ESTIMATION OF LINEAR AUTOREGRESSIVE MODELS WITH MARKOV-SWITCHING, THE E.M. ALGORITHM REVISITED
Joseph Rynkiewicz*, SAMOS/MATISSE, University of Paris I, Paris, France

ABSTRACT
This work concerns estimation of linear autoregressive models with Markov-switching using expectation maximisation (E.M.) algorithm. Our method generalises the method introduced by Elliott for general hidden Markov models and avoids using backward recursion.

Key words: Hidden Markov models, Switching models.
MSC: 62M09

RESUMEN
Este trabajo concierne la estimación de modelos lineales autoregresivos con cambios Markov usando el algoritmo de maximización de la esperanza. Nuestro método generaliza el método introducido por Elliott para modelos de Markov ocultos y evita el uso de la recursión descendente.

1. INTRODUCTION

In the present paper we consider an extension of basic (HMM). Let \((X_t, Y_t)_{t \in \mathbb{Z}}\) be the process such that

1. \((X_t)_{t \in \mathbb{Z}}\) is a Markov chain in a finite state space \(IE = \{e_1, \ldots, e_N\}\), which can be identified without loss of generality with the simplex of \(\mathbb{R}^N\), where \(e_i\) is a unit vector in \(\mathbb{R}^N\), with unity as the \(i\)th element and zeros elsewhere.

2. Given \((X_t)_{t \in \mathbb{Z}}\), the process \((Y_t)_{t \in \mathbb{Z}}\) is a sequence of linear autoregressive model in \(\mathbb{R}\) and the distribution of \(Y_n\) depends only of \(X_n\) and \(Y_{n-1}, \ldots, Y_{n-p}\).

Hence, for a fixed \(t\), the dynamic of the model is:

\[
Y_{t+1} = Y_{X_{t+1}} + \sigma_{X_{t+1}} \varepsilon_{t+1}
\]

with \(F_{X_{t+1}} \in \{F_{e_1}, \ldots, F_{e_N}\}\) linear functions, \(\sigma_{X_{t+1}} \in \{\sigma_{e_1}, \ldots, \sigma_{e_N}\}\) strictly positive numbers and \((\varepsilon_t)_{t \in \mathbb{N}}\) a i.i.d. sequence of Gaussian random variable \(N(0,1)\).

Definition 1. Write \(F_t = \sigma(X_0, \ldots, X_t)\), for the \(\sigma\)-field generated by \(X_0, \ldots, X_t\) and \(Y_t = \sigma(Y_0, \ldots, Y_t)\), for the \(\sigma\)-field generated by \(Y_0, \ldots, Y_t\) and \(G_t = \sigma((X_0, Y_0), \ldots, (X_t, Y_t))\), for the \(\sigma\)-field generated by \(X_0, \ldots, X_t\) and \(Y_0, \ldots, Y_t\).

The Markov property implies here that \(P(X_{t+1} = e_i|F_t) = P(X_{t+1} = e_i|X_t)\). Write \(a_{ij} = P(X_{t+1} = e_i|X_t = e_j)\) and \(A = (a_{ij}) \in \mathbb{R}^{N \times N}\) and define: \(V_{t+1} = X_{t+1} - E[X_{t+1}|F_t] = X_{t+1} - AX_t\). With the previous notations, we obtain the general equation of the model, for \(t \in \mathbb{N}\):

\[
\begin{align*}
X_{t+1} &= AX_t + V_{t+1} \\
Y_{t+1} &= F_{X_{t+1}}(Y_{t-p+1}) + \sigma_{X_{t+1}} \varepsilon_{t+1}
\end{align*}
\]  

(1)

The parameters of the model are the transition probabilities of the matrix \(A\), the coefficients of the linear functions \(F_{e_i}\) and the variances \(\sigma_{e_i}\). A successful method for estimating such model is to compute the maximum likelihood estimator\(^1\) with the E.M. algorithm introduced by Demster, Lair and Rubin (1977).

E-mail: *rynkiewi@univ-paris1.fr

\(^1\)This likelihood is computed conditionally to the first “p” observations.
Generally, this algorithm demands the calculus of the conditional expectation of the hidden states knowing the observations (the E-step), this can be done with the Baum and Welch forward-backward algorithm (see Baum et al. (1970)). The derivation of the M-step of the E.M. algorithm is then immediate since we can compute the optimal parameters of the regression functions thanks weighted linear regression.

However we show here that we can also embed these two steps in only one. Namely we can compute, for each step of the E.M. algorithm, directly the optimal coefficients of the regression functions as the variance and the transition matrix thanks a generalisation of the method introduced by Elliott (1994).

2. CHANGE OF MEASURE

The fundamental technique employed throughout this paper is the discrete time change of measure. Write \( \sigma = (\sigma_{e_1}, \ldots, \sigma_{e_n}) \), \( \phi(\cdot) \) for the density of \( N(0,1) \) and \( \langle \cdot, \cdot \rangle \) the inner product in \( \mathbb{R}^N \).

We wish to introduce a new probability measure \( \overline{P} \), using a density \( \Lambda \), so that \( \frac{d\overline{P}}{dP} = \Lambda \) and under \( \overline{P} \) the random variables \( y_t \) are \( N(0,1) \) i.i.d. random variables.

Define

\[
\lambda_l = \frac{\langle \sigma, X_{l-1} \rangle \phi(y_l) \phi(e_l)}{\phi(e_l)}, \quad l \in \mathbb{N}^*, \text{ with } \Lambda_0 = 1 \text{ and } \Lambda_t = \prod_{l=1}^{t} \lambda_l
\]

and construct a new probability measure \( \overline{P} \) by setting the restriction of the Random-Nikodym derivative to \( G_t \) equal to \( \Lambda_t \). Then the following lemma is a straightforward adaptation of lemma 4.1 of Elliott (1994) (see annexe).

**Lemma 1.** Under \( \overline{P} \) the \( Y_t \) are \( N(0,1) \) i.i.d. random variables.

Conversely, suppose we start with a probability measure \( \overline{P} \) such that under \( \overline{P} \)

1. \( (X_t)_{t \in \mathbb{N}} \) is a Markov chain with transition matrix \( A \).
2. \( (Y_t)_{t \in \mathbb{N}} \) is a sequence of \( N(0,1) \) i.i.d. random variable.

We construct a new probability measure \( P \) such that under \( P \) we have \( Y_{t+1} = F_{X_t} (Y_{t-p}) + \sigma_{X_t} \alpha_{t+1} \). To construct \( P \) from \( \overline{P} \), we introduce \( \overline{\lambda}_l := (\lambda_l)^{-1} \) and \( \overline{\Lambda}_t := (\Lambda_t)^{-1} \) and define \( P \) by putting \( \left( \frac{dP}{d\overline{P}} \right)_{G_t} = \overline{\Lambda}_t \).

**Definition 2.** Let \((H_t), t \in \mathbb{N}\) be a sequence adapted to \((G_t)\), we shall write:

\[
\gamma_t(H_t) = \mathbb{E} [ \overline{\Lambda}_t H_t | y_1 ] \text{ and } \Gamma^t(Y_{t+1}) = \frac{\phi \left( Y_{t+1} - F_{X_t} (Y_{t-p+1}) \right)}{\langle \phi(e_l) \rangle}.
\]

The proof of the following theorem is a detailed adaptation of the proof of theorem 5.3 of Elliott (1994) (see annexe).

**Theorem 1.** Suppose \( H_t \) is a scalar G-adapted process of the form: \( H_0 = F_0 \) measurable, \( H_{t+1} = H_t + \alpha_{t+1} + \langle \beta_{t+1}, V_{t+1} \rangle + \delta_{t+1} f(Y_{t+1}) \), \( k \geq 0 \), where \( V_{t+1} = X_{t+1} - AX_t \), \( f \) is a scalar valued function and \( \alpha, \beta, \delta \) are G predictable process (\( \beta \) will be \( N \)-dimensional vector process). Then:
\[ \gamma_{t+1}(H_{t+1}, X_{t+1}) = \gamma_{t+1,t+1}(H_{t+1}) \]
\[ = \sum_{i=1}^{N} \{ \langle \gamma_i(H_t, X_t), \Gamma'(y_{t+1}) \rangle a_i \] 
\[+ \gamma_i(a_{t+1}(X_t, \Gamma'(y_{t+1})) a_i \]
\[+ \gamma_i(\delta_{t+1}(X_t, \Gamma'(y_{t+1}))) f(y_{t+1}) a_i \] 
\[+ (\text{diag}(a_i) + a_i a_i^T) \gamma_i(\beta_{t+1}(X_t, \Gamma'(y_{t+1}))) \] (2)

where \( a_i : = Ae_i, a_i^T \) is the transpose of \( a_i \) and \( \text{diag}(a_i) \) is the matrix with vector \( a_i \) for diagonal and zeros elsewhere.

We will now consider special cases of processes \( H \). In all cases, we will calculate the quantity \( \gamma_{t,t}(H_t) \) and deduce \( \gamma_i(H_t) \) by summing the components of \( \gamma_{t,t}(H_t) \). Then, we deduce from the conditional Bayes' theorem the conditional expectation of \( H_t \):
\[ \hat{H}_t := E[H_t|y_t] = \frac{\gamma_t(H_t)}{\gamma_t(1)}. \]

3. APPLICATION TO THE EXPECTATION (E.-STEP) OF THE E.M. ALGORITHM

We will use the previous theorem in order to compute conditional quantities needed by the E.M. algorithm.

Let \( J_{rs}^t = \sum_{l=1}^{t} \langle X_{t-l}, e_r \rangle \langle X_l, e_s \rangle \) be the number of jump from state \( e_r \) to state \( e_s \) at time \( t \), we obtain:
\[ \gamma_{t+1,t+1}(J_{rs}^t) = \sum_{l=1}^{N} \{ \langle \gamma_{l,t}^i(J_{rs}^t), \Gamma^i(Y_{t+1}) \rangle a_i \] 
\[+ \langle \gamma_i(X_t), \Gamma^i(Y_{t+1}) \rangle a_{se} \].
\[ (3) \]

Write now \( O_{t}^r = \sum_{n=1}^{t-1} \langle X_{n}, e_r \rangle \) for the number of times, up to \( t \), that \( X \) occupies the state \( e_r \). We obtain
\[ \gamma_{t+1,t+1}(O_{t}^r) = \sum_{l=1}^{N} \{ \langle \gamma_{l,t}^i(O_{t}^r), \Gamma^i(Y_{t+1}) \rangle a_i \] 
\[+ \langle \gamma_i(X_t), \Gamma^i(Y_{t+1}) \rangle a_r \].
\[ (4) \]

For the regression functions, the M-Step of the E.M. algorithm is achieved by finding the parameters minimising the weighted sum of squares:
\[ \sum_{l=1}^{n} \gamma_i(t)(y_t - (a_0 + a_1 y_{t-1} + \ldots + a_p y_{t-1-p})^2) \]
where \( \gamma_i(t) \) is the conditional expectation of the hidden \( e_i \) at time \( t \) knowing the observations \( y_{t+1}, \ldots, y_n \).

Write \( \psi^T(t) = (1, y_{t-1}, \ldots, y_{t-p}) \) and \( \theta_i = (a_0^i, \ldots, a_p^i) \), suppose that the matrix \( \sum_{t=1}^{n} \gamma_i(t) \psi(t) \psi^T(t) \) is invertible.

The estimator \( \hat{\theta}_i(n) \) of \( \theta_i \) is given by:
\[ \hat{\theta}_i(n) = \left[ \sum_{t=1}^{n} \gamma_i(t) \psi(t) \psi^T(t) \right]^{-1} \sum_{t=1}^{n} \gamma_i(t) \psi(t) Y_t. \]

Hence, in order to \( \hat{\theta}_i(n) \), we need to estimate the conditional expectation of the following processes:

1. \( T \ A_{t+1}^i(j) = \sum_{l=1}^{t+1} \langle X_t, e_r \rangle Y_l Y_{t+1} \)
   for \(-1 \leq j \leq p \) and \( 1 \leq r \leq N. \)

2. \( T \ B_{t+1}^i(i,j) = \sum_{l=1}^{t+1} \langle X_t, e_r \rangle Y_l Y_{t-i} \)
   for \( 0 \leq j, i \leq p \) and \( 1 \leq r \leq N. \)

3. \( T \ C_{t+1}^i = \sum_{l=1}^{t+1} \langle X_t, e_r \rangle Y_{t-l} \)

4. \( T \ D_{t+1}^i(j) = \sum_{l=1}^{t+1} \langle X_t, e_r \rangle Y_l \)
   for \( 0 \leq j, p \) and \( 1 \leq r \leq N. \)

Applying theorem (2) with \( H_{t+1}(j) = \ A_{t+1}^r(j), H_0 = 0, \alpha_{t+1} = 0, \beta_{t+1} = 0, \delta_{t+1} = \langle X_t, e_r \rangle Y_{t-j} \) and \( f(Y_{t+1}) = Y_{t+1}, \) if \( j \neq -1 \) or \( \delta_{t+1} = \langle X_t, e_r \rangle \) and \( f(Y_{t+1}) = Y^2_{t+1} \) if \( j = -1, \) gives us

\[ \gamma_{t+1, t+1}(T \ A_{t+1}^r(j)) = \sum_{l=1}^{N} \langle \gamma_{t+1}(T \ A_{t+1}^r(j)), \Gamma_i(Y_{t+1}) \rangle a_i + \langle \gamma_{t+1}(X_t), \Gamma_f(Y_{t+1}) \rangle Y_{t-j} Y_{t+1}, \]  

where \( a_i \) is the \( r \)-th column of \( A. \)

Then, applying theorem (2) with \( H_{t+1}(j) = \ B_{t+1}^r(i,j), H_0 = 0, \alpha_{t+1} = 0, \beta_{t+1} = 0, \delta_{t+1} = \langle X_t, e_r \rangle Y_{t-j} Y_{t-i} \) and \( f(Y_{t+1}) = 1 \) gives:

\[ \gamma_{t+1, t+1}(T \ B_{t+1}^r(j)) = \sum_{l=1}^{N} \langle \gamma_{t+1}(T \ B_{t+1}^r(j)), \Gamma_i(Y_{t+1}) \rangle a_i + \langle \gamma_{t+1}(X_t), \Gamma_f(Y_{t+1}) \rangle Y_{t-j} Y_{t-i}, \]  

Next, applying theorem (2) with

\( H_{t+1} = T \ C_{t+1}^r, H_0 = 0, \alpha_{t+1} = 0, \beta_{t+1} = 0, \delta_{t+1} = \langle X_t, e_r \rangle \) and \( f(Y_{t+1}) = Y_{t+1} \) gives:

\[ \gamma_{t+1, t+1}(T \ C_{t+1}^r) = \sum_{l=1}^{N} \langle \gamma_{t+1}(T \ C_{t+1}^r), \Gamma_i(Y_{t+1}) \rangle a_i + \langle \gamma_{t+1}(X_t), \Gamma_f(Y_{t+1}) \rangle Y_{t-j}, \]  

Finally, applying theorem (2) with

\( H_{t+1}(j) = T \ D_{t+1}^r, H_0 = 0, \alpha_{t+1} = 0, \beta_{t+1} = 0, \delta_{t+1} = \langle X_t, e_r \rangle \) \( Y_{t-j} \) and \( f(Y_{t+1}) = 1 \) gives:
The "Maximisation" pass of the E.M. algorithm is now achieved by updating the parameters in the following way.

**Parameters of the transition matrix.** The parameter of the transition matrix will be updated with the formula:

$$\hat{a}_{ir} = \frac{\gamma_T(J_{ir}^r)}{\gamma_T(O_{ir}^r)}$$

(9)

**Parameters of the regression functions.** For $1 \leq r \leq N$, let $R_i^r = (R_{ij}^r)_{1 \leq i \leq p-1}$ be symmetric with $R_{11}^r = 1$, $R_{ij}^r = R_{ji}^r = \hat{T}D_i^r(j)$, $R_{ij} = \hat{T}B_i'(i-1,j-1)$ and $C_i' = \{\hat{T}C_i'(\hat{T}A_i'(i))_{0 \leq i \leq p}\}$. We can then compute the update parameter $\hat{\theta}_r$ of the regression $F_{r0}$ with the formula:

$$\hat{\theta}_r = (R_i^r)^{-1}C_i'$$

(10)

**Parameters of the variances:** Finally, thanks to the previous conditional expectations, we can directly calculate the parameters $\hat{\sigma}_{1r},..,\hat{\sigma}_{Nm}$ since for $1 \leq r \leq N$ the conditional expectation of the mean square error of the $r$th model is

$$\hat{\sigma}_r^2 = \frac{1}{O_r}\left(\hat{T}A_r'(-1) + \hat{O}_rR_r\hat{\theta}_r - 2\hat{\theta}_r^TC_r'\right).$$

(11)

This completes the M-step of the E.M. algorithm.

4. CONCLUSION

Using the discrete Girsanov measure transform, we propose a new way to apply the E.M. algorithm in the case of Markov-switching linear autoregressions.

Note that, contrary to the Baum and Welch algorithm, we don’t use backward recurrence, although the cost of calculus slightly increases since the number of operations is multiplied by $\frac{N}{2}$, where $N$ is the number of hidden states of the Markov chain.

ANNEXE

**Proof of lemma 1**

**Lemma 2.** Under $\overline{P}$ the $Y_t$ are $N(0,1)$ i.i.d. random variables.

**Proof.** The proof is based on the conditional Bayes’ Theorem, it is a simple rewriting of the Proof of Elliott, hence we have

$$\overline{P}\{Y_{t+1} \leq \tau \mid G_i\} = \overline{E}[1_{\{Y_{t+1} \leq \tau\}} \mid G_i]$$

Thanks to the conditional Bayes’ Theorem we have:

$$\overline{E}[1_{\{Y_{t+1} \leq \tau\}} \mid G_i] = \frac{E[\Lambda_{t+1}1_{Y_{t+1} \leq \tau}\mid G_i]}{E[\Lambda_{t+1}\mid G_i]} = \frac{\Lambda_t}{\Lambda_t} \times \frac{E[\Lambda_{t+1}1_{Y_{t+1} \leq \tau}\mid G_i]}{E[\Lambda_{t+1}\mid G_i]}$$

Now
\[
E[\lambda_{t+1} \mid G_t] = \int_{-\infty}^{\infty} \frac{\langle \sigma, X_t \rangle \phi(Y_{t+1})}{\phi(\varepsilon_{t+1})} \times \phi(\varepsilon_{t+1}) d\varepsilon_{t+1}
\]

\[
= \int_{-\infty}^{\infty} \frac{\langle \sigma, X_t \rangle \phi(F_{X_t}(Y_{t+1}^t) + \langle \sigma, X_t \rangle \times \varepsilon_{t+1})}{\phi(\varepsilon_{t+1})} d\varepsilon_{t+1} = 1
\]

and since \( \varepsilon_{t+1} = \frac{Y_{t+1} - F_{X_t}(Y_{t+1}^t)}{\langle \sigma, X_t \rangle} \):

\[
\bar{P}(Y_{t+1} \leq \tau \mid G_t) = E[\lambda_{t+1} 1_{\{Y_{t+1} \leq \tau\}} \mid G_t]
\]

\[
= \int_{-\infty}^{\infty} \frac{\langle \sigma, X_t \rangle \phi(Y_{t+1})}{\phi(\varepsilon_{t+1})} \times 1_{\{Y_{t+1} \leq \tau\}} \times \phi(\varepsilon_{t+1}) d\varepsilon_{t+1}
\]

\[
= \int_{-\infty}^{\infty} \phi(Y_{t+1}) dy_{t+1} = \bar{P}(Y_{t+1} \leq \tau)
\]

\[\blacksquare\]

**Proof of Theorem 2**

**Theorem 2.** Suppose \( H_t \) is a scalar \( G \)-adapted process of the form: \( H_0 \) is \( F_t \) measurable, \( H_{t+1} = H_t + \alpha_{t+1} + \langle \beta_{t+1}, V_{t+1} \rangle + \delta_{t+1} f(Y_{t+1}), k \geq 0 \), where \( V_{t+1} = X_{t+1} - A X_t, f \) is a scalar valued function and \( \alpha, \beta, \delta \) are \( G \) predictable process (\( \beta \) will be \( N \)-dimensional vector process). Then:

\[
\gamma_{t+1}(H_{t+1}X_{t+1}) := \gamma_{t+1,t+1}(H_{t+1})
\]

\[
= \sum_{i=1}^{N} \langle \gamma_t(H_t X_t), \Gamma_{t+1}(Y_{t+1}) \rangle a_i
\]

\[
+ \gamma_{t}(\alpha_{t+1}(X_t, \Gamma_{t}(y_{t+1}))) a_i
\]

\[
+ \gamma_{t}(\delta_{t+1}(X_t, \Gamma_{t}(y_{t+1}))) f(y_{t+1}) a_i
\]

\[
+ \left( \text{diag}(a_i) - a_i a_i^T \right) \gamma_{t}(\beta_{t+1}(X_t, \Gamma_{t}(y_{t+1})))
\]

where \( a_i = A e_i, a_i^T \) is the transpose of \( a_i \) and \( \text{diag}(a_i) \) is the matrix with vector \( a_i \) for diagonal and zeros elsewhere.

**Proof.** Here again it is only a rewriting of the proof of Elliott.

We begin with the two following results:

**Result 1**

\[
\bar{E} \left[ V_{t+1} \mid Y_{t+1} \right] = \bar{E} \left[ \bar{E} \left[ V_{t+1} \mid G_t, Y_{t+1} \right] \mid Y_{t+1} \right] = \bar{E} \left[ \bar{E} \left[ V_{t+1} \mid G_t \right] \mid Y_{t+1} \right] = 0.
\]

(13)

**Result 2**

\[
X_{t+1}X_{t+1}^T = AX_t(AX_t)^T + AX_t V_{t+1}^T + V_{t+1}(AX_t)^T + V_{t+1} V_{t+1}^T.
\]

Since \( X_t \) is of the form \((0,...,0,1,0,...,0)\) we have
\[ X_{t+1} X_{t+1}^T = \text{diag}(X_{t+1}) = \text{diag}(AX_t) + \text{diag}(V_{t+1}) \]

so

\[ V_{t+1} V_{t+1}^T = \text{diag}(AX_t) + \text{diag}(V_{t+1}) - A \text{ diag}(X_t) A^T - AX_t V_{t+1}^T - V_{t+1} (AX_t)^T. \]

Finally we obtain the result

\[
\langle V_{t+1} \rangle = E[V_{t+1} V_{t+1}^T | F_t] = E[V_{t+1} V_{t+1}^T | X_t]
\]

(14)

\[
= \text{diag}(AX_t) - A \text{ diag}(X_t) A^T.
\]

**Main proof** We have

\[
\gamma_{t+1,t+1}(H_{t+1}) = E[\Lambda_{t+1} H_{t+1} X_{t+1} | Y_{t+1}] = E[(AX_t + V_{t+1}) (H_t + \alpha_{t+1} + \delta_{t+1} f(y_{t+1})) \times \Lambda_{t+1} | Y_{t+1}].
\]

Thanks equation (13),

\[
\gamma_{t+1,t+1}(H_{t+1}) = E[((H_t + \alpha_{t+1} + \delta_{t+1} f(y_{t+1})) AX_t + <\beta_{t+1}, V_{t+1}>) \times \Lambda_{t+1} | Y_{t+1}]
\]

so:

\[
\gamma_{t+1,t+1}(H_{t+1}) = \sum_{i=1}^N \sum_{j=1}^N \{ E[((H_t + \alpha_{t+1} + \delta_{t+1} f(y_{t+1})) X_t, e_i, e_j) \Lambda_{t+1} | Y_{t+1}] + E[<\beta_{t+1}, V_{t+1}> \times \Lambda_{t+1} | Y_{t+1}] \}
\]

hence

\[
\gamma_{t+1,t+1}(H_{t+1}) = \sum_{i=1}^N \sum_{j=1}^N \{ E[((H_t + \alpha_{t+1} + \delta_{t+1} f(y_{t+1})) X_t, e_i, e_j) \Lambda_{t+1} | Y_{t+1}] + E[<\beta_{t+1}, V_{t+1}> \times \Lambda_{t+1} | Y_{t+1}] \}
\]

we have noted \( a_i = A e_i \), so

\[
\gamma_{t+1,t+1}(H_{t+1}) = \sum_{i=1}^N \{ E[((H_t + \alpha_{t+1} + \delta_{t+1} f(y_{t+1})) X_t, e_i) \Lambda_{t+1} | Y_{t+1}] + E[<\beta_{t+1}, V_{t+1}> \times \Lambda_{t+1} | Y_{t+1}] \}
\]

Since for an adapted process \( H_t \) to the sigma-algebra \( G_t \)

\[
E[\Lambda_{t+1} H_t | Y_{t+1}] = \sum_{i=1}^N \langle \gamma_i(H_t, X_t), \Gamma_i(y_{t+1}) \rangle
\]

So, for all \( e_t \in IE \)

\[
E[\Lambda_{t+1} H_t < X_t, e_t | Y_{t+1}] = \sum_{i=1}^N \langle \gamma_i(H_t, X_t < X_t, e_t), \Gamma_i(y_{t+1}) \rangle
\]

\[
= \sum_{i=1}^N \langle \gamma_i(H_t, X_t e_t), e_i) e_i^T, \Gamma_i(y_{t+1}) \rangle
\]

But we also:

\[
\gamma_i(H_t, X_t e_t) = \sum_{i=1}^N \langle \gamma_i(H_t, X_t e_t) e_i, e_i^T, \Gamma_i(y_{t+1}) \rangle
\]

So we have:
\[ \mathbb{E}[\Lambda_{t+1} | H_t, X_t, e_t > | Y_{t+1}] = \sum_{i=1}^{N} \langle \gamma_i (H_t X_t^T e_t), \Gamma^i (y_{t+1}) \rangle = \langle \gamma_{11} (H_t X_t), \Gamma (y_{t+1}) \rangle. \]

Since \( \alpha, \beta, \delta \) and \( G \) predictable and \( f(y_{t+1}) \) measurable with respect to \( Y_{t+1} \), the result (14) yields us the conclusion. □

REFERENCES

BAUM, L.E.; T. PETRIE; G. SOULES and N.A. WEISS (1970): “A maximisation technique occurring in the statistical estimation of probabilistic functions of Markov process”, Annals of Mathematical statistics, 41:164-171.

DEMSTER, A.P.; N.M. LAIR and D.B. RUBIN (1966): “Maximum likelihood from incomplete data via the E.M. algorithm”, Journal of the Royal statistical society of London, Series B:39:1-38.

ELLIOTT, R.J. (1994): “Exact Adaptive Filters for Markov chains observed in Gaussian Noise”, Automatica, 30(9):1399-1408.