C. Kieffer.
Asymptotically mean stationary measures.

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.  

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

Claude Elwood Shannon and Warren Weaver.  
*The mathematical theory of communication.*  