International Journal of Applied Mathematics

**Volume 31 No. 1 2018, 53-62**
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
**doi:** http://dx.doi.org/10.12732/ijam.v31i1.5

# PARAMETER ESTIMATION OF EXPONENTIAL HIDDEN MARKOV MODEL AND CONVERGENCE OF ITS PARAMETER ESTIMATOR SEQUENCE

Muhammad Firmasyah[1 §], Berlian Setiawaty[2],
I. Gusti Putu Purnaba[3]

[1,2,3] First Department of Mathematics
Bogor Agricultural University
Meranti Street, Bogor - 16680, INDONESIA

**Abstract:** An exponential hidden Markov model (EHMM) is a hidden Markov model which consists of a pair of stochastic processes $\{X_t, Y_t\}_{t \in N}$.$\{Y_t\}_{t \in N}$ is influenced by $\{X_t\}_{t \in N}$, which is assumed to form a Markov chain. $\{X_t\}_{t \in N}$ is not observed. $\{Y_t\}_{t \in N}$ is an observation process and $Y_t$ given $X_t$ has exponential distribution. In this paper, we estimate the parameter of EHMM and study the convergence of the parameter estimator sequence. EHMM is characterized by a parameter $\phi = (A, \lambda)$ where $A$ is a transition matrix of $X_t$ and $\lambda$ is a vector of parameters of probability density function of $Y_t$ given $X_t$. To determine the parameter estimator, a maximum likelihood method is used. Numerical approximation is used through an Expectation Maximization (EM) algorithm. Under the continuous assumption, the sequence $\{\phi^{(k)}\}$ obtained by the EM algorithm, converges to $\phi^*$ which is the stationary point of ln $L_t(\phi)$ and the sequence $\{\ln L_t(\phi^{(k)})\}$ increasingly converges to ln $L_t(\phi^*)$.

**AMS Subject Classification:** 30B30, 62L12
**Key Words:** convergence, EM algorithm, exponential hidden Markov, forward-backward algorithm

## 1. Introduction

An exponential hidden Markov model (EHMM) is a continuous hidden Markov

Received: December 7, 2017 

§Correspondence author

model which consists of a pair of stochastic processes $\{X_t, Y_t\}_{t\in N}$. $\{Y_t\}_{t\in N}$ is influenced by $\{X_t\}_{t\in N}$, which is not observed. $\{X_t\}_{t\in N}$ is assumed to form a Markov chain. $\{Y_t\}_{t\in N}$ is an observation process which $Y_t$ given $X_t$ has exponential distribution. Let $S_X = \{1, 2, 3, ..., m\}$ be a state space of $\{X_t\}_{t\in N}, A = [a_{ij}]_{m\times m}$ be a transition probability matrix with $a_{ij} = P(X_t = j|X_{t-1} = i) = P(X_2 = j|X_1 = i)$, where $a_{ij} \geq 0, 1 \leq i, j \leq m$ and $\sum_{j=1}^{m} a_{ij} = 1$ for $i \in S_X$. $\varphi = [\varphi_i]_{m\times 1}$ is an initial state probability vector with $\varphi_i = P(X_i = 1)$ for $i = 1, 2, 3, ..., m, \sum_{i=1}^{m} \varphi_i = 1$ and $A\varphi = \varphi$. $\lambda = (\lambda_i)_{m\times 1}$ is a vector that characterizes the probability density function of $Y_t$ given $X_t = i$, that is $\gamma_{yi} = f(y) = \frac{1}{\lambda_i} e^{-\frac{1}{\lambda_i} y}$ for $y > 0$. So the EHMM can be characterized by a parameter $\phi = (A, \lambda)$.

The aims of this paper are:

1. To estimate the parameter $\phi$ for an observation $\{y_t\}$ which is assumed to be generated by the EHMM.
2. To determine the convergence of parameter estimator sequence.

## 2. Parameter Estimation (see [1])

Let $T$ be an observation number, $y = (y_1, y_2, ..., y_T)$ be an observation sequence, and $x = (i_1, i_2, ..., i_T)$ be a sequence which is not observed. Let $\epsilon > 0$ be a number close to 0, and $\Phi = \{\phi = (A, \lambda) : A \in [0, 1]^{m^2}, \lambda \in [\epsilon, \frac{1}{\epsilon}]^m\}$ be the EHMM parameter space.

Assume that:

1. $a_{ij} : \Phi \to \mathbb{R}$ with $a_{ij} = a_{ij}(\phi)$ is a continuous function in $\Phi, \forall i, j \in S_X$.
2. $\lambda_i : \Phi \to \mathbb{R}$ with $\lambda_i = \lambda_i(\phi)$ is a continuous function in $\Phi, \forall i \in S_X$.
3. $\varphi_i : \Phi \to \mathbb{R}$ with $\varphi_i = \varphi_i(\phi)$ is a continuous function in $\Phi, \forall i \in S_X$.

Define the likelihood function for the observation process $Y$ as follows:

$$\begin{aligned} L_T(\phi) &= f(y_1, y_2, ..., y_T|\phi) \\ &= \sum_{i_1=1}^{m} ... \sum_{i_T=1}^{m} f(Y_T = y_T, X_T = i_T, Y_{T-1} = y_{T-1}, X_{T-1} = i_{T-1}, ..., \\ &\qquad Y_1 = y_1, X_1 = i_1|\phi) \end{aligned} \tag{1}$$

$$= \sum_{i_1=1}^{m} \dots \sum_{i_T=1}^{m} \varphi_{i_1} \gamma_{y_1 i_1} \prod_{t=2}^{T} a_{i_{t-1} i_t} \gamma_{y_t i_t}.$$

Define also:

$$\begin{aligned} L_T^c(\phi) &= f(y_T, i_T, y_{T-1}, i_{T-1}, \dots, y_1, i_1|\phi) \\ &= f(y_T|i_T, y_{T-1}, i_{T-1}, \dots, y_1, i_1, \phi) f(i_T, y_{T-1}, i_{T-1}, \dots, y_1, i_1|\phi) \\ &= \varphi_{i_1} \gamma_{y_1 i_1} \prod_{t=2}^{T} a_{i_{t-1} i_t} \gamma_{y_t i_t}. \end{aligned} \tag{2}$$

From (1) and (2), we have

$$\begin{aligned} L_T(\phi) &= \sum_{i_1=1}^{m} \dots \sum_{i_T=1}^{m} \varphi_{i_1} \gamma_{y_1 i_1} \prod_{t=2}^{T} a_{i_{t-1} i_t} \gamma_{y_t i_t} \\ &= \sum_{x} f(y, x|\phi) = \sum_{x} L_T^c(\phi). \end{aligned}$$

Calculating the likelihood function directly is very complicated. So, a Forward-Backward algorithm is used to solve the problem.

### 2.1. Forward-Backward Algorithm

A Forward-backward algorithm is an iterative algorithm which is used to calculate the joint probability of observation process sequence $(y_1, y_2, \dots, y_T)$. The Forward-Backward algorithm is used to speed up the computing process.

Define the forward probability for $t = 1, 2, \dots, T$ and $i = 1, 2, \dots, m$ as

$$\alpha_t(i|\phi) = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_t = y_t, X_t = i|\phi)$$

and the backward probability for $t = T-1, T-2, \dots, 1$ and $i = 1, 2, \dots, m$ as

$$\beta_t(i|\phi) = P(Y_{t+1} = y_{t+1}, \dots, Y_T = y_T|X_t = i, \phi).$$

Then, we have

$$\alpha_1(i|\phi) = \gamma_{y_1 i} \varphi_i,$$

$$\alpha_{t+1}(j|\phi) = \left( \sum_{i \in S_X} \alpha(i|\phi) a_{ij} \right) \gamma_{y_{t+1} j},$$

for $t = 1, 2, \dots, T-1$ and

$$\beta_T(j|\phi) = 1,$$

$$\beta_t(j|\phi) = \sum_{i \in S_X} \beta_{t+1}(i|\phi)\gamma_{y_{t+1}i}a_{ij},$$

for $t = T-1, T-2, ..., 1$ and $i, j \in S_X$.

**Proposition.** (see [3]) *For each* $t = 1, 2, ..., T$ :

$$L_T(\phi) = \sum_{i \in S_X} \alpha_t(i|\phi)\beta_t(i|\phi).$$

The problem is to find $\phi^* \in \Phi$ which maximizes $L_T(\phi)$. We modify the problem becomes to find $\phi^* \in \Phi$ which maximizes $\ln L_T(\phi)$. The EM algorithm is then used to find them. As a result of EM algorithm, we obtain a sequence $\{\phi^{(k)}\}$ in $\Phi$ such that a sequence $\{\ln L_T(\phi^{(k)})\}$ increases and converges to $\ln L_T(\phi)$.

It is known that

$$f(x|y,\phi) = \frac{f(y,x|\phi)}{f(y|\phi)} = \frac{L_T^c(\phi)}{L_T(\phi)},$$

then

$$\ln f(x|y,\phi) = \ln \frac{L_T^c(\phi)}{L_T(\phi)} = \ln L_T^c(\phi) - \ln L_T(\phi),$$
$$\ln L_T(\phi) = \ln L_T^c(\phi) - \ln f(x|y,\phi).$$

From above, for each $\hat{\phi} \in \Phi$,

$$E_{\hat{\phi}}(\ln L_T(\phi)|y) = E_{\hat{\phi}}(\ln L_T^c(\phi)|y) - E_{\hat{\phi}}(\ln f(x|y,\phi)|y) \tag{3}$$

and

$$\begin{aligned} E_{\hat{\phi}}(\ln L_T(\phi)|y) &= \sum_x \ln L_T(\phi) f(x|y,\phi) = \sum_x \ln f(y|\phi) \frac{f(x,y|\hat{\phi})}{f(y|\hat{\phi})} \\ &= \frac{f(y|\phi)}{f(y|\hat{\phi})} \sum_x f(x,y|\hat{\phi}) = \frac{\ln f(y|\phi)}{f(y|\hat{\phi})} f(y|\hat{\phi}) \\ &= \ln f(y|\phi) = \ln L_T(\phi). \end{aligned} \tag{4}$$

Define

$$Q(\phi|\hat{\phi}) = E_{\hat{\phi}}(\ln L_T^c(\phi)|y)$$

and

$$H(\phi|\hat{\phi}) = E_{\hat{\phi}}(\ln f(x|y,\phi)|y).$$

From (3) and (4),

$$\ln L_T(\phi) = Q(\phi|\hat{\phi}) - H(\phi|\hat{\phi}). \tag{5}$$

**Theorem 2.1.** (see [2]) *Let* $\epsilon > 0$ *be a number close to* 0, *and* $\Phi = \{\phi = (A, \lambda) : A \in [0,1]^{m^2}, \lambda \in [\epsilon, \frac{1}{\epsilon}]^m\}$ *be the EHMM parameter space. Then:*

1. $\Phi$ *is a bounded subset in* $\mathbb{R}^{m^3}$.
2. $\ln L_T(\phi)$ *is a continuous function in* $\Phi$ *and differentiable in the interior of* $\Phi$.
3. $\Phi_{\phi^{(0)}} = \{\phi \in \Phi : \ln L_T(\phi) \geq \ln L_T(\phi^{(0)})\}$ *is compact for each* $\ln L_T(\phi^{(0)}) > -\infty$.
4. $Q(\phi|\hat{\phi})$ *is continuous in* $\phi$.

*Proof.* 1. $a_{ij} \in [0,1]$ for each $i, j$ since $a_{ij} = P(X_t = j|X_{t-1} = i)$ and $\lambda_i \in [\epsilon, \frac{1}{\epsilon}]$. Therefore $\Phi \subseteq [0,1]^{m^2} \times [\epsilon, \frac{1}{\epsilon}]^m$ which is a bounded subset in $\mathbb{R}^{m^3}$.

2. Since $L_T(\phi)$ is obtained from an addition and multiplication of continuous and differentiable function in interior $\Phi$, then $L_T(\phi)$ is continuous.

3. Set $\phi^{(0)} \in \Phi$. It will be proven that $\Phi_{\phi^{(0)}}$ is compact. It is enough to prove that $\Phi_{\phi^{(0)}}$ is closed and bounded in $\Phi$. Since $\Phi_{\phi^{(0)}} \subset \Phi$ and $\Phi$ is bounded then $\Phi_{\phi^{(0)}}$ is bounded. $\Phi_{\phi^{(0)}}$ is closed $\leftrightarrow \Phi_{\phi^{(0)}} = \overline{\Phi_{\phi^{(0)}}}$. Since $\Phi_{\phi^{(0)}} \subset \overline{\Phi_{\phi^{(0)}}}$, it is enough to prove $\overline{\Phi_{\phi^{(0)}}} \subset \Phi_{\phi^{(0)}}$. Let $\phi^* \in \overline{\Phi_{\phi(0)}}$ then $\phi^*$ is a limit point of $\Phi_{\phi^{(0)}}$. Thus, there is a sequence $\{\phi^{(k)}\}$ in $\Phi_{\phi^{(0)}}$ with $\ln L_T(\phi^{(k)}) > \ln L_T(\phi^{(0)})$ and $\lim\limits_{k\to\infty} \phi^{(k)} = \phi^*$. If $\phi^* \notin \Phi_{\phi^{(0)}}$ then $\ln L_T(\phi^{(k)}) < \ln L_T(\phi^{(0)})$. Let $\epsilon = L_T(\phi^{(0)}) - L_T(\phi^*) > 0$, since $\lim\limits_{k\to\infty} \phi^{(k)} = \phi^*$ and $\ln L_T(\phi)$ is continuous in $\Phi$, then $\lim\limits_{k\to\infty} L_T(\phi^{(k)}) = L_T(\phi^*)$. For each $\epsilon > 0$, there is $k^*$ such that for each $k \geq k^*$ then $L_T(\phi^{(k)}) - \ln L_T(\phi^*) < \epsilon = L_T(\phi^{(0)}) - L_T(\phi^*)$. So $L_T(\phi^{(k)}) < L_T(\phi^{(0)})$. It is contradicted to the assumption, this implies that $\Phi_{\phi^{(0)}}$ is closed.

4. Since $Q(\phi|\phi^{(k)})$ is an addition and multiplication of

$$\alpha_t(i|\phi^{(k)}), \beta_t(i|\phi^{(k)}), a_{ij}(\phi), \lambda(\phi), \ln \varphi(\phi), \ln \lambda_i(\phi), \ln \gamma_{ij}(\phi),$$

which are continuous in $\Phi$, then $Q(\phi|\phi^{(k)})$ is continuous in $\Phi$.

□

**Corollary 2.1.** *The sequence* $\{\phi^{(k)}\}$ *is well defined in* $\Phi$.

### 2.2. EM Algorithm

1. Set a value $\phi^{(k)}$ for $k = 0$.
2. E step : compute $Q(\phi|\phi^{(k)}) = E_{\phi^{(k)}}(\ln L_T^c(\phi)|Y = y)$.
3. M step : find the value $\phi^{(k+1)}$ which maximizes $Q(\phi|\phi^{(k)})$ so that
$$Q(\phi^{(k+1)}|\phi^{(k)}) \geq Q(\phi|\phi^{(k)}), \forall \phi \in \Phi.$$
4. Replace $k$ by $k+1$ and repeat steps 2 to 4 until $|\ln L_T(\phi^{(k+1)}) - \ln L_T(\phi^{(k)})|$ less than desirable error. In other words the sequence $\{\ln L_T(\phi^{(k)})\}$ is convergent.

**Lemma 2.1.** *$\partial_\phi(\ln L_T(\phi)) = E_{\hat{\phi}}(\partial_\phi \ln L_T(\phi)|y)$, $\partial_\phi Q(\phi|\hat{\phi}) = E_{\hat{\phi}}(\partial_\phi \ln L_T^c(\phi)|y)$ and $\partial_\phi H(\phi|\hat{\phi}) = E_{\hat{\phi}}(\partial_\phi \ln f(x|y, \phi)|y)$.*

From (5) and Lemma 2.1,

$$\begin{aligned} \partial_\phi(\ln L_T(\phi)) &= E_{\hat{\phi}}(\partial_\phi \ln L_T(\phi)|y) \\ &= E_{\hat{\phi}}(\partial_\phi \ln L_T^c(\phi)|y) - E_{\hat{\phi}}(\partial_\phi \ln f(x|y, \phi)|y). \end{aligned} \tag{6}$$

Define

$$D^{10}Q(\phi|\hat{\phi}) = E_{\hat{\phi}}(\partial_\phi \ln L_T^c(\phi)|y), \tag{7}$$

and

$$D^{10}H(\phi|\hat{\phi}) = E_{\hat{\phi}}(\partial_\phi \ln f(x|y, \phi)|y). \tag{8}$$

From (6), (7) and (8)

$$\partial_\phi(\ln L_T(\phi)) = D^{10}Q(\phi|\hat{\phi}) - D^{10}H(\phi|\hat{\phi}). \tag{9}$$

**Lemma 2.2.** *For each $\phi, \hat{\phi} \in \Phi$ then $H(\phi|\hat{\phi}) \leq H(\hat{\phi}|\hat{\phi})$.*

**Lemma 2.3.** *For each $\hat{\phi} \in \Phi$ then $D^{10}H(\hat{\phi}|\hat{\phi}) = 0$.*

Based on Lemma 2.2, Lemma 2.3 and the properties of a maximum and minimum value in a compact metric space (see [4]), it implies the following corollary.

**Corollary 2.2.** *$H(\phi|\hat{\phi})$ attains global maximum at $\hat{\phi}$.*

**Theorem 2.2.** (see [2]) *For each* $\phi^{(k)} \in \Psi$ *we have* $\ln L_T(\phi^{(k+1)}) \geq \ln L_T(\phi^{(k)})$.

*Proof.* It is given an initial value $\phi^{(0)} \in \Phi$. From the EM algorithm, there is $\phi^{(1)} \in \Phi$ so that $\ln L_T(\phi^{(1)}) \geq \ln L_T(\phi^{(0)})$. By obtaining $\phi^{(1)} \in \Phi$, it is obtained $\phi^{(2)}$ so that $\ln L_T(\phi^{(2)}) \geq \ln L_T(\phi^{(1)})$ etc. There is the sequence $\{\phi^{(k)}\}$ which $\ln L_T(\phi^{(k+1)}) \geq \ln L_T(\phi^{(k)})$ and $\{\ln L_T(\phi^{(k)})\}$ which is increasing. It implies for each $\phi^{(k)} \in \Psi$,

$$\begin{aligned} &\ln L_T(\phi^{(k+1)}) - \ln L_T(\phi^{(k)}) \\ &= \Big(Q(\phi^{(k+1)}|\phi^{(k)}) - H(\phi^{(k+1)}|\phi^{(k)})\Big) - \Big(Q(\phi^{(k)}|\phi^{(k)}) - H(\phi^{(k)}|\phi^{(k)})\Big) \\ &= \Big(Q(\phi^{(k+1)}|\phi^{(k)}) - Q(\phi^{(k)}|\phi^{(k)})\Big) - \Big(H(\phi^{(k+1)}|\phi^{(k)}) - H(\phi^{(k)}|\phi^{(k)})\Big). \quad (10) \end{aligned}$$

According to the M step, it is defined $Q(\phi^{(k+1)}|\phi^{(k)}) \geq Q(\phi|\phi^{(k)})$ and from Lemma 2.2, $H(\phi^{(k)}|\phi^{(k)}) \geq H(\phi|\phi^{(k)})$ for each $\phi \in \Phi$ and $H(\phi^{(k+1)}|\phi^{(k)}) - H(\phi^{(k)}|\phi^{(k)}) \leq 0$. So it can be said that

$$\ln L_T(\phi^{(k+1)}) - \ln L_T(\phi^{(k)}) \geq 0$$

$$\ln L_T(\phi^{(k+1)}) \geq \ln L_T(\phi^{(k)}). \quad (11)$$

□

**Theorem 2.3.** (see [2]) *For each* $\phi^{(k)} \notin \Psi$,

$$\ln L_T(\phi^{(k+1)}) - \ln L_T(\phi^{(k)}).$$

*Proof.* From (8), it is known

$$\partial_{\phi^{(k)}}(\ln L_T(\phi^{(k)})) = D^{10}Q(\phi^{(k)}|\phi^{(k)}) - D^{10}H(\phi^{(k)}|\phi^{(k)}).$$

Since $D^{10}H(\phi^{(k)}|\phi^{(k)}) = 0$, then $\partial_{\phi^{(k)}}(\ln L_T(\phi^{(k)})) = D^{10}Q(\phi^{(k)}|\phi^{(k)})$. For $\phi^{(k)} \in \Psi$, $\partial_{\phi^{(k)}}(\ln L_T(\phi^{(k)})) \neq 0$ and $D^{10}Q(\phi^{(k)}|\phi^{(k)}) \neq 0$ so that $\phi^{(k)}$ is not a local maximum of $Q(\phi|\phi^{(k)}))$. According to the M step, it is defined $Q(\phi^{(k+1)}|\phi^{(k)}) \geq Q(\phi|\phi^{(k)})$ for each $\phi \in \Phi$. Thus, it is obtained $Q(\phi^{(k+1)}|\phi^{(k)}) \geq Q(\phi|\phi^{(k)})$ and it implies

$$\ln L_T(\phi^{(k+1)}) \geq \ln L_T(\phi^{(k)}). \quad (12)$$

□

**Corollary 2.3.** *The sequence* $\{\ln L_T(\phi^{(k)})\}$ *is an increasing sequence.*

**Theorem 2.4.** *Let* $E_{\phi^{(k)}}(\ln L_T^c(\phi)|Y=y) = Q(\phi|\phi^{(k)})$, *then*

$$Q(\phi|\phi^{(k)}) = \sum_{i\in S_X} \frac{\alpha_1(i|\phi^{(k)})\beta_1(i|\phi^{(k)}))}{\sum_{l\in S_X}\alpha_t(l|\phi^{(k)})\beta(l|\phi^{(k)})} \ln \varphi_i(\phi)$$

$$+ \sum_{i\in S_X} \frac{\sum_{t=1}^{T}\alpha_t(i|\phi^{(k)})\beta_t(i|\phi^{(k)})}{\sum_{l\in S_X}\alpha_t(l|\phi^{(k)})\beta_t(l|\phi^{(k)})} \ln f(Y_t = y_t|X_t = i, \phi)$$

$$+ \sum_{i\in S_X}\sum_{j\in S_X} \frac{\sum_{t=1}^{T-1} a_{ij}(\phi^{(k)})\alpha_t(i|\phi^{(k)})\beta_{t+1}(j|\phi^{(k)})J(y)}{\sum_{l\in S_X}\alpha_t(l|\phi^{(k)})\beta_t(l|\phi^{(k)})} \ln a_{ij}(\phi),$$

*where* $J(y) = f(Y_{t+1} = y_{t+1}|X_{t+1} = j, \phi^{(k)})$.

**Theorem 2.5.** *Let* $\phi = (A, \lambda)$ *be the parameter of* $Q(\phi|\phi^{(k)})$ *with* $A = [a_{ij}]$ *and* $\lambda = \lambda_i$, *then*

$$a_{ij}(\phi^{(k+1)}) = \frac{\sum_{t=1}^{T-1} a_{ij}(\phi^{(k)})\alpha_t(i|\phi^{(k)})\beta_{t+1}(j|\phi^{(k)})J(y)}{\sum_{t=1}^{T-1}\alpha_t(i|\phi^{(k)})\beta(i|\phi^{(k)})},$$

*where* $J(y) = f(Y_{t+1} = y_{t+1}|X_{t+1} = j, \phi^{(k)})$ *and*

$$\lambda(\phi^{(k+1)}) = \frac{\sum_{t=1}^{T}\alpha_t(i|\phi^{(k)})\beta_t(i|\phi^{(k)})(y_t)}{\sum_{t=1}^{T}\alpha_t(i|\phi^{(k)})\beta_t(i|\phi^{(k)})}.$$

## 3. Convergence of Parameter Estimator EHMM

Let $\{\phi^{(k)}\}$ be the sequence which is obtained from the EM algorithm. It will be proven that the sequence $\{\ln L_T(\phi^{(k)})\}$ converges to $\ln L_T(\phi^*)$ which $\phi^*$ is a stationary point of $\ln L_T(\phi)$.

Based on the properties of a continuous function in a compact metric space (see [4]), we have the following corollaries.

**Corollary 3.1.** *Let* $h : \Phi \to \mathbb{R}^1$ *be a function with* $h(\phi) = \ln L_T(\phi)$. *Then the range of* $h(\phi)$ *is a compact metric space in* $\mathbb{R}^1$.

**Corollary 3.2.** *The range of* $h(\phi)$ *is bounded.*

**Corollary 3.3.** *The sequence* $\{\ln L_T(\phi^{(k)})\}$ *is an increasing and convergent sequence in* $h(\phi)$ *which is convergent. Since* $h(\phi)$ *is compact, there is* $\phi^* \in \Phi$ *such that* $\lim_{k\to\infty} \ln L_T(\phi^{(k)}) = \ln L_T(\phi^*)$.

**Theorem 3.1.** (see [2]) *Let* $g(\hat{\phi}) = \{\delta' \in \Phi : Q(\delta'|\hat{\phi}) \geq Q(\delta|\hat{\phi})$ *for each* $\delta \in \Phi\}$ *then* $g$ *is a closed set in* $\Phi\backslash\Psi$.

*Proof.* Since $g$ is a set value function (see [5]), from $Q(\delta'|\phi')$ it is known that $\delta' \in g(\phi')$ for $\delta', \phi' \in \Phi$. For each $\bar{\phi} \in \Phi\backslash\Psi$ and from Theorem 2.1 (4), $Q(\delta|\phi)$ is a continuous function for $\delta$ and $\phi$ in $\Phi \times \Phi$, if $\phi^{(k)} \to \bar{\phi}$ and $\delta^{(k)} \to \bar{\delta}$ then $Q(\delta^{(k)}|\phi^{(k)}) \to Q(\bar{\delta}|\bar{\phi})$ for $k \to \infty$. So that, it is obtained $\delta^{(k)} \in g(\phi^{(k)})$ for $k = 1, 2, ...$ and it satisfies if $\phi^{(k)} \to \bar{\phi}$ and $\delta^{(k)} \to \bar{\delta}$, then $\bar{\delta} \in g(\bar{\phi})$, for $k \to \infty$. So that $g$ is a closed function, it is satisfied by the EM algorithm i.e $\delta^{(k)}$ corresponding to $\phi^{(k+1)}$. □

**Theorem 3.2.** (see [2]) *Let* $Q(\phi|\phi^{(k)})$ *be a continuous function of* $\phi, \phi^{(k)} \in \Phi \times \Phi$. *Let* $\{\phi^{(k)}\}$ *be the EHMM estimator sequence which is obtained from the EM algorithm,*

1. $\lim_{k\to\infty} \ln L_T(\phi^{(k)}) = \ln L_T(\phi^*)$, *which the convergence is increasing.*
2. *If* $\lim_{k\to\infty} \phi^{(k)} = \phi^*$, *then* $\phi^*$ *is a stationary point of* $\ln L_T(\phi)$.

*Proof.* 1. From Corollary 3.1, Theorem 2.2 and Theorem 2.3,

$$\lim_{k\to\infty} \ln L_T(\phi^{(k)}) = \ln L_T(\phi^*).$$

The sequence $\{\ln L_T(\phi^{(k)})\}$ is an increasing sequence.

2. Let $\lim_{k\to\infty} \phi^{(k)} = \phi^*$, if $\phi^*$ is not a stationary point ($\phi^* \notin \Psi$). We consider a sequence $\{\phi^{(k+1)}\}$ so that $\phi^{(k+1)} \in g(\phi^{(k)})$ for each $k$ and the sequence $\{\phi^{(k+1)}\}$ in a compact set according to Theorem 2.1 (3). It implies that there is the sequence $\{\phi^{(k+1)}\}$ so that for $m \to \infty$ then $\phi^{(k+1)_m} \to \hat{\phi}$ and for $k \to \infty$ then $\phi^{(k+1)} \to \hat{\phi}$. From Theorem 3.1, $g$ is a closed function in $\Phi\backslash\Psi$ and the assumption $\phi^* \notin \Psi$ thus $\hat{\phi} \in g(\phi^*)$. From (12), it implies

$$\ln L_T(\hat{\phi}) > \ln L_T(\phi^*). \tag{13}$$

Since $\ln L_T(\phi)$ in a continuous function, from Theorem 3.2 (1) and $\phi^{(k+1)} \to \hat{\phi}$ for $k \to \infty$, then $\lim_{k\to\infty} \ln L_T(\phi^{(k)}) = \lim_{k\to\infty} \ln L_T(\phi^{(k+1)})$. It implies

$\ln L_T(\phi^*) = \ln L_T(\hat{\phi})$ and it is contradicted by (13). So $\phi^*$ is not a stationary point.

□

## References

[1] A.P Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, **39** (1997), 1-38.

[2] C.J.F. Wu, On the convergence properties of the EM algorithm, *The Annals of Statistics*, **11** (1983), 95-103.

[3] L.R. Rabiner, B.H. Juang, An introduction of hidden Markov models, *IEEE Assp Magazine*, (1997).

[4] R.R. Goldberg, *Methods of Real Analysis.* 2nd Ed., John Wiley and Sons, New York (1976).

[5] W.I. Zangwill, *Nonlinear Programming*, Englewood Cliffs, New Jersey, Prentice Hall (1969).