

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 stationary process, the observed process has several properties, the most important is ergodicity. The ergodicity is essential for limit theorems. 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 $\{X_t : t \in \mathbf{N}\}$ be a finite state Markov chain defined on a probability space (Ω, \mathcal{F}, P) . Suppose that $\{X_t\}$ is not observed directly, but rather there is an *observation* process $\{Y_t : t \in \mathbf{N}\}$ defined on (Ω, \mathcal{F}, P) . Then consequently, the Markov chain is said to be *hidden* in the observations. A pair of stochastic processes $\{(X_t, Y_t) : t \in \mathbf{N}\}$ 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 $\{(X_t, Y_t) : t \in \mathbf{N}\}$ defined on a probability space (Ω, \mathcal{F}, P) and taking values in a set $\mathcal{S} \times \mathcal{Y}$, 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 \mathbf{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 (Ω, \mathcal{F}, P) , taking values on $\mathbf{S} \times \mathcal{Y}$, where $\mathbf{S} = \{1, \dots, K\}$ and $\mathcal{Y} = \mathbf{R}$.

Let Λ be the set of all realizations $\{(x_t, y_t)\}$ of the hidden Markov model $\{(X_t, Y_t)\}$. Let \mathcal{B}_Λ be the Borel σ -field of Λ . For each $t \in \mathbf{N}$, define mappings

$$\tilde{X}_t : \Lambda \longrightarrow \mathbf{S},$$

by

$$\tilde{X}_t(\lambda) = x_t$$

and

$$\tilde{Y}_t : \Lambda \longrightarrow \mathcal{Y},$$

by

$$\tilde{Y}_t(\lambda) = y_t,$$

for $\lambda = \{(x_t, y_t)\} \in \Lambda$. For $t \in \mathbf{N}$, \tilde{X}_t, \tilde{Y}_t are *coordinate projections* on Λ . When later a probability measure is defined on \mathcal{B}_Λ , 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}_Λ 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}_Λ 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.

The hidden Markov model $\{(X_t, Y_t)\}$ is defined on the probability space (Ω, \mathcal{F}, P) . For each $\omega \in \Omega$, let

$$\begin{aligned} X_t(\omega) &= x_t, & t \in \mathbf{N} \\ Y_t(\omega) &= y_t, & t \in \mathbf{N}. \end{aligned}$$

For each $k \in \mathbf{N}$ and distinct $t_1, \dots, t_k \in \mathbf{N}$, let ν_{t_1, \dots, t_k} be the joint distribution of $X_{t_1}, \dots, X_{t_k}; Y_{t_1}, \dots, Y_{t_k}$,

$$\nu_{t_1, \dots, t_k}(A \times B) = P\left\{(X_{t_1}, \dots, X_{t_k}) \in A, (Y_{t_1}, \dots, Y_{t_k}) \in B\right\}, \quad (2.1)$$

for $A \in \mathcal{S}_k$ and $B \in \mathcal{B}_k$, where \mathcal{S}_k and \mathcal{B}_k are the Borel σ -field of \mathbf{S}^k and \mathcal{Y}^k respectively.

Define a mapping

$$\zeta : \Omega \rightarrow \Lambda,$$

by the requirement

$$\begin{aligned} \tilde{X}_t(\zeta(\omega)) &= X_t(\omega) = x_t \\ \tilde{Y}_t(\zeta(\omega)) &= Y_t(\omega) = y_t, \end{aligned}$$

for $\omega \in \Omega$ and $t \in \mathbf{N}$. Clearly,

$$\begin{aligned} &\zeta^{-1}\left\{\lambda \in \Lambda : (\tilde{X}_{t_1}(\lambda), \dots, \tilde{X}_{t_k}(\lambda)) \in A, (\tilde{Y}_{t_1}(\lambda), \dots, \tilde{Y}_{t_k}(\lambda)) \in B\right\} \\ &= \left\{\omega \in \Omega : (\tilde{X}_{t_1}(\zeta(\omega)), \dots, \tilde{X}_{t_k}(\zeta(\omega))) \in A, (\tilde{Y}_{t_1}(\zeta(\omega)), \dots, \tilde{Y}_{t_k}(\zeta(\omega))) \in B\right\} \\ &= \left\{\omega \in \Omega : (X_{t_1}(\omega), \dots, X_{t_k}(\omega)) \in A, (Y_{t_1}(\omega), \dots, Y_{t_k}(\omega)) \in B\right\} \\ &\in \mathcal{F}, \end{aligned} \quad (2.2)$$

if $A \in \mathcal{S}_k$ and $B \in \mathcal{B}_k$. Thus ζ is measurable.

Define probability measure $\tilde{P} = P\zeta^{-1}$ on \mathcal{B}_Λ , then from (2.1) and (2.2),

$$\begin{aligned} &\tilde{P}\left\{\lambda \in \Lambda : (\tilde{X}_{t_1}(\lambda), \dots, \tilde{X}_{t_k}(\lambda)) \in A, (\tilde{Y}_{t_1}(\lambda), \dots, \tilde{Y}_{t_k}(\lambda)) \in B\right\} \\ &= P\zeta^{-1}\tilde{P}\left\{\lambda \in \Lambda : (\tilde{X}_{t_1}(\lambda), \dots, \tilde{X}_{t_k}(\lambda)) \in A, (\tilde{Y}_{t_1}(\lambda), \dots, \tilde{Y}_{t_k}(\lambda)) \in B\right\} \\ &= P\left\{\omega \in \Omega : (X_{t_1}(\omega), \dots, X_{t_k}(\omega)) \in A, (Y_{t_1}(\omega), \dots, Y_{t_k}(\omega)) \in B\right\} \\ &= \nu_{t_1, \dots, t_k}(A \times B). \end{aligned} \quad (2.3)$$

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 ν_{t_1, \dots, 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 $t \in \mathbf{Z}$, unique up to equivalence, such that $\{(X_t, Y_t) : t \in \mathbf{N}\}$ and $\{(\bar{X}_t, \bar{Y}_t) : t \in \mathbf{N}\}$ are equivalent processes.*

Proof :

The proof follows from [1], page 21.

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

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

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

$$\begin{aligned} \mathcal{A}_{k'} &= \left\{ \lambda \in \Lambda : (Y_{-k'}(\lambda), \dots, Y_{k'}(\lambda)) \in B_{2k'} \right\} \\ &= \left\{ \lambda \in \Lambda : (y_{-k'}, \dots, y_{k'}) \in B_{2k'} \right\}, \end{aligned}$$

where $B_{2k'} \in \mathcal{B}_{2k'}$, that is the Borel σ -field of $\mathcal{Y}^{2k'}$, such that $\forall k \in \mathbf{N}$,

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

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

$$\begin{aligned} P(Y \in \mathcal{A} \Delta \mathcal{A}_{k'}) &= P(T^{2k'}Y \in \mathcal{A} \Delta \mathcal{A}_{k'}) \\ &= P(Y \in \mathcal{A} \Delta T^{-2k'}\mathcal{A}_{k'}) \\ &= P(Y \in \mathcal{A} \Delta \tilde{\mathcal{A}}_{k'}), \end{aligned} \quad (3.2)$$

where

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

Let

$$\tilde{\mathcal{A}} = \bigcap_{k \in \mathbf{N}} \bigcup_{j \geq k} \tilde{\mathcal{A}}_{j'},$$

then

$$\begin{aligned} \mathcal{A}^c \cap \tilde{\mathcal{A}} &= \mathcal{A}^c \cap \left(\bigcap_{k \geq 1} \bigcup_{j \geq k} \tilde{\mathcal{A}}_{j'} \right) \\ &= \bigcap_{k \geq 1} \bigcup_{j \geq k} (\mathcal{A}^c \cap \tilde{\mathcal{A}}_{j'}) \\ &= \limsup_{k' \rightarrow \infty} \mathcal{A}^c \cap \tilde{\mathcal{A}}_{k'} \end{aligned}$$

and

$$\begin{aligned} \mathcal{A} \cap \tilde{\mathcal{A}}^c &= \mathcal{A} \cap \left(\bigcap_{k \geq 1} \bigcup_{j \geq k} \tilde{\mathcal{A}}_{j'} \right)^c \\ &= \mathcal{A} \cap \left(\bigcup_{k \geq 1} \bigcap_{j \geq k} \tilde{\mathcal{A}}_{j'}^c \right) \\ &= \bigcup_{k \geq 1} \bigcap_{j \geq k} (\mathcal{A} \cap \tilde{\mathcal{A}}_{j'}^c) \\ &= \liminf_{k' \rightarrow \infty} \mathcal{A} \cap \tilde{\mathcal{A}}_{k'}^c. \end{aligned}$$

Hence,

$$\begin{aligned}
\mathcal{A} \Delta \tilde{\mathcal{A}} &= (\mathcal{A} \cap \tilde{\mathcal{A}}^c) \cup (\mathcal{A}^c \cap \tilde{\mathcal{A}}) \\
&= \left(\liminf_{k' \rightarrow \infty} \mathcal{A} \cap \tilde{\mathcal{A}}_{k'}^c \right) \cup \left(\limsup_{k' \rightarrow \infty} \mathcal{A}^c \cap \tilde{\mathcal{A}}_{k'} \right) \\
&\subset \left(\limsup_{k' \rightarrow \infty} \mathcal{A} \cap \tilde{\mathcal{A}}_{k'}^c \right) \cup \left(\limsup_{k' \rightarrow \infty} \mathcal{A}^c \cap \tilde{\mathcal{A}}_{k'} \right) \\
&= \limsup_{k' \rightarrow \infty} \left((\mathcal{A} \cap \tilde{\mathcal{A}}_{k'}^c) \cup (\mathcal{A}^c \cap \tilde{\mathcal{A}}_{k'}) \right) \\
&= \limsup_{k' \rightarrow \infty} (\mathcal{A} \Delta \tilde{\mathcal{A}}_{k'}). \tag{3.3}
\end{aligned}$$

From (3.1) and (3.2)

$$\begin{aligned}
\sum_k P(Y \in \mathcal{A} \Delta \tilde{\mathcal{A}}_{k'}) &= \sum_k P(Y \in \mathcal{A} \Delta \mathcal{A}_{k'}) \\
&\leq \sum_k 2^{-k} \\
&= 1,
\end{aligned}$$

then by Borrel Cantelli's Lemma,

$$0 \leq P(Y \in \mathcal{A} \Delta \tilde{\mathcal{A}}) \leq P(Y \in \limsup \mathcal{A} \Delta \tilde{\mathcal{A}}_{k'}) = 0,$$

implying

$$P(\mathcal{A} \Delta \tilde{\mathcal{A}}) = 0. \tag{3.4}$$

Since (3.4) holds, then to prove $P(Y \in \mathcal{A}) = 0$ or 1 , is equivalent with showing that

$$P(Y \in \tilde{\mathcal{A}}) = 0 \text{ or } 1.$$

By definition, $\tilde{\mathcal{A}} = \bigcap_{k \geq 1} \bigcup_{j \geq k} \tilde{\mathcal{A}}_{j'}$, so $\tilde{\mathcal{A}}$ is in the tail σ -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

$$P(Y \in \tilde{\mathcal{A}}|x) = 0 \text{ or } 1.$$

Let

$$B = \{x : P(Y \in \tilde{\mathcal{A}}|x) = 1\},$$

so

$$\begin{aligned}
P(Y \in \tilde{\mathcal{A}}) &= E [1_{Y \in \tilde{\mathcal{A}}}] \\
&= E [E[1_{Y \in \tilde{\mathcal{A}}}|x]] \\
&= E [P(Y \in \tilde{\mathcal{A}}|x)] \\
&= 0 + 1 \cdot P(X \in B) \\
&= P(X \in B). \tag{3.5}
\end{aligned}$$

But, B is invariant, as

$$\begin{aligned} P(Y \in \tilde{\mathcal{A}}|x) &= P(TY \in \tilde{\mathcal{A}}|Tx) \\ &= P(Y \in \tilde{\mathcal{A}}|Tx). \end{aligned}$$

Since the Markov chain $\{X_t\}$ is stationary and irreducible, then $\{X_t\}$ is ergodic, implying

$$P(X \in B) = 0 \quad \text{or} \quad 1.$$

Hence, by (3.5),

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

■

REFERENCES

- [1] Azencott, R. and Dacunha-Castelle, D. 1986. *Series of irregular observation: forecasting and model building*. Applied Probability. Springer Verlag, New York.
- [2] Billingsley, P. 1986. *Probability and Measure*. John Wiley, New York.
- [3] Elliott, R. J, Aggoun, L. and Moore J.B. 1993. *Hidden Markov models: estimation and control*, Volume 29 of Application of Mathematics. Springer Verlag, New York.
- [4] Leroux, B.G. 1992. Maximum likelihood estimation for hidden Markov models. *Stochastic processes and their applications*, 40: 127-143.
- [5] Ryden, T. 1996. On recursive estimation for hidden Markov models. *Stochastic Processes and their Applications*, 66, 79-96.
- [6] Setiawaty, B. 2002. Equivalent representations of hidden Markov models. *Journal of Mathematics and its Applications*, Volume 2, No.1, 1-12.
- [7] Walters, P. 1982. *An Introduction to ergodic Theory*. Springer Verlag, New York.