CHARACTERISTICS OF A TRUE PARAMETER
OF A HIDDEN MARKOV MODEL

BERLIAN SETIAWATY

Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University
Jln. Raya Pajajaran, Kampus IPB Baranangsiang,
Bogor, 16143 Indonesia

Abstract. Representation which generates the observed process
of a hidden Markov model is not unique. The simplest one, that is,
the one with minimum size is called a true parameter. This article
is aimed to present characteristics of this parameter.
Key words: Hidden Markov, representations, true parameter.

1. Introduction

According to [3], representation for a hidden Markov model is not
unique. Our main interest is to find the simplest one, that is, the one
with minimum size. Such representation will be called a true param-
eter. Our task is to identify a true parameter and its size. Therefore,
the main aim of this article is to collect facts concerning the true pa-
rameter.

For this purpose, we begin with definition of a hidden Markov model,
representations and equivalent representations in the first section. The
second section will present definition of a true parameter of a hidden
Markov model and its characteristics.

2. A hidden Markov model and its representations

Let \( \{ X_t : t \in \mathbb{N} \} \) be a finite state Markov chain defined on a prob-
ability 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 [1], 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 \(S \times Y\), 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.

Since the Markov chain \(\{X_t\}\) in a hidden Markov model \(\{(X_t, Y_t)\}\) is not observable, then inference concerning the hidden Markov model has to be based on the information of \(\{Y_t\}\) alone. By knowing the finite dimensional joint distributions of \(\{Y_t\}\), parameters which characterize the hidden Markov model can then be analysed.

From [3], it can be seen that the law of the hidden Markov model \(\{(X_t, Y_t)\}\) is completely specified by:

(a). The size \(K\).

(b). The transition probability matrix \(A = (\alpha_{ij})\), satisfying

\[
\alpha_{ij} \geq 0, \quad \sum_{j=1}^{K} \alpha_{ij} = 1, \quad i, j = 1, \ldots, K.
\]

(c). The initial probability distribution \(\pi = (\pi_i)\) satisfying

\[
\pi_i \geq 0, \quad i = 1, \ldots, K, \quad \sum_{i=1}^{K} \pi_i = 1.
\]

(d). The vector \(\theta = (\theta_i)^T, \theta_i \in \Theta, i = 1, \ldots, K\), which describes the conditional densities of \(Y_t\) given \(X_t = i, i = 1, \ldots, K\).

**Definition 2.2.** Let

\[
\phi = (K, A, \pi, \theta).
\]

The parameter \(\phi\) is called a **representation** of the hidden Markov model \(\{(X_t, Y_t)\}\).

Thus, the hidden Markov model \(\{(X_t, Y_t)\}\) can be represented by a representation \(\phi = (K, A, \pi, \theta)\).

On the other hand, we can also generate a hidden Markov model \(\{(X_t, Y_t)\}\) from a representation \(\phi = (K, A, \pi, \theta)\), by choosing a Markov
chain \( \{X_t\} \) which takes values on \( \{1, \ldots, K\} \) and its law is determined by the \( K \times K \)-transition probability matrix \( A \) and the initial probability \( \pi \), and an observation process \( \{Y_t\} \) taking values on \( \mathcal{Y} \), where the density functions of \( Y_t \) given \( X_t = i, i = 1, \ldots, K \) are determined by \( \theta \).

Let \( \phi = (K, A, \pi, \theta) \) and \( \hat{\phi} = (\hat{K}, \hat{A}, \hat{\pi}, \hat{\theta}) \) be two representations which respectively generate hidden Markov models \( \{(X_t, Y_t)\} \) and \( \{\hat{(X)_t, Y_t}\} \). The \( \{(X_t, Y_t)\} \) takes values on \( \{1, \ldots, K\} \times \mathcal{Y} \) and \( \{\hat{(X)_t, Y_t}\} \) takes values on \( \{1, \ldots, \hat{K}\} \times \mathcal{Y} \). For any \( n \in \mathbb{N} \), let \( p_\phi(\cdot, \ldots, \cdot) \) and \( p_\hat{\phi}(\cdot, \ldots, \cdot) \) be the \( n \)-dimensional joint density function of \( Y_1, \ldots, Y_n \) with respect to \( \phi \) and \( \hat{\phi} \). Suppose that for every \( n \in \mathbb{N} \),

\[
p_\phi(Y_1, \ldots, Y_n) = p_\hat{\phi}(Y_1, \ldots, Y_n).
\]

Then \( \{Y_t\} \) has the same law under \( \phi \) and \( \hat{\phi} \). Since in hidden Markov models \( \{(X_t, Y_t)\} \) and \( \{\hat{(X)_t, Y_t}\} \), the Markov chains \( \{X_t\} \) and \( \{\hat{X}_t\} \) are not observable and we only observed the values of \( \{Y_t\} \), then theoretically, the hidden Markov models \( \{(X_t, Y_t)\} \) and \( \{\hat{(X)_t, Y_t}\} \) are indistinguishable. In this case, it is said that \( \{(X_t, Y_t)\} \) and \( \{\hat{(X)_t, Y_t}\} \) are equivalent. The representations \( \phi \) and \( \hat{\phi} \) are also said to be equivalent, and will be denoted as \( \phi \sim \hat{\phi} \).

For each \( K \in \mathbb{N} \), define

\[
\Phi_K = \left\{ \phi : \phi = (K, A, \pi, \theta), \text{ where } A, \pi \text{ and } \theta \text{ satisfy :} \right. \\
A = (\alpha_{ij}), \quad \alpha_{ij} \geq 0, \quad \sum_{j=1}^{K} \alpha_{ij} = 1, \quad i, j = 1, \ldots, K \\
\pi = (\pi_i), \quad \pi_i \geq 0, \quad i = 1, \ldots, K, \quad \sum_{i=1}^{K} \pi_i = 1 \\
\theta = (\theta_i)^T, \quad \theta_i \in \Theta, \quad i = 1, \ldots, K \left. \right\} \tag{2.1}
\]

and

\[
\Phi = \bigcup_{K \in \mathbb{N}} \Phi_K. \tag{2.2}
\]

The relation \( \sim \) is now defined on \( \Phi \) as follows.

**Definition 2.3.** Let \( \phi, \hat{\phi} \in \Phi \). Representations \( \phi \) and \( \hat{\phi} \) are said to be equivalent, denoted as

\[
\phi \sim \hat{\phi}
\]

if and only if for every \( n \in \mathbb{N} \),

\[
p_\phi(Y_1, Y_2, \ldots, Y_n) = p_\hat{\phi}(Y_1, Y_2, \ldots, Y_n).
\]
Remarks 2.4. It is clear that relation \( \sim \) forms an equivalence relation on \( \Phi \).

Let \( \phi = (K, A, \pi, \theta) \in \Phi_K \), then under \( \phi \), \( Y_1, \ldots, Y_n \), for any \( n \), has joint density

\[
p_\phi(y_1, \ldots, y_n) = \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \pi_{x_1} f(y_1, \theta_{x_1}) \cdot \prod_{t=2}^{n} \alpha_{x_{t-1}, x_t} f(y_t, \theta_{x_t}). \tag{2.3}
\]

Let \( \sigma \) be any permutation of \( \{1, 2, \ldots, K\} \). Define

\[
\sigma(A) = (\alpha_{\sigma(i), \sigma(j)}) \\
\sigma(\pi) = (\pi_{\sigma(i)}) \\
\sigma(\theta) = (\theta_{\sigma(i)})^T.
\]

Let

\[
\sigma(\phi) = (K, \sigma(A), \sigma(\pi), \sigma(\theta)),
\]

then \( \sigma(\phi) \in \Phi_K \) and easy to see from (2.3) that

\[
p_\phi(y_1, \ldots, y_n) = p_{\sigma(\phi)}(y_1, \ldots, y_n).
\]

implying \( \phi \sim \sigma(\phi) \). So we have the following lemma.

**Lemma 2.5.** Let \( \phi \in \Phi_K \), then for every permutation \( \sigma \) of \( \{1, 2, \ldots, K\} \),

\[
\sigma(\phi) \sim \phi.
\]

from [3], we have the following lemmas.

**Lemma 2.6.** Let \( \phi = (K, A, \pi, \theta) \in \Phi_K \), where \( \pi \) is a stationary probability distribution of \( A \). Let \( N \) be the number of non-zero \( \pi_i \). Then there is \( \hat{\phi} = (N, \hat{A}, \hat{\pi}, \hat{\theta}) \in \Phi_N \), such that :

1. \( \hat{\pi}_i > 0 \), for \( i = 1, \ldots, N \).
2. \( \hat{\pi} \) is a stationary probability distribution of \( \hat{A} \).
3. \( \phi \sim \hat{\phi} \).

**Lemma 2.7.** For any \( K \in \mathbb{N} \) and \( \phi \in \Phi_K \), there is \( \hat{\phi} \in \Phi_{K+1} \), such that \( \phi \sim \hat{\phi} \).

By Lemma 2.7, we can define an order \( \prec \) in \( \{\Phi_K\} \).

**Definition 2.8.** Define an order \( \prec \) on \( \{\Phi_K\} \) by

\[
\Phi_K \prec \Phi_L, \quad K, L \in \mathbb{N},
\]

if and only if for every \( \phi \in \Phi_K \), there is \( \hat{\phi} \in \Phi_L \) such that \( \phi \sim \hat{\phi} \).
As a consequence of Lemma 2.7, Lemma 2.9 follows.

**Lemma 2.9.** For every $K \in \mathbb{N}$,
\[ \Phi_K \prec \Phi_{K+1} \]

From Lemma 2.9, the families of hidden Markov models represented by $\{\Phi_K\}$ are *nested families*.

3. A true parameter and its characteristics

We begin this section with a formal definition of a true parameter.

**Definition 3.1.** Let $\{(X_t, Y_t)\}$ be a hidden Markov model with representation $\phi \in \Phi$. A representation $\phi^o = (K^o, A^o, \pi^o, \theta^o) \in \Phi$, is called a *true parameter* of the hidden Markov model $\{(X_t, Y_t)\}$ if and only if
1. $\phi^o \sim \phi$.
2. $K^o$ is minimum, that is, there is no $\hat{\phi} \in \Phi_K$, with $K < K^o$, such that $\hat{\phi} \sim \phi^o$.

A true parameter $\phi^o = (K^o, A^o, \pi^o, \theta^o)$ of a hidden Markov model $\{(X_t, Y_t)\}$ is not unique, by Lemma 2.5, for every permutation $\sigma$ of $\{1, \ldots, K^o\}$,
\[ \sigma(\phi^o) \sim \phi^o. \]
So $\sigma(\phi^o)$ is also a true parameter of the hidden Markov model $\{(X_t, Y_t)\}$.

As a straight consequence of Definition 3.1, we have the following lemma.

**Lemma 3.2.** Let $\phi^o = (K^o, A^o, \pi^o, \theta^o)$ be a true parameter of a hidden Markov model $\{(X_t, Y_t)\}$. If $\pi^o_i$ is a stationary probability distribution of $A^o$, then $\pi^o_i > 0$, for $i = 1, \ldots, K^o$.

**Proof:**
Let $N^o$ be the number of non-zero $\pi^o_i$’s, then $1 \leq N^o \leq K$. If $N^o < K^o$, then by Lemma 2.6, there is $\phi = (N^o, A, \pi, \theta) \in \Phi_{N^o}$, such that $\phi \sim \phi^o$, contradicting with Lemma 3.2. Thus, it must be $N^o = K^o$. 

\[ \square \]
Lemma 3.4. Let \( \phi^o = (K^o, A^o, \pi^o, \theta^o) \) be a true parameter of a hidden Markov model \( \{ (X_t, Y_t) \} \), where \( \pi^o \) is a stationary probability distribution of \( A^o \). Let \( \phi = (K, A, \pi, \theta) \in \Phi_K \), where \( \phi \sim \phi^o \) and \( N \) be the number of non-zero \( \pi_i \).

1. If \( K = K^o \), then \( N = K^o \).
2. If \( K > K^o \), then \( N \geq K^o \).

Proof:
Let \( \phi = (K, A, \pi, \theta) \in \Phi_K \), where \( \phi \sim \phi^o \). By Lemma 3.2,
\[ K \geq K^o. \]

Let \( N \) be the number of non-zero \( \pi_i \), then
\[ 1 \leq N \leq K. \]
Suppose that \( N < K^o \), since \( \phi \sim \phi^o \), then \( \pi \) is a stationary probability distribution of \( A \). By Lemma 2.6, there is \( \hat{\phi} = (N, \hat{A}, \hat{\pi}, \hat{\theta}) \in \Phi_N \), such that \( \phi \sim \hat{\phi} \), implying \( \hat{\phi} \sim \phi^o \), contradicting with Lemma 3.2. Thus, it must be
\[ K^o \leq N \leq K. \quad (3.1) \]
If \( K = K^o \), then by (3.1), \( N = K^o \). If \( K > K^o \), then \( N \geq K^o \). \[ \blacksquare \]

Corollary 1. let \( \phi^o = (K^o, A^o, \pi^o, \theta^o) \) be a true parameter of a hidden Markov model \( \{ (X_t, Y_t) \} \), where \( \pi^o \) is a stationary probability distribution of \( A^o \). Let \( \phi = (K^o, A, \pi, \theta) \in \Phi_{K^o} \). If \( \phi \sim \phi^o \), then
\[ \pi_i > 0, \quad \text{for} \ i = 1, \ldots, K^o. \]

Proof:
This is part (a) of Lemma 3.4.

References

