THE ERGODICITY OF THE OBSERVED PROCESS OF
A HIDDEN MARKOV MODEL

BERLIAN SETIAWATY

Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University
Jl. Raya Pajajaran, Kampus IPB Baranangsiang, Bogor, Indonesia

Abstract. This paper presents some properties of a stationary
hidden Markov model. The most important is the ergodicity of
the observed process which is essential for limit theorems.

Key words: Hidden Markov, stationary, ergodic.

1. Introduction

According to [6], if the Markov chain of a hidden Markov model
is stationary, then the observed process is also stationary. As a sta-
tionary process, the observed process has several properties, the most
important is ergodicity. The ergodicity is essential for limit theo-
rems. Therefore, finding sufficient conditions for the ergodicity of the
observed process will be the focus of this paper.

We will begin with definition of a hidden Markov model and some
properties of a stationary hidden Markov model. Then in the last
section, we show sufficient conditions for the ergodicity of the observed
process of a hidden Markov model.

2. A Stationary Hidden Markov Model

Let \( \{X_t : t \in \mathbb{N}\} \) be a finite state Markov chain defined on a pro-
bability space \((\Omega, \mathcal{F}, P)\). Suppose that \(\{X_t\}\) is not observed directly,
but rather there is an observation process \(\{Y_t : t \in \mathbb{N}\}\) defined on
\((\Omega, \mathcal{F}, P)\). Then consequently, the Markov chain is said to be hidden
in the observations. A pair of stochastic processes \(\{(X_t, Y_t) : t \in \mathbb{N}\}\)
is called a hidden Markov model. Precisely, according to [5], a hidden
Markov model is formally defined as follows.

**Definition 2.1.** A pair of discrete time stochastic processes \(\{(X_t, Y_t) : \)
t \in \mathbb{N}\} \) defined on a probability space \((\Omega, \mathcal{F}, P)\) and taking values in
a set \(S \times \mathcal{Y}\), is said to be a hidden Markov model (HMM), if it
satisfies the following conditions.
1. \{X_t\} is a finite state Markov chain.
2. Given \{X_t\}, \{Y_t\} is a sequence of conditionally independent random variables.
3. The conditional distribution of \(Y_n\) depends on \{X_t\} only through \(X_n\).
4. The conditional distribution of \(Y_t\) given \(X_t\) does not depend on \(t\).

Assume that the Markov chain \{X_t\} is not observable. The cardinality \(K\) of \(S\), will be called the size of the hidden Markov model.

Let \{(X_t, Y_t)\} be a hidden Markov model defined on a probability space \((\Omega, \mathcal{F}, P)\), taking values on \(S \times \mathcal{Y}\), where \(S = \{1, \ldots, K\}\) and \(\mathcal{Y} = \mathbb{R}\).

Let \(\Lambda\) be the set of all realizations \{(x_t, y_t)\} of the hidden Markov model \{(X_t, Y_t)\}. Let \(\mathcal{B}_\Lambda\) be the Borel \(\sigma\)-field of \(\Lambda\). For each \(t \in \mathbb{N}\), define mappings

\[
\tilde{X}_t : \Lambda \rightarrow S, \\
\tilde{Y}_t : \Lambda \rightarrow \mathcal{Y},
\]

by

\[
\tilde{X}_t(\lambda) = x_t \\
\tilde{Y}_t(\lambda) = y_t,
\]

for \(\lambda = \{(x_t, y_t)\} \in \Lambda\). For \(t \in \mathbb{N}\), \(\tilde{X}_t, \tilde{Y}_t\) are coordinate projections on \(\Lambda\). When later a probability measure is defined on \(\mathcal{B}_\Lambda\), then \(\tilde{X}_t, \tilde{Y}_t\) will be random variables.

Next lemma shows that there is a probability measure \(\tilde{P}\) defined on \(\mathcal{B}_\Lambda\) such that the hidden Markov model \{(X_t, Y_t)\} is equivalent with the pair of processes \{(\tilde{X}_t, \tilde{Y}_t)\}, that is, \{(X_t, Y_t)\} and \{(\tilde{X}_t, \tilde{Y}_t)\} have the same law.

**Lemma 2.2.** There exists a probability measure \(\tilde{P}\) defined on \(\mathcal{B}_\Lambda\) such that the pair of coordinate projections \{(\tilde{X}_t, \tilde{Y}_t)\} and the hidden Markov model \{(X_t, Y_t)\} are equivalent.

**Proof:**

The idea of the proof comes from [2], page 511.
For each $k \in \mathbf{N}$ and distinct $t_1, \ldots, t_k \in \mathbf{N}$, let $\nu_{t_1,\ldots, t_k}$ be the joint distribution of $X_{t_1}, \ldots, X_{t_k}; Y_{t_1}, \ldots, Y_{t_k}$,
\[
\nu_{t_1,\ldots, t_k}(A \times B) = \mathbb{P}\{(X_{t_1}, \ldots, X_{t_k}) \in A, (Y_{t_1}, \ldots, Y_{t_k}) \in B\}, \quad (2.1)
\]
for $A \in \mathcal{S}_k$ and $B \in \mathcal{B}_k$, where $\mathcal{S}_k$ and $\mathcal{B}_k$ are the Borel $\sigma$-field of $\mathbf{S}^k$ and $\mathcal{Y}^k$ respectively.

Define a mapping
\[
\zeta : \Omega \rightarrow \Lambda,
\]
by the requirement
\[
\tilde{X}_t(\zeta(\omega)) = X_t(\omega) = x_t \quad \text{and} \quad \tilde{Y}_t(\zeta(\omega)) = Y_t(\omega) = y_t,
\]
for $\omega \in \Omega$ and $t \in \mathbf{N}$. Clearly,
\[
\zeta^{-1}\{\lambda \in \Lambda : (\tilde{X}_{t_1}(\lambda), \ldots, \tilde{X}_{t_k}(\lambda)) \in A, (\tilde{Y}_{t_1}(\lambda), \ldots, \tilde{Y}_{t_k}(\lambda)) \in B\}
= \{\omega \in \Omega : (\tilde{X}_{t_1}(\zeta(\omega)), \ldots, \tilde{X}_{t_k}(\zeta(\omega))) \in A, (\tilde{Y}_{t_1}(\zeta(\omega)), \ldots, \tilde{Y}_{t_k}(\zeta(\omega))) \in B\}
= \{\omega \in \Omega : (X_{t_1}(\omega), \ldots, X_{t_k}(\omega)) \in A, (Y_{t_1}(\omega), \ldots, Y_{t_k}(\omega)) \in B\}
\in \mathcal{F}, \quad (2.2)
\]
if $A \in \mathcal{S}_k$ and $B \in \mathcal{B}_k$. Thus $\zeta$ is measurable.

Define probability measure $\tilde{P} = \mathbb{P}_\zeta^{-1}$ on $\mathcal{B}_{\Lambda}$, then from (2.1) and (2.2),
\[
\tilde{P}\{\lambda \in \Lambda : (\tilde{X}_{t_1}(\lambda), \ldots, \tilde{X}_{t_k}(\lambda)) \in A, (\tilde{Y}_{t_1}(\lambda), \ldots, \tilde{Y}_{t_k}(\lambda)) \in B\}
= \mathbb{P}_\zeta^{-1}\tilde{P}\{\lambda \in \Lambda : (\tilde{X}_{t_1}(\lambda), \ldots, \tilde{X}_{t_k}(\lambda)) \in A, (\tilde{Y}_{t_1}(\lambda), \ldots, \tilde{Y}_{t_k}(\lambda)) \in B\}
= \mathbb{P}\{\omega \in \Omega : (X_{t_1}(\omega), \ldots, X_{t_k}(\omega)) \in A, (Y_{t_1}(\omega), \ldots, Y_{t_k}(\omega)) \in B\}
= \nu_{t_1,\ldots, t_k}(A \times B). \quad (2.3)
\]

The equation (2.3) shows that $\{(\tilde{X}_t, \tilde{Y}_t)\}$, defined on $(\Lambda, \mathcal{B}_\Lambda, \tilde{P})$ also has finite dimensional distribution $\nu_{t_1,\ldots, t_k}$. Thus $\{(X_t, Y_t)\}$ and $\{(\tilde{X}_t, \tilde{Y}_t)\}$ are equivalent.

Remarks 2.3. By Lemma 2.2, from now on, the hidden Markov model $\{(X_t, Y_t)\}$ may be considered as the pair of coordinate projection processes $\{(\tilde{X}_t, \tilde{Y}_t)\}$, defined on $(\Lambda, \mathcal{B}_\Lambda, \tilde{P})$. For convenience, we will drop the tilde.

Suppose that the Markov chain $\{X_t\}$ is stationary, then from [6], the hidden Markov model $\{(X_t, Y_t)\}$ is also stationary. We want to build a past for the hidden Markov model $\{(X_t, Y_t) : t \in \mathbf{N}\}$ without losing its stationarity. The problem is to find a pair of stochastic processes $\{(\tilde{X}_t, \tilde{Y}_t) : t \in \mathbf{Z}\}$ such that $\{(X_t, Y_t) : t \in \mathbf{N}\}$ and $\{(\tilde{X}_t, \tilde{Y}_t) : t \in \mathbf{N}\}$ have the same law.
Lemma 2.4. There is a stationary process \{ (\bar{X}_t, \bar{Y}_t) \} indexed by \( t \in \mathbb{Z} \), unique up to equivalence, such that \{ (X_t, Y_t) : t \in \mathbb{N} \} and \{ (\bar{X}_t, \bar{Y}_t) : t \in \mathbb{N} \} are equivalent processes.

**Proof:**
The proof follows from [1], page 21.

Let \( I = \{ t_1, t_2, \ldots, t_k \} \in \mathbb{Z} \). For all \( r \) large enough, the integers \( I_r = \{ r+t_1, r+t_2, \ldots, r+t_k \} \subset \mathbb{N} \) and the joint law of \{ (X_t, Y_t) : t \in I_r \} is independent of \( r \), since \{ (X_t, Y_t) \} is stationary. Let \( \Pi_I \) be this law. The family \( \Pi_I \) is consistent. Kolmogorov consistency theorem ([1], page 6) grants the existence of the process \{ (\bar{X}_t, \bar{Y}_t) \} indexed by \( \mathbb{Z} \), such that for all \( I \) as above \( \Pi_I \) is the joint law of \{ (\bar{X}_t, \bar{Y}_t) : t \in I \}. Clearly \{ (X_t, Y_t) : t \in \mathbb{N} \} and \{ (\bar{X}_t, \bar{Y}_t) : t \in \mathbb{N} \} are equivalent processes.

**Remarks 2.5.** Without loss of generality, by Lemma 2.4, now we have the stationary hidden Markov model \{ (X_t, Y_t) : t \in \mathbb{Z} \}, defined on the probability space \( (\Lambda, \mathcal{B}_\Lambda, P) \), where \( \Lambda \) is the set of realizations \( \lambda = \{ (x_t, y_t) \} \), \( \mathcal{B}_\Lambda \) is the Borel \( \sigma \)-field of \( \Lambda \) and \( X_t, Y_t \) are coordinate projections defined on \( \Lambda \).

3. The Ergodicity of the Observed Process

If \( z = \{ z_t \} \) is a real sequence, let \( Tz \) denote the shifted sequence \( \{ z_{t+1} \} \). \( T \) is called the shift operator. A set of \( \mathcal{A} \) of real sequences is called shift invariant, when \( Tz \in \mathcal{A} \) if and only if \( z \in \mathcal{A} \). A stationary process \( Z = \{ Z_t \} \) is said to be ergodic if

\[
P(Z \in \mathcal{A}) = 0 \quad \text{or} \quad 1,
\]

whenever \( \mathcal{A} \) is shift invariant.

From [7], page 33, a stationary and irreducible Markov chain is ergodic. Based on this, Leroux [4] derived the ergodicity of the observed process \{ \( Y_t \) \}.

**Lemma 3.1 (Leroux [4]).** If the Markov chain \( \{ X_t \} \) is stationary and irreducible, then the observed process \( \{ Y_t \} \) is stationary and ergodic.

**Proof:**
Let \( \mathcal{A} \) be a shift invariant set of sequences \( y = \{ y_t \} \) of possible realizations of \( Y = \{ Y_t \} \). It will be proved that

\[
P(Y \in \mathcal{A}) = 0 \quad \text{or} \quad 1.
\]

By the approximation theorem ([2], page 167), there is a subsequence \( \{ k' \} \) and cylinder set \( \mathcal{A}_{k'} \) having form

\[
\mathcal{A}_{k'} = \left\{ \lambda \in \Lambda : (Y_{-k'}(\lambda), \ldots, Y_{k'}(\lambda)) \in B_{2k'} \right\}
= \left\{ \lambda \in \Lambda : (y_{-k'}, \ldots, y_{k'}) \in B_{2k'} \right\},
\]
where $B_{2^k} \in \mathcal{B}_{2^k}$, that is the Borel $\sigma$-field of $\mathcal{Y}_{2^k}$, such that $\forall k \in \mathbb{N}$,

$$P(Y \in \mathcal{A} \triangle \mathcal{A}_{k'}) < 2^{-k}. \tag{3.1}$$

Since $Y$ is stationary and $\mathcal{A}$ is invariant, then

$$P(Y \in \mathcal{A} \triangle \mathcal{A}_{k'}) = P(T^{2^k}Y \in \mathcal{A} \triangle \mathcal{A}_{k'})$$
$$= P(Y \in \mathcal{A} \triangle T^{-2^k}\mathcal{A}_{k'})$$
$$= P(Y \in \mathcal{A} \triangle \tilde{\mathcal{A}}_{k'}), \tag{3.2}$$

where

$$\tilde{\mathcal{A}}_{k'} = T^{-2^k}\mathcal{A}_{k'}$$
$$= \left\{ \lambda \in \Lambda : (Y_{k'}(\lambda), \ldots, Y_{3k'}(\lambda)) \in B_{2^k} \right\}$$
$$= \left\{ \lambda \in \Lambda : (y_{k'}, \ldots, y_{3k'}) \in B_{2^k} \right\}.$$

Let

$$\tilde{\mathcal{A}} = \bigcap_{k \in \mathbb{N}} \bigcup_{j \geq k} \tilde{\mathcal{A}}_j,$$

then

$$\mathcal{A}^c \cap \tilde{\mathcal{A}} = \mathcal{A}^c \cap \left( \bigcap_{k \geq 1} \bigcup_{j \geq k} \tilde{\mathcal{A}}_j \right)$$
$$= \bigcup_{k \geq 1} \left( \mathcal{A}^c \cap \tilde{\mathcal{A}}_j \right)$$
$$= \limsup_{k' \to \infty} \mathcal{A}^c \cap \tilde{\mathcal{A}}_{k'}$$

and

$$\mathcal{A} \cap \tilde{\mathcal{A}}^c = \mathcal{A} \cap \left( \bigcap_{k \geq 1} \bigcup_{j \geq k} \tilde{\mathcal{A}}_j \right)^c$$
$$= \mathcal{A} \cap \left( \bigcup_{k \geq 1} \bigcap_{j \geq k} \tilde{\mathcal{A}}_j \right)^c$$
$$= \bigcup_{k \geq 1} \left( \mathcal{A} \cap \tilde{\mathcal{A}}_j^c \right)$$
$$= \liminf_{k' \to \infty} \mathcal{A} \cap \tilde{\mathcal{A}}_{k'}^c.$$
Hence,
\[
\mathcal{A} \triangle \tilde{A} = (\mathcal{A} \cap \tilde{A}^c) \cup (\mathcal{A}^c \cap \tilde{A})
\]
\[
= \left( \liminf_{k' \to \infty} \mathcal{A} \cap \tilde{A}^c_{k'} \right) \cup \left( \limsup_{k' \to \infty} \mathcal{A}^c \cap \tilde{A}_{k'} \right)
\]
\[
\subset \left( \limsup_{k' \to \infty} \mathcal{A} \cap \tilde{A}^c_{k'} \right) \cup \left( \limsup_{k' \to \infty} \mathcal{A}^c \cap \tilde{A}_{k'} \right)
\]
\[
= \limsup_{k' \to \infty} \left( (\mathcal{A} \cap \tilde{A}^c_{k'}) \cup (\mathcal{A}^c \cap \tilde{A}_{k'}) \right)
\]
\[
= \limsup_{k' \to \infty} \left( \mathcal{A} \triangle \tilde{A}_{k'} \right). \tag{3.3}
\]

From (3.1) and (3.2)
\[
\sum_k P(Y \in \mathcal{A} \triangle \tilde{A}_{k'}) = \sum_k P(Y \in \mathcal{A} \triangle \tilde{A}_{k'}) 
\leq \sum_k 2^{-k}
\]
\[
= 1,
\]
then by Borrel Cantelli’s Lemma,
\[
0 \leq P(Y \in \mathcal{A} \triangle \tilde{A}) \leq P(Y \in \limsup \mathcal{A} \triangle \tilde{A}_{k'}) = 0,
\]
implying
\[
P(\mathcal{A} \triangle \tilde{A}) = 0. \tag{3.4}
\]

Since (3.4) holds, then to prove \( P(Y \in \mathcal{A}) = 0 \) or 1, is equivalent with showing that
\[
P(Y \in \tilde{A}) = 0 \quad \text{or} \quad 1.
\]

By definition, \( \tilde{A} = \bigcap_{k \geq 1} \bigcup_{j \geq k} \tilde{A}_j \), so \( \tilde{A} \) is in the tail \( \sigma \)-field. Since \( Y_t \) are conditionally independent given a realization \( x = \{x_t\} \) of the underlying Markov chain \( X = \{X_t\} \), then the zero-one law implies
\[
P(Y \in \tilde{A}|x) = 0 \quad \text{or} \quad 1.
\]

Let
\[
B = \{x : P(Y \in \tilde{A}|x) = 1\},
\]
so
\[
P(Y \in \tilde{A}) = E \left[ 1_{Y \in \tilde{A}} \right]
\]
\[
= E \left[ E[1_{Y \in \tilde{A}}|x] \right]
\]
\[
= E \left[ P(Y \in \tilde{A}|x) \right]
\]
\[
= 0 + 1 \cdot P(X \in B)
\]
\[
= P(X \in B). \tag{3.5}
\]
But, \( B \) is invariant, as
\[
P(Y \in \tilde{A} | x) = P(TY \in \tilde{A} | Tx) = P(Y \in \tilde{A} | Tx).
\]
Since the Markov chain \( \{X_t\} \) is stationary and irreducible, then \( \{X_t\} \) is ergodic, implying
\[
P(X \in B) = \begin{cases} 0 & \text{or} \ 1. 
\end{cases}
\]
Hence, by (3.5),
\[
P(Y \in \tilde{A}) = \begin{cases} 0 & \text{or} \ 1. 
\end{cases}
\]

References