A HIDDEN MARKOV MODEL: DEPENDENCIES BETWEEN RANDOM VARIABLES AND ITS REPRESENTATION

BERLIAN SETIAWATY

Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University
Jl. Meranti, Kampus IPB Darmaga, Bogor, 16680 Indonesia

Abstract. This article shows the nature of dependencies between random variables in a hidden Markov model. Using these properties,we will show that the law of a hidden Markov model is completely specified by a set of four parameters which is called a representation of the hidden Markov model.
Key words: Hidden Markov, conditional joint distributions, representations.

1. Introduction

The purposes of this article are to introduce hidden Markov models and to show the nature of dependencies between the random variables in a hidden Markov model. Based on these, the finite dimensional joint distributions of the observed process are derived. So the parameters which characterize the model can be analysed. Such parameters will be referred to as a *representations* of the model.

Later, we will show that a hidden markov model can be represented by a representation,and on the other hand, a representation can be used also to generate a hidden Markov model.

A hidden Markov model is formally defined in section 2 and an example is also given in this section. In section 3, the nature of dependencies between random variables in a hidden Markov model is discussed. We show four parameters which specify completely the law of a hidden Markov model in section 4.

2. Hidden Markov Models

Let $\{X_t : t \in \boldsymbol{N}\}$ be a finite state Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\{X_t\}$ is not observed directly, but rather there is an *observation* process $\{Y_t : t \in \boldsymbol{N}\}$ defined on $(\Omega, \mathcal{F}, P)$. Then consequently, the Markov chain is said to be *hidden* in the observations. A pair of stochastic processes $\{(X_t, Y_t) : t \in \boldsymbol{N}\}$ 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 $\{(X_t, Y_t) : t \in \boldsymbol{N}\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in a set $\boldsymbol{S} \times \mathcal{Y}$, 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 $\boldsymbol{S}$, will be called the* ***size*** *of the hidden Markov model.*

The following is an example of a hidden Markov model which is adapted from [2].

Example 2.2. Let $\{X_t\}$ be a Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values on $\boldsymbol{S} = \{1, \dots, K\}$. The observed process $\{Y_t\}$ is defined by

$$Y_t = c(X_t) + \sigma(X_t)\omega_t, \qquad t \in \boldsymbol{N}, \tag{2.1}$$

where c and σ are real valued functions and positive real valued function on $\boldsymbol{S}$ respectively, and $\{\omega_t\}$ is a sequence of $N(0,1)$ independent, identically distributed (i.i.d.) random variables.

Since $\{\omega_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables, then given $\{X_t\}$, $\{Y_t\}$ is a sequence of independent random variables. From (2.1), it is clear that Y_t is a function of X_t only, then the third condition of Definition 2.1 holds. Let $y \in \mathcal{Y}$ and $i \in \boldsymbol{S}$. Let $c_i = c(i)$ and $\sigma_i = \sigma(i)$, then

$$\begin{aligned} P(Y_t \le y | X_t = i) &= P(c_i + \sigma_i \omega_t \le y) \\ &= P(\sigma_i \omega_t \le y - c_i) \\ &= \int_{-\infty}^{y-c_i} \phi_i(z)\, dz, \end{aligned} \tag{2.2}$$

where

$$\phi_i(z) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z}{\sigma_i}\right)^2}. \tag{2.3}$$

Thus from (2.2) and (2.3), the conditional density of Y_t given $X_t = i$ is $\phi_i(\cdot - c_i)$ which does not depend on t. Therefore it can be concluded that $\{(X_t, Y_t)\}$ is a hidden Markov model.

3. Dependencies between Random Variables

This section shows the nature of dependencies between the random variables in a hidden Markov model.

Let $\{(X_t, Y_t)\}$ be a hidden Markov model defined on a probability space $(\Omega, \mathcal{F}, P)$, where the Markov chain $\{X_t\}$ taking values in a set $\boldsymbol{S} = \{1, \dots, K\}$ and the observed process $\{Y_t\}$ taking values on $\mathcal{Y}$. Throughout the thesis, we will assume that Y_t is scalar valued and without loss of generality, we will suppose that $\mathcal{Y} = \boldsymbol{R}$. The generalization to vector case is straight forward.

Assume that the conditional density of Y_t given $X_t = i$, for all $t \in \boldsymbol{N}$ and $i = 1, \dots, K$ are dominated by a σ-finite measure μ. The conditional density of Y_t given $X_t = i$, with respect to μ, will be denoted by $p(\cdot|i)$. This means that for all $t \in \boldsymbol{N}$ and $i = 1, \dots, K$,

$$P(Y_t \le y | X_t = i) = \int_{-\infty}^{y} p(z|i)\, d\mu(z).$$

Notation 3.1. Here and in the sequel, p will be used as s generic symbol for a probability density function. If there is no confusion, for random variables U and V defined on $(\Omega, \mathcal{F}, P)$, the joint density function of U and V, $p_{U,V}(\cdot,\cdot)$ will be denoted by $p(\cdot,\cdot)$ and the conditional density function of U given V, $p_{U|V}(\cdot|\cdot)$ will simply be denoted by $p(\cdot|\cdot)$.

Let U and V be any random variables defined on $(\Omega, \mathcal{F}, P)$. Notice that the joint density function of U and V and the conditional density function of U given V can be expressed as

$$\begin{aligned} p(u,v) &= p(U(\omega),V(\omega)) \\ &= p(U,V)(\omega) \\ p(u|v) &= p(U(\omega)|V(\omega)) \\ &= p(U|V)(\omega), \end{aligned}$$

where $U(\omega) = u$ and $V(\omega)$, for some $\omega \in \Omega$.

First we prove some general rules for conditional densities.

Lemma 3.2. *Let U, V and W be any random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, then*
(a). $p(U|V,W) = \dfrac{p(U|V)\cdot p(W|U,V)}{p(W|V)}$.
(b). $p(U|V,W) = \dfrac{p(U,V|W)}{p(V|W)}$.
(c). $p(U,V|W) = p(V|W)\cdot p(U|V,W)$.

Proof :
The conditional probability density function of U given V is defined by

$$p(u|v) = \frac{p(u,v)}{p(v)}, \tag{3.1}$$

for all u and for all v such that $p(v) > 0$. By equation (3.1), we have

$$p(U|V) = \frac{p(U,V)}{p(V)}. \tag{3.2}$$

Analog with (3.2),

$$p(W|U,V) = \frac{p(U,V,W)}{p(U,V)} \tag{3.3}$$

$$p(U|V,W) = \frac{p(U,V,W)}{p(V,W)}. \tag{3.4}$$

By equations (3.2), (3.3) and (3.4),

$$\begin{aligned} \frac{p(U|V)\cdot p(W|U,V)}{p(W|V)} &= \frac{\dfrac{p(U,V)}{p(V)}\cdot\dfrac{p(U,V,W)}{p(U,V)}}{\dfrac{p(V,W)}{p(V)}} \\ &= \frac{p(U,V,W)}{p(V,W)} \\ &= p(U|V,W), \end{aligned}$$

$$\begin{aligned} \frac{p(U,V|W)}{p(V|W)} &= \frac{\dfrac{p(U,V,W)}{p(W)}}{\dfrac{p(V,W)}{p(W)}} \\ &= \frac{p(U,V,W)}{p(V,W)} \\ &= p(U|V,W), \end{aligned}$$

and

$$\begin{aligned} p(V|W) \cdot p(U|V,W) &= \frac{p(V,W)}{p(W)} \cdot \frac{p(U,V,W)}{p(V,W)} \\ &= \frac{p(U,V,W)}{p(W)} \\ &= p(U,V|W). \end{aligned}$$

So the lemma is proved. ■

Using the general rules from Lemma 3.2, we prove the following lemmas which describes the nature of dependencies between random variables in the hidden Markov model.

Notation 3.3. For convinience, sometimes $X_m, \ldots, X_n$ and its realizations $x_m, \ldots, x_n$ will be abbreviated X_m^n and x_m^n respectively. Similar notations are also applied for the $\{Y_t\}$ process and its realizations.

Lemma 3.4. *Let* $1 \le m \le t < n$.
(a). $p(X_{t+1}, Y_{t+1}|X_m^t, Y_m^t) = p(X_{t+1}, Y_{t+1}|X_t)$.
(b). $p(X_t, Y_t|X_{t+1}^n, Y_{t+1}^n) = p(X_t, Y_t|X_{t+1})$.

Proof :
By the third part of Lemma 3.2,

$$p(X_{t+1}, Y_{t+1}|X_m^t, Y_m^t) = p(X_{t+1}|X_m^t, Y_m^t) \cdot p(Y_{t+1}|X_m^{t+1}, Y_m^t). \tag{3.5}$$

By the first part of Lemma 3.2 and the Markov property,

$$\begin{aligned} p(X_{t+1}|X_m^t, Y_m^t) &= \frac{p(X_{t+1}|X_m^t) \cdot p(Y_m^t|X_m^{t+1})}{p(Y_m^t|X_m^t)} \\ &= \frac{p(X_{t+1}|X_t) \cdot p(Y_m^t|X_m^{t+1})}{p(Y_m^t|X_m^t)}. \end{aligned} \tag{3.6}$$

Also by the first part of Lemma 3.2 and condition (c) of Definition 2.1,

$$\begin{aligned} p(Y_{t+1}|X_m^{t+1}, Y_m^t) &= \frac{p(Y_{t+1}|X_m^{t+1}) \cdot p(Y_m^t|X_m^{t+1}, Y_{t+1})}{p(Y_m^t|X_m^{t+1})} \\ &= \frac{p(Y_{t+1}|X_{t+1}) \cdot p(X_m^{t+1}, Y_m^{t+1})}{p(Y_m^t|X_m^{t+1}) \cdot p(X_m^{t+1}, Y_{t+1})} \\ &= \frac{p(Y_{t+1}|X_{t+1}) \cdot p(Y_m^{t+1}|X_m^{t+1})}{p(Y_m^t|X_m^{t+1}) \cdot p(Y_{t+1}|X_m^{t+1})} \\ &= \frac{p(Y_{t+1}|X_{t+1}) \cdot p(Y_m^{t+1}|X_m^{t+1})}{p(Y_m^t|X_m^{t+1}) \cdot p(Y_{t+1}|X_{t+1})} \\ &= \frac{p(Y_m^{t+1}|X_m^{t+1})}{p(Y_m^t|X_m^{t+1})}. \end{aligned} \tag{3.7}$$

From (3.5), (3.6), (3.7) and conditions (b) and (c) of Definition 2.1,

$$\begin{aligned} p(X_{t+1}|X_m^t, Y_m^t) &= \frac{p(X_{t+1}|X_t) \cdot p(Y_m^{t+1}|X_m^{t+1})}{p(Y_m^t|X_m^t)} \\ &= p(X_{t+1}|X_t) \cdot p(Y_{t+1}|X_{t+1}) \\ &= p(X_{t+1}, Y_{t+1}|X_t). \end{aligned}$$

The proof for (b) is similar using the first part of Lemma 3.2, the Markov property and conditions (b) and (c) of Definition 2.1. ■

Corollary 3.5. *Let* $1 \leq m < t < n$.
(a). $p(X_{t+1}|X_m^t, Y_m^t) = p(X_{t+1}|X_t)$.
(b). $p(Y_{t+1}|X_m^t, Y_m^t) = p(Y_{t+1}|X_t)$.
(c). $p(X_t|X_{t+1}^n, Y_{t+1}^n) = p(X_t|X_{t+1})$.
(d). $p(Y_t|X_{t+1}^n, Y_{t+1}^n) = p(Y_t|X_{t+1})$.

Proof :
For (a), using the first part of Lemma 3.4,

$$\begin{aligned} p(x_{t+1}|x_m^t, y_m^t) &= \int_{-\infty}^{\infty} p(x_{t+1}, y_{t+1}|x_m^t, y_m^t)\, d\mu(y_{t+1}) \\ &= \int_{-\infty}^{\infty} p(x_{t+1}, y_{t+1}|x_t)\, d\mu(y_{t+1}) \\ &= p(x_{t+1}|x_t), \end{aligned} \tag{3.8}$$

which gives

$$p(X_{t+1}|X_m^t, Y_m^t) = p(X_{t+1}|X_t).$$

The proofs for (b), (c) and (d) are similar using Lemma 3.4. ■

Lemma 3.6. *Let* $1 \leq m < t < n$.
(a). $p(X_{t+1}, Y_{t+1}|X_m^t, Y_{m+1}^t) = p(X_{t+1}, Y_{t+1}|X_t)$.
(b). $p(X_t, Y_t|X_{t+1}^n, Y_{t+1}^{n-1}) = p(X_t, Y_t|X_{t+1})$.

Proof :
For (a), by the first part of Lemma 3.2, Lemma 3.4 and the third part of Corollary 3.5,

$$\begin{aligned} p(X_{t+1}, Y_{t+1}|X_m^t, Y_{m+1}^t) &= p(X_{t+1}, Y_{t+1}|X_m, X_{m+1}^t, Y_{m+1}^t, X_m) \\ &= \frac{p(X_{t+1}, Y_{t+1}|X_{m+1}^t, Y_{m+1}^t) \cdot p(X_m|X_{m+1}^{t+1}, Y_{m+1}^{t+1})}{p(X_m|X_{m+1}^t, Y_{m+1}^t)} \\ &= \frac{p(X_{t+1}, Y_{t+1}|X_t) \cdot p(X_m|X_{m+1})}{p(X_m|X_{m+1})} \\ &= p(X_{t+1}, Y_{t+1}|X_t). \end{aligned}$$

The proof for (b) is similar using the first part of Lemma 3.2, Lemma 3.4 and Corollary 3.5. ■

Lemma 3.7. *Let* $1 \leq m \leq t < n$.
(a). $p(X_{t+1}^n, Y_{t+1}^n|X_m^t, Y_m^t) = p(X_{t+1}^n, Y_{t+1}^n|X_t)$.
(b). $p(X_m^t, Y_m^t|X_{t+1}^n, Y_{t+1}^n) = p(X_m^t, Y_m^t|X_{t+1})$.

Proof :
For (a), using the third part of Lemma 3.2 and the first parts of Lemma 3.4 and Lemma 3.6,

$$\begin{aligned} & p(X_{t+1}^n, Y_{t+1}^n) \\ &\quad= p(X_{t+1}, Y_{t+1}|X_m^t, Y_m^t) p(X_{t+2}, Y_{t+2}|X_m^{t+1}, Y_m^{t+1}) \cdots p(X_n, Y_n|X_m^{n-1}, Y_m^{n-1}) \\ &\quad= p(X_{t+1}, Y_{t+1}|X_t) p(X_{t+2}, Y_{t+2}|X_{t+1}) \cdots p(X_n, Y_n|X_{n-1}) \\ &\quad= p(X_{t+1}, Y_{t+1}|X_t) p(X_{t+2}, Y_{t+2}|X_t^{t+1}, Y_{t+1}) \cdots p(X_n, Y_n|X_t^{n-1}, Y_{t+1}^{n-1}) \\ &\quad= p(X_{t+1}^n, Y_{t+1}^n|X_t). \end{aligned}$$

The proof for (b) is similar using the third part of Lemma 3.2 and the second parts of Lemma 3.4 and Lemma 3.6. ■

Lemma 3.8. *Let* $1 \leq k, l \leq t < m, n$.
(a). $p(X_{t+1}^m, Y_{t+1}^n|X_k^t, Y_l^t) = p(X_{t+1}^m, Y_{t+1}^n|X_t)$.
(b). $p(X_k^t, Y_l^t|X_{t+1}^m, Y_{t+1}^n) = p(X_k^t, Y_l^t|X_{t+1})$.

Proof :
For (a), let $1 \leq k, l, \leq t < m, n$ and suppose that $k < l$ and $m \leq n$, then by the first part of Lemma 3.7

$$\begin{aligned}
& p(x_{t+1}^m, y_{t+1}^n | x_k^t, y_l^t) \\
&= \frac{p(x_k^t, x_{t+1}^m), y_l^t, y_{t+1}^n}{p(x_k^t, y_l^t)} \\
&= \frac{\sum_{x_{m+1}=1}^K \cdots \sum_{x_n=1}^K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_k^t, x_{t+1}^n, y_k^t, y_{t+1}^n) d\mu(y_k) \cdots d\mu(y_{l-1})}{p(x_k^t, y_l^t)} \\
&= \frac{\sum_{x_{m+1}=1}^K \cdots \sum_{x_n=1}^K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_{t+1}^n, y_{t+1}^n | x_k^t, y_k^t) p(x_k^t, y_k^t) d\mu(y_k) \cdots d\mu(y_{l-1})}{p(x_k^t, y_l^t)} \\
&= \frac{\sum_{x_{m+1}=1}^K \cdots \sum_{x_n=1}^K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_{t+1}^n, y_{t+1}^n | x_t) p(x_k^t, y_k^t) d\mu(y_k) \cdots d\mu(y_{l-1})}{p(x_k^t, y_l^t)} \\
&= \frac{p(x_{t+1}^m, y_{t+1}^n | x_t) p(x_k^t, y_l^t)}{p(x_k^t, y_l^t)} \\
&= p(x_{t+1}^m, y_{t+1}^n | x_t). \qquad (3.9)
\end{aligned}$$

The proofs for the other 3 possibilities of k and l are similar. So from (3.9), (a) follows.

The proof for (b) is similar using the second part of Lemma 3.7. ∎

Corollary 3.9. *Let* $1 \leq k, l \leq t < m, n$.
(a). $p(X_{t+1}^m | X_k^t, Y_l^t) = p(X_{t+1}^m | X_t)$.
(b). $p(Y_{t+1}^n | X_k^t, Y_l^t) = p(X_{t+1}^m | X_t)$.
(c). $p(X_k^t | X_{t+1}^m, Y_{t+1}^n) = p(X_k^t | X_{t+1})$.
(d). $p(Y_l^t | X_{t+1}^m, Y_{t+1}^n) = p(Y_l^t | X_{t+1})$.

Proof :
This lemma is a direct consequences of Lemma 3.8 which is obtained by integrating part (a) and (b) of Lemma 3.8 with respect to x and y. ∎

Corollary 3.10. *Let* $1 \leq k < t$, *then*

$$p(Y_t | X_t, Y_k^{t-1}) = p(Y_t | X_t).$$

Proof :
By the first part of Lemma 3.2 and the third part of Corollary 3.9,

$$\begin{aligned}
p(Y_t | X_t, Y_k^{t-1}) &= \frac{p(Y_t | X_t) \cdot p(Y_k^{t-1}) | X_t, Y_t)}{p(Y_k^{t-1} | X_t)} \\
&= \frac{p(Y_t | X_t) \cdot p(Y_k^{t-1}) | X_t)}{p(Y_k^{t-1} | X_t)} \\
&= p(Y_t | X_t).
\end{aligned}$$

∎

Lemma 3.11.

1. *If* $1 \leq k \leq l < t \leq m, n$, *then* $p(X_t^m | X_k^l, Y_1^n) = p(X_t^m | X_l, Y_{l+1}^n)$.
2. *If* $1 \leq l \leq t < m \leq n_1, n_2$, *then* $p(X_l^t | X_m^{n_1}, Y_1^{n_2}) = p(X_l^t | X_m, Y_1^{m-1})$.

Proof :
For(a), by the first part of Lemma 3.8, the second parts of Corollary 3.9 and Lemma

3.2,

$$
\begin{aligned}
p(x_t^m|x_k^l, y_1^n) &= \frac{p(x_k^l, x_t^m, y_1^n)}{p(x_k^l, y_1^n)} \\
&= \frac{p(x_k^l, x_t^m, y_1^l, y_{l+1}^n)}{p(x_k^l, y_1^l, y_{l+1}^n)} \\
&= \frac{\sum_{x_{l+1}=1}^K \cdots \sum_{x_{t-1}=1}^K p(x_k^l, x_{l+1}^m, y_1^l, y_{l+1}^n)}{p(x_k^l, y_1^l, y_{l+1}^n)} \\
&= \frac{\sum_{x_{l+1}=1}^K \cdots \sum_{x_{t-1}=1}^K p(x_{l+1}^m, y_{l+1}^n|x_k^l, y_1^l)}{p(y_{l+1}^n|x_k^l, y_1^l)} \\
&= \frac{\sum_{x_{l+1}=1}^K \cdots \sum_{x_{t-1}=1}^K p(x_{l+1}^m, y_{l+1}^n|x_l)}{p(y_{l+1}^n|x_l)} \\
&= \frac{p(x_t^m, y_{l+1}^n|x_l)}{p(y_{l+1}^n|x_l)} \\
&= p(x_t^m|x_l, y_{l+1}^n).
\end{aligned}
$$

Thus (a) follows.

The proof for (b) is similar, using the second part of Lemma 3.8, the last part of Corollary 3.9 and the second part of Lemma 3.2. ■

4. Representations of a Hidden Markov Model

The aim of this section is to find parameters which determine the characteristics of a 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.

Let $\{(X_t, Y_t)\}$ be a hidden Markov model defined on the probability space $(\Omega, \mathcal{F}, P)$, taking values on $\boldsymbol{S} \times \mathcal{Y}$, where $\boldsymbol{S} = \{1, \dots, K\}$ and $\mathcal{Y} = \boldsymbol{R}$. Let $A = (\alpha_{ij})$ be the transition probability matrix and $\pi = (\pi_i)$ be the initial probability distribution of the Markov chain $\{X_t\}$. Assume for $i = i, \dots, K$, the conditional densities of Y_t given $X_t = i$ with respect to the measure μ, $p(\cdot|i)$, belong to the same family $\boldsymbol{\mathcal{F}}$, where $\boldsymbol{\mathcal{F}} = \{f(\cdot|\theta) : \theta \in \Theta\}$ is a family of densities on a Euclidean space with respect to the measure μ, indexed by $\theta \in \Theta$. This means that for each $i = 1, \dots, K$,

$$p(\cdot|i) = f(\cdot, \theta_i),$$

for some $\theta_i \in \Theta$.

For $y \in \mathcal{Y}$ and $i, j = 1, \dots, K$, define

$$m_{ij}(y) = \alpha_{ij} \cdot f(y, \theta_j).$$

For every $y \in \mathcal{Y}$, let $M(y)$ be the $K \times K$-matrix defined by

$$M(y) = (m_{ij}(y)).$$

Then

$$M(y) = A \cdot B(y), \qquad y \in \mathcal{Y}, \tag{4.1}$$

where

$$B(y) = \begin{pmatrix} f(y,\theta_1) & 0 & 0 & \cdots & 0 \\ 0 & f(y,\theta_2) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f(y,\theta_K) \end{pmatrix}.$$

Observe that

$$\begin{aligned} \int_{-\infty}^{\infty} M(y)\, d\mu(y) &= \left(\int_{-\infty}^{\infty} m_{ij}(y)\, d\mu(y) \right) \\ &= \left(\int_{-\infty}^{\infty} \alpha_{ij} f(y,\theta_j)\, d\mu(y) \right) \\ &= (\alpha_{ij}) \\ &= A\,. \end{aligned} \tag{4.2}$$

Theorem 4.1. *For each $n \in \boldsymbol{N}$, the n-dimensional joint density fuction of $Y_1, Y_2, \ldots, Y_n$ is*

$$p(Y_1, Y_2, \ldots, Y_n) = \pi M(Y_1) M(Y_2) \cdots M(Y_n) e, \tag{4.3}$$

where $e = (1, 1, 1, \ldots, 1)^T$.

Proof :
By Lemma 3.2, Corollary 3.5, Lemma 3.6 and Lemma 3.7, the joint density function of $Y_1, Y_2, \ldots, Y_n$ can be expressed as,

$$\begin{aligned} p(y_1, y_2, \ldots, y_n) &= \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} p(x_1, y_1, x_2, y_2, \ldots, x_n, y_n) \\ &= \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \Big\{ p(x_1) \cdot p(y_1|x_1) \\ &\qquad \times p(x_2|x_1, y_1) \cdot p(y_2|x_1^2, y_1) \\ &\qquad \times \cdots \times p(x_n|x_1^{n-1}, y_1^{n-1}) \cdot p(y_n|x_1^n, y_1^{n-1}) \Big\} \\ &= \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \Big\{ p(x_1) \cdot p(y_1|x_1) \\ &\qquad \times p(x_2|x_1) \cdot p(y_2|x_2) \\ &\qquad \times \cdots \times p(x_n|x_{n-1}) \cdot p(y_n|x_n) \} \\ &= \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \Big\{ P(X_1 = x_1) \cdot f(y_1, \theta_{x_1}) \\ &\qquad \times P(X_2 = x_2 | X_1 = x_1) \cdot f(y_2, \theta_{x_2}) \\ &\qquad \times \cdots \times P(X_n = x_n | X_{n-1} = x_{n-1}) \cdot f(y_n, \theta_{x_n}) \Big\} \\ &= \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \pi_{x_1} \cdot f(y_1, \theta_{x_1}) \prod_{t=2}^{n} \alpha_{x_{t-1}, x_t} \cdot f(y_t, \theta_{x_t}) \\ &= \pi B(y_1) M(y_2) \cdots M(y_n) e, \end{aligned}$$

so the conclusion of the lemma follows. ■

Corollary 4.2. *If $\{X_t\}$ is a stationary Markov chain, then for each $n \in \boldsymbol{N}$, the n-dimensional joint density function of $Y_1, Y_2, \ldots, Y_n$ is*

$$p(Y_1, Y_2, \ldots, Y_n) = \pi M(Y_1) M(Y_2) \cdots M(y_n) e.$$

Proof :
Since $\{X_t\}$ is a stationary Markov chain, then the initial probability distribution π satisfies

$$\pi A = A. \tag{4.4}$$

By Theorem 4.1 and equation (4.4), for any $n \in \boldsymbol{N}$, the n-dimensional joint density function of $Y_1, Y_2, \ldots, Y_n$ is

$$\begin{aligned} p(Y_1, Y_2, \ldots, Y_n) &= \pi B(Y_1)M(Y_2)\cdots M(Y_n)e \\ &= \pi AB(Y_1)M(Y_2)\cdots M(Y_n)e \\ &= \pi M(Y_1)M(Y_2)\cdots M(Y_n)e. \end{aligned}$$

■

Since for $i = 1, \ldots, K$,

$$P(X_n = i) = \pi_i \qquad \forall n \in \boldsymbol{N},$$

when $\{X_t\}$ is a stationary Markov chain, then using a similar proof as in the proofs of Theorem 4.1 and Corollary 4.2, for any $m, n \in \boldsymbol{N}$, the n-dimensional joint density function of $Y_m, Y_{m+1}, \ldots, Y_{m+n-1}$ has the form

$$p(Y_m, Y_{m+1}, \ldots, Y_{m+n-1}) = \pi M(Y_m)M(Y_{m+1})\cdots M(Y_{m+n-1})e. \tag{4.5}$$

Equation (4.5) shows that the observation process $\{Y_t\}$ is a stationary process. This implies the pair of stochastic processes $\{(X_t, Y_t)\}$ is also (stricly) stationary. So we have the following corollary.

Corollary 4.3. *If $\{X_t\}$ is a stationary Markov chain, then the hidden Markov model $\{(X_t, Y_t)\}$ is also stationary.*

Lemma 4.4. *For each $n \in \boldsymbol{N}$, the conditional density function of $Y_1, Y_2, \ldots,$ Y_n given $X_1 = i$, for $i = 1, \ldots, K$, is*

$$p(Y_1, Y_2, \ldots, Y_n | X_1 = i) = e_i^T B(Y_1)M(Y_2)\cdots M(Y_n)e,$$

where $e_i^T = (0, \ldots, 0, 1, 0, \ldots, 0)$.

Proof :
Let $n \in \boldsymbol{N}$ and $i \in \{1, \ldots, K\}$, then by Lemma 3.2, Corollary 3.5, Lemma 3.6 and

Lemma 3.7, the conditional density of $Y_1, Y_2, \dots, Y_n$ given $X_1 = i$ is

$$\begin{aligned}
p(y_1, y_2, \dots, y_n | i) &= \sum_{x_2=1}^{K} \cdots \sum_{x_n=1}^{K} p(y_1, x_2, y_2, \dots, x_n, y_n | i) \\
&= \sum_{x_2=1}^{K} \cdots \sum_{x_n=1}^{K} \Big\{ p(y_1|i) p(x_2|i, y_1) p(y_2|i, x_2, y_1) \\
&\qquad \times \cdots \times p(x_n|i, x_2^{n-1}, y_1^{n-1}) p(y_n|i, x_2^n, y_1^{n-1}) \Big\} \\
&= \sum_{x_2=1}^{K} \cdots \sum_{x_n=1}^{K} \Big\{ p(y_1|i) p(x_2|i) p(y_2|x_2) \\
&\qquad \times \cdots \times p(x_n|x_{n-1}) p(y_n|x_n) \} \\
&= \sum_{x_2=1}^{K} \cdots \sum_{x_n=1}^{K} \Big\{ f(y_1, \theta_i) P(X_2 = x_2 | X_1 = i) f(y_2, \theta_{x_2}) \\
&\qquad \times \cdots \times P(X_n = x_n | X_{n-1} = x_{n-1}) f(y_n, \theta_{x_n}) \Big\} \\
&= \sum_{x_2=1}^{K} \cdots \sum_{x_n=1}^{K} f(y_1, \theta_i) \alpha_{i, x_2} f(y_2, \theta_{x_2}) \prod_{t=3}^{n} \alpha_{x_{t-1}, x_t} f(y_t, \theta_{x_t}) \\
&= e_i^T B(y_1) M(y_2) \cdots M(y_n) e.
\end{aligned}$$

So the conclusion of the lemma follows. ■

From Theorem 4.1, 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

$$\alpha_{ij} \geq 0 \qquad , \sum_{j=1}^{K} \alpha_{ij} = 1, \qquad i, j = 1, \dots, K.$$

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

$$\pi_i \geq 0, \qquad i = 1, \dots, K, \qquad \sum_{i=1}^{K} \pi_i = 1.$$

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

Definition 4.5. *Let*

$$\phi = (K, A, \pi, \theta).$$

The parameter ϕ 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 otherhand, 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 $\{1, \dots, K\}$ and its law is determined by the $K \times K$-transition probability matrix A and the initial probability π, and an observation process $\{Y_t\}$ taking values on $\mathcal{Y}$, where the density functions of Y_t given $X_t = i$, $i = 1, \dots, K$ are determined by θ.

References

[1] Ryden, T. 1996. On recursive estimation for hidden Markov models. *Stochastic Processes and their Applications*, 66, 79-96.

[2] R.J. Elliott, L. Aggoun, and J.B. Moore. 1993. *Hidden Markov models: estimation and control*, Volume 29 of Application of Mathematics. Springer Verlag, New York.