RECURSIVE ESTIMATION OF THE CLAIM RATES AND SIZES IN AN INSURANCE MODEL

LAKHDAR AGGOUN

Received 19 October 2003 and in revised form 16 May 2004

It is a common fact that for most classes of general insurance, many possible sources of heterogeneity of risk exist. Premium rates based on information from a heterogeneous portfolio might be quite inadequate. One way of reducing this danger is by grouping policies according to the different levels of the various risk factors involved. Using measure change techniques, we derive recursive filters and predictors for the claim rates and claim sizes for the different groups.

1. Introduction

All processes are defined on a measurable space $(\Omega, \mathcal{F})$, with probability measure P. Consider a portfolio of L policyholders of, for instance, automobile insurance. Each policyholder belongs to one of a finite number G of risk level groups classified by age, sex, type of automobile owned, and so forth.

Under the two assumptions that the initial distribution of the rate of claims is $\Gamma(\alpha_0, \beta_0)$ and that the number of claims y and the number of policies N are Poisson random variables, it is easily seen that the posterior probability density of the rate of claims, given new data y, N, is $\Gamma(\alpha_0 + y, \beta_0 + N)$.

More precisely, we will be using the following notation and assumptions.

(i) Let N_n^c be the total number of new policies purchased by individuals classified in group c during the nth year and let y_n^c be the number of claims reported by the cth group during the same year.

(ii) The rate of claims reported by policyholders in the cth group during the nth year, δ_n^c, is a random variable with conditional Γ-distribution

$$P(\delta_n^c \in dx \mid \alpha_{n-1}^c, \beta_{n-1}^c, y_{n-1}^c, N_{n-1}^c) = \frac{\beta_n^c}{\Gamma(\alpha_n^c)} (\beta_n^c x)^{\alpha_n^c - 1} e^{-\beta_n^c x} dx, \tag{1.1}$$

which is close to a normal distribution when α_n^c and β_n^c are large enough, where α_0^c, β_0^c are initial guesses and, for $n \geq 1$,

$$\alpha_n^c = \alpha_{n-1}^c + y_{n-1}^c, \qquad \beta_n^c = \beta_{n-1}^c + N_{n-1}^c. \tag{1.2}$$

Journal of Applied Mathematics and Stochastic Analysis 2004:3 (2004) 245–259
2000 Mathematics Subject Classification: 60K30, 60J10, 90B05
URL: http://dx.doi.org/10.1155/S1048953304310014

In this paper, we assume that

$$\begin{aligned}
P&\left(\delta_n^c \in dx \mid \delta_{n-1}^c = \delta^c, \alpha_{n-1}^c, \beta_{n-1}^c, y_{n-1}^c, N_{n-1}^c,\ c = 1,\dots,G\right) \\
&= \frac{1}{\sigma_n^c\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{x_c - \sum_{i=1}^G d_i^c \delta^j}{\sigma_n^c} \right)^2 \right\} dx \qquad (1.3)\\
&\triangleq \eta_n\left(x_c, \delta^1, \dots, \delta^G, \sigma_n^c\right) dx.
\end{aligned}$$

Here, d_i^c, $c, i = 1,\dots,G$, are real numbers expressing some dependence between claim sizes from the different groups and

$$\sigma_n^c = \frac{\alpha_{n-1}^c + y_{n-1}^c}{\left(\beta_{n-1}^c + N_{n-1}^c\right)^2} = \frac{\alpha_n^c}{\left(\beta_n^c\right)^2}. \qquad (1.4)$$

(iii) The random variables y_n^c, N_n^c are Poisson random variables such that

$$\begin{aligned}
P\left[y_n^c = \ell \mid \mathcal{F}_{n-1}, \delta_n^c\right] &= \frac{e^{-\delta_n^c}\left(\delta_n^c\right)^\ell}{\ell!}, \\
P\left[N_n^c = m \mid \mathcal{F}_{n-1}\right] &= \frac{e^{-\mu_n^c}\left(\mu_n^c\right)^m}{m!}.
\end{aligned} \qquad (1.5)$$

Here, $\mathcal{F}_n = \sigma\{\delta_k^c, y_k^c, N_k^c,\ c = 1,\dots,G,\ X_k,\ k \le n\}$ is a complete filtration. We assume here that μ_n^c is either known or $\mathcal{F}_n$-predictable.

(iv) Let $\overline{S}_n^c$ be the mean claim size of group c by the end of year n. It is usually assumed that the lognormal distribution is suitable for claim sizes. (See, e.g., [4, 5].) The central limit theorem suggests the following (conditional) normal distribution for $\overline{S}_n^c$:

$$\begin{aligned}
P\left(\overline{S}_n^c \in dz \mid \overline{S}_{n-1}^c = s, X_n\right) &= \frac{1}{\Sigma_n^c\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{z - a^c(X_n)s}{\Sigma_n^c} \right)^2 \right\} dz \\
&\triangleq \phi_n^c(z, s, X_n)\, dx.
\end{aligned} \qquad (1.6)$$

Here, $a^c(X_n) = \langle a^c, X_n\rangle$, where $a^c = (a_1^c,\dots,a_K^c)$ may represent the year index [5] which, for simplicity, belongs to the finite set of real numbers a^c. The probability density function of $\overline{S}_n^c$ is modulated by an unobserved finite-state Markov chain X, that is, the mean number of policies purchased every year is changing from year to year due to many economical factors and the changes are modeled by a finite-state Markov chain X. Without loss of generality, let the state space of X be the standard basis $\{e_1,\dots,e_K\}$ of $\mathbb{R}^K$.

Write $P = \{p_{j,i}\}$, $i, j = 1,\dots,K$, where $\sum_{j=1}^K p_{j,i} = 1$ and

$$p_{j,i} = P\left[X_n = e_j \mid X_{n-1} = e_i\right]. \qquad (1.7)$$

Then, we have the following dynamical representation [3]:

$$X_n = PX_{n-1} + V_n, \qquad (1.8)$$

where V_n is a martingale increment with respect to the complete filtration generated by X.

(v) Credibility theory deals with adjusting insurance premiums as claim experience is obtained [4]. The technique consists of using a credibility factor $Z \in (0,1)$ to obtain a convex linear combination of some data obtained from past experience, which may not be very reliable, and data from recently reported claims. In this paper, we propose the following (conditional) normal distribution for $\overline{S}_n^c$:

$$
\begin{aligned}
P(\overline{S}_n^c &\in ds \mid \overline{S}_{n-1}^c = s_1, \overline{S}_{n-2}^c = s_2) \\
&= \frac{1}{\Psi_n^c \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{s - Z^c s_1 - (1 - Z^c) s_2}{\Psi_n^c} \right)^2 \right\} dz \qquad (1.9) \\
&\triangleq \chi_n^c(s, s_1, s_2)\, dx.
\end{aligned}
$$

The parameter Z is reestimated in Section 6.

In Sections 3 and 4, recursive estimates for the rates of claims are derived under a suitable "reference" probability measure.

In Sections 4 and 5, recursive estimates of the claim sizes are derived under a different "reference" probability measure. The reason was to separate between the distributions of the claim rates and the claim sizes. Note that the changes in the economical environment, expressed by the jumps of the Markov chain X, link the claim sizes of the whole portfolio, therefore creating some dependence between the different risk groups.

In Section 6, the expectation maximization (EM) is used to update the parameters of the model.

2. Recursive estimation

In this section, we choose a probability measure $P^\dagger$, on the measurable space $(\Omega, \mathcal{F})$, under which the processes y^c, N^c, $c = 1, \ldots, G$, are sequences of stochastically independent and identically distributed (i.i.d.) random variables. The probability measure P is referred to as the "real world" measure, that is, under this measure, (1.5), (1.6), and (1.8) hold.

Suppose that under the measure $P^\dagger$, processes y^c, N^c, $c = 1, \ldots, G$, are sequences of i.i.d. Poisson random variables with rate 1 independent of everything else. Further, under the measure $P^\dagger$, (1.3) and (1.8) hold.

Define

$$\Lambda_n = \prod_{m=0}^{n} \lambda_m, \qquad (2.1)$$

where $\lambda_0 = 1$ and

$$\lambda_m = \prod_{c=1}^{G} \exp\{1 - \delta_m^c\} (\delta_m^c)^{y_m^c} \exp\{1 - \mu_m^c\} (\mu_m^c)^{N_m^c}. \qquad (2.2)$$

Define the "real world" measure P in terms of $P^\dagger$ by setting $dP/dP^\dagger|_{\mathcal{F}_n} \triangleq \Lambda_n$. Define the measure-valued process

$$g_n(x_1, \ldots, x_G)\, dx_1 \cdots dx_G = E^\dagger[\Lambda_n I(\delta_n^c \in dx_1, \ldots, \delta_n^G \in dx_G) \mid \mathcal{Y}_n]. \qquad (2.3)$$

Here, $\mathcal{Y}_n = \sigma\{y_k^c, N_k^c, c = 1, \ldots, G, k \le n\}$.

Remark 2.1. By Bayes' theorem [3],

$$P(\delta_n^1 \in dx_1,\dots,\delta_n^G \in dx_G \mid \mathcal{Y}_n) = \frac{g_n(x_1,\dots,x_G)\,dx_1\cdots dx_G}{\int_{\mathbb{R}_+^G} g_n(u_1,\dots,u_G)\,du_1\cdots du_G}. \tag{2.4}$$

THEOREM 2.2. *Denote by $g_0(x)$ the initial probability density function of δ. The unnormalized probability density functions $g_n(\cdot) \in \mathbb{R}_+$ satisfy the recursion*

$$\begin{aligned} g_n(x_1,\dots,x_G) = \prod_{c=1}^{G} & \frac{\exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c}\exp\{1-x_c\}(x_c)^{y_n^c}}{\sigma_n^c\sqrt{2\pi}} \\ & \times \int_{\mathbb{R}_+^G} \exp\left\{-\frac{1}{2}\sum_{c=1}^{G}\left(\frac{x_c-\sum_{i=1}^{G} d_i^c u^i}{\sigma_n^c}\right)^2\right\} g_{n-1}(u)\,du. \end{aligned} \tag{2.5}$$

Proof. Let f be a "test" function, $\delta_n = (\delta^1,\dots,\delta^G)$, and write

$$E^\dagger[f(\delta_n)\Lambda_n \mid \mathcal{Y}_n] = \int_{\mathbb{R}_+^G} f(x_1,\dots,x_G)g_n(x_1,\dots,x_G)\,dx_1\cdots dx_G. \tag{2.6}$$

However, in view of (2.1), (2.2), (1.8), and (1.3),

$$\begin{aligned} & E^\dagger\left[\Lambda_n \prod_{c=1}^{G} f(\delta_n) \mid \mathcal{Y}_n\right] \\ & = E^\dagger\left[\Lambda_{n-1}\prod_{c=1}^{G}\exp\{1-\delta_n^c\}(\delta_n^c)^{y_n^c}\exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c} f(\delta_n) \mid \mathcal{Y}_n\right] \\ & = E^\dagger\Bigg[E^\dagger\Bigg[\Lambda_{n-1}\prod_{c=1}^{G}\exp\{1-\delta_n^c\}(\delta_n^c)^{y_n^c} \\ & \qquad \times \exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c} f(\delta_n) \mid \mathcal{Y}_n, \delta_{n-1}^c,\ c=1,\dots,G\Bigg] \mid \mathcal{Y}_n\Bigg] \\ & = \prod_{c=1}^{G}\exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c} E^\dagger\Bigg[\Lambda_{n-1}\int_{\mathbb{R}_+^G}\prod_{c=1}^{G}\eta_n(x_c,\delta_{n-1}^1,\dots,\delta_{n-1}^G,\sigma_n^c) \\ & \qquad \times \exp\{1-x_c\}(x_c)^{y_n^c} f(x_1,\dots,x_G)\,dx_1\cdots dx_G \mid \mathcal{Y}_{n-1}\Bigg] \\ & = \prod_{c=1}^{G}\exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c}\int_{\mathbb{R}_+^G}\int_{\mathbb{R}_+^G}\prod_{c=1}^{G}\eta_n(x,u_1,\dots,u_G,\sigma_n^c) \\ & \qquad \times \exp\{1-x_c\}(x_c)^{y_n^c} g_{n-1}(u_1,\dots,u_G) \\ & \qquad \times f(x_1,\dots,x_G)\,dx_1\cdots dx_G du_1\cdots du_G \quad \text{by (2.6)}. \end{aligned} \tag{2.7}$$

Since f is an arbitrary test function, this finishes the proof. □

3. Predicting future claim rates

In this section, we wish to derive predictors for the rates of claims within the subgroups of policyholders. That is, we wish to compute the conditional probability of δ^c_{n+1} given the history up to the nth year. Define the process

$$h_{n+1,n}(x_1,\dots,x_G)\,dx_1\cdots dx_G = E^\dagger\left[\Lambda_{n+1} I\left(\delta^1_{n+1}\in dx_1,\dots,\delta^G_{n+1}\in dx_G\right) \mid \mathcal{Y}_n\right]. \tag{3.1}$$

Let f be a "test" function and write

$$E^\dagger\left[\Lambda_{n+1} f(\delta_{n+1}) \mid \mathcal{Y}_n\right] = \int_{\mathbb{R}^G_+} f(x_1,\dots,x_G)\, h_{n+1,n}(x_1,\dots,x_G)\,dx_1\cdots dx_G. \tag{3.2}$$

Lemma 3.1.

$$\begin{aligned} h_{n+1,n}&(x_1,\dots,x_G) \\ &= \prod_{c=1}^{G} \exp\{-x_c\}\exp\{-\mu^c_{n+1}\}\sum_{k=0}^{\infty}(x_c)^k(\mu_{n+1})^k\frac{1}{(k!)^2} \\ &\qquad \times \int_{\mathbb{R}^G_+}\exp\left\{-\frac{1}{2}\sum_{c=1}^{G}\left(\frac{x_c-\sum_{i=1}^{G} d^c_i u^i}{\sigma^c_{n+1}}\right)^2\right\} g_n(u_1,\dots,u_G)\,du_1\cdots du_G. \end{aligned} \tag{3.3}$$

Proof.

$$\begin{aligned} &E^\dagger\left[\Lambda_{n+1} f(\delta_{n+1}) \mid \mathcal{Y}_n\right] \\ &\quad = E^\dagger\left[\Lambda_n f(\delta_{n+1})\prod_{c=1}^{G}\exp\{1-\delta^c_{n+1}\}\left(\delta^c_{n+1}\right)^{y^c_{n+1}}\exp\{1-\mu^c_n\}\left(\mu^c_n\right)^{N^c_{n+1}} \mid \mathcal{Y}_n\right] \end{aligned}$$

(since y^c_{n+1}, N^c_{n+1} are not in $\mathcal{Y}_n$, therefore we use their distributions under $P^\dagger$)

$$\begin{aligned} &= E^\dagger\Bigg[E^\dagger\Bigg[\Lambda_n f(\delta_{n+1})\prod_{c=1}^{G}\exp\{1-\delta^c_{n+1}\}\left(\delta^c_{n+1}\right)^{y^c_{n+1}} \\ &\qquad\qquad \times\exp\{1-\mu^c_{n+1}\}\left(\mu_{n+1}\right)^{N^c_{n+1}} \mid \mathcal{Y}_n,\delta^c_{n+1},\mu_{n+1}\Bigg] \mid \mathcal{Y}_n\Bigg] \\ &= E^\dagger\left[\Lambda_n f(\delta_{n+1})\prod_{c=1}^{G}\exp\{1-\delta^c_{n+1}\}\exp\{1-\mu^c_{n+1}\}\sum_{k=0}^{\infty}\left(\delta^c_{n+1}\right)^k(\mu_{n+1})^k\frac{e^{-2}}{(k!)^2} \mid \mathcal{Y}_n\right] \\ &= E^\dagger\Bigg[\Lambda_n\int_{\mathbb{R}^G_+} f(x_1,\dots,x_G)\prod_{c=1}^{G}\eta_{n+1}\left(x_c,\delta^1_n,\dots,\delta^G_n,\sigma^c_{n+1}\right)\exp\{-x_c\} \\ &\qquad\qquad \times\exp\{-\mu^c_{n+1}\}\sum_{k=0}^{\infty}(x_c)^k(\mu_{n+1})^k\frac{1}{(k!)^2}\,dx_1\cdots dx_G \mid \mathcal{Y}_n\Bigg] \end{aligned}$$

(assuming here that μ^c_{n+1} is either known or predictable with respect to $\mathcal{Y}_n$ and using (3.2), this is)

$$= \int_{\mathbb{R}^G_+}\int_{\mathbb{R}^G_+}\prod_{c=1}^{G} f(x_1,\dots,x_G)\eta_{n+1}(x_c,u_1,\dots,u_G,\sigma^c_{n+1})\exp\{-x_c-\mu^c_{n+1}\}$$
$$\times\sum_{k=0}^{\infty}(x_c)^k(\mu_{n+1})^k\frac{1}{(k!)^2}g_n(u_1,\dots,u_G)\,dx_1\cdots dx_G du_1\cdots du_G. \tag{3.4}$$

The unnormalized density $g_n(u_1,\dots,u_G)$ is given recursively in (2.5). Since f is arbitrary, the result follows. □

4. A second change of measure

In this section, we choose a probability measure $\overline{P}$, on the measurable space $(\Omega,\mathcal{F})$, under which the processes $\overline{S}^1,\dots,\overline{S}^G$ are sequences of stochastically i.i.d. random variables with the standard normal distribution.

Define

$$\Gamma_n = \prod_{m=0}^{n}\gamma_m, \tag{4.1}$$

where $\gamma_0 = 1$ and

$$\gamma_m = \prod_{i=1}^{K}\left[\prod_{c=1}^{G}\frac{\phi^c_m(\overline{S}^c_m,\overline{S}^c_{m-1},i)}{\psi^c(\overline{S}^c_m)}\right]^{\langle X_n,e_i\rangle}. \tag{4.2}$$

Here, ψ^c is the density function of the standard normal distribution and $\langle\cdot,\cdot\rangle$ is the inner product of two vectors in $\mathbb{R}^K$.

Now set $dP/d\overline{P}|_{\mathcal{G}_n} \triangleq \Gamma_n$, where

$$\mathcal{G}_n = \sigma\{\delta^c_k, y^c_k, N^c_k, X_k, \overline{S}^c_k,\ c=1,\dots,G,\ k\le n\}. \tag{4.3}$$

Define the measure-valued process

$$\zeta_n(j) = \overline{E}[\Gamma_n\langle X_n,e_j\rangle \mid \mathcal{S}_n]. \tag{4.4}$$

Here, $\mathcal{S}_n = \sigma\{\overline{S}^c_k,\ c=1,\dots,G,\ k\le n\}$.

THEOREM 4.1. *Denote by $\zeta_0(j)$ the initial joint probability density function of X. The unnormalized probability $\zeta_n(j) \in \mathbb{R}_+$ satisfies the recursion*

$$\zeta_n(j) = \prod_{c=1}^{G} \frac{1}{\Sigma_n^c} \exp\left\{-\frac{1}{2}\left(\frac{\overline{S}_n^c - a_j^c \overline{S}_{n-1}^c}{\Sigma_n^c}\right)^2 + \frac{1}{2}(\overline{S}_n^c)^2\right\} \sum_{i=1}^{K} p_{ji}\zeta_{n-1}(i). \tag{4.5}$$

Proof. In view of (4.1), (4.2), and (1.8), we have

$$\begin{aligned} &\overline{E}[\Gamma_n\langle X_n, e_j\rangle \mid \mathcal{S}_n] \\ &\quad = \prod_{c=1}^{G} \frac{\phi_m^c(\overline{S}_n^c, \overline{S}_{n-1}^c, j)}{\psi^c(\overline{S}_n^c)} \overline{E}[\Gamma_{n-1}\langle X_n, e_j\rangle \mid \mathcal{S}_n] \\ &\quad = \prod_{c=1}^{G} \frac{\phi_m^c(\overline{S}_n^c, \overline{S}_{n-1}^c, j)}{\psi^c(\overline{S}_n^c)} \sum_{i=1}^{K} p_{ji}\overline{E}[\Gamma_{n-1}\langle X_{n-1}, e_i\rangle \mid \mathcal{S}_{n-1}] \\ &\quad = \prod_{c=1}^{G} \frac{\phi_m^c(\overline{S}_n^c, \overline{S}_{n-1}^c, j)}{\psi^c(\overline{S}_n^c)} \sum_{i=1}^{K} p_{ji}, \zeta_{n-1}(i) \quad \text{by (4.4)}. \end{aligned} \tag{4.6}$$

□

5. Predicting future claim sizes

In this section, we wish to derive one-year-ahead predictors for the claim size. That is, we wish to compute the joint conditional probability of $\overline{S}_{n+1}^1, \ldots, \overline{S}_{n+1}^G$ given the history up to the nth year. Define the process

$$\xi_{n+1,n}(x_1, \ldots, x_G)\,dx_1 \cdots dx_G = \overline{E}\Big[\Gamma_{n+1} I\Big(\overline{S}_{n+1}^1 \in dx_1, \ldots, \overline{S}_{n+1}^G \in dx_G\Big) \mid \mathcal{S}_n\Big]. \tag{5.1}$$

Let f be a "test" function, $\overline{S}_{n+1} = (\overline{S}_{n+1}^1, \ldots, \overline{S}_{n+1}^G)$, and write

$$E^\dagger[\Gamma_{n+1} f(\overline{S}_{n+1}) \mid \mathcal{Y}_n] = \int_{\mathbb{R}_+^G} f(x_1, \ldots, x_G)\xi_{n+1,n}(x_1, \ldots, x_G)\,dx_1 \cdots dx_G. \tag{5.2}$$

LEMMA 5.1. *The one-step (unnormalized) predictor for the claim sizes is given by the measure*

$$\xi_{n+1,n}(x_1, \ldots, x_G) = \sum_{j=1}^{K} \prod_{c=1}^{G} \frac{1}{\Sigma_{n+1}^c \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x_c - a_j^c \overline{S}_n^c}{\Sigma_{n+1}^c}\right)^2\right\} \sum_{i=1}^{K} p_{ji}\zeta_n(i). \tag{5.3}$$

The unnormalized density $\zeta_n(i)$ is given recursively in Theorem 4.1.

Proof.

$$
\begin{aligned}
&\overline{E}[\Gamma_{n+1} f(\overline{S}_{n+1}) \mid \mathcal{S}_n] \\
&= \sum_{j=1}^{K} \overline{E}[\Gamma_{n+1}\langle X_{n+1}, e_j\rangle f(\overline{S}_{n+1}) \mid \mathcal{S}_n] \\
&= \sum_{j=1}^{K} \overline{E}\left[\Gamma_n \langle X_{n+1}, e_j\rangle f(\overline{S}_{n+1}) \prod_{c=1}^{G} \frac{\phi_{n+1}^c(\overline{S}_{n+1}^c, \overline{S}_n^c, j)}{\psi^c(\overline{S}_{n+1}^c)} \mid \mathcal{S}_n\right] \\
&= \sum_{j=1}^{K}\sum_{i=1}^{K} p_{ji}\overline{E}\left[\Gamma_n\langle X_n, e_i\rangle \int_{\mathbb{R}_+^G} f(x_1,\dots,x_G) \prod_{c=1}^{G} \psi^c(x_c) \frac{\phi_{n+1}^c(x_c, \overline{S}_n^c, j)}{\psi^c(x)} dx_1 \cdots dx_G \mid \mathcal{S}_n\right] \\
&= \sum_{j=1}^{K} \int_{\mathbb{R}_+^G} f(x_1,\dots,x_G) \prod_{c=1}^{G} \phi_{n+1}^c(x_c, \overline{S}_n^c, j)\, dx_1 \cdots dx_G \sum_{i=1}^{K} p_{ji} \overline{E}[\Gamma_n \langle X_n, e_i\rangle \mid \mathcal{S}_n] \\
&= \sum_{j=1}^{K} \int_{\mathbb{R}_+^G} f(x_1,\dots,x_G) \prod_{c=1}^{G} \phi_{n+1}^c(x_c, \overline{S}_n^c, j)\, dx_1 \cdots dx_G \sum_{i=1}^{K} p_{ji} \zeta_n(i).
\end{aligned}
\tag{5.4}
$$

This finishes the proof. □

6. The EM algorithm

The EM algorithm (see [1, 2]) is a widely used iterative numerical method for computing maximum likelihood parameter estimates (MLEs) of partially observed models such as linear Gaussian state-space models. For such models, direct computation of the MLE is difficult. The EM algorithm has the appealing property that successive iterations yield parameter estimates with nondecreasing values of the likelihood function.

Suppose that we have observations $y_1, \dots, y_K$ available, where K is a fixed positive integer. Let $\{P_\theta, \theta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{F})$, all absolutely continuous with respect to a fixed probability measure P_0. The log-likelihood function for computing an estimate of the parameter θ based on the information available in $\mathcal{Y}_K$ is

$$
\mathcal{L}_K(\theta) = E_0\left[\log \frac{dP_\theta}{dP_0} \mid \mathcal{Y}_K\right] \tag{6.1}
$$

and the MLE is defined by

$$
\hat{\theta} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \mathcal{L}_K(\theta). \tag{6.2}
$$

Let $\hat{\theta}_0$ be the initial parameter estimate. The EM algorithm generates a sequence of parameter estimates $\{\hat{\theta}_j\}$, $j \geq 1$, as follows.

Each iteration of the algorithm consists of two steps.

Step 1 (E-step). Set $\widetilde{\theta} = \hat{\theta}_j$ and compute $\mathcal{Q}(\theta, \widetilde{\theta})$, where

$$\mathcal{Q}(\theta, \widetilde{\theta}) = E_{\widetilde{\theta}}\left[\log \frac{dP_\theta}{dP_{\widetilde{\theta}}} \mid \mathcal{Y}_K\right]. \tag{6.3}$$

Step 2 (M-step). Find $\hat{\theta}_{j+1} \in \operatorname{argmax}_{\theta \in \Theta} \mathcal{Q}(\theta, \hat{\theta}_j)$.

Using Jensen's inequality, it can be shown (see [2, Theorem 1]) that the sequence of model estimates $\{\hat{\theta}_j,\ j \geq 1\}$ from the EM algorithm is such that the sequence of likelihoods $\{\mathcal{L}_K(\hat{\theta}_j)\}$, $j \geq 1$, is monotonically increasing with equality if and only if $\hat{\theta}_{j+1} = \hat{\theta}_j$.

Sufficient conditions for convergence of the EM algorithm are given in [6]. We briefly summarize them here: assume that

(i) the parameter space Θ is a subset of some finite-dimensional Euclidean space $\mathbb{R}^r$;
(ii) $\{\theta \in \Theta : \mathcal{L}_K(\theta) \geq \mathcal{L}_K(\hat{\theta}_0)\}$ is compact for any $\mathcal{L}_K(\hat{\theta}_0) > -\infty$;
(iii) $\mathcal{L}_K$ is continuous in Θ and differentiable in the interior of Θ (as a consequence of (i), (ii), and (iii), clearly $\mathcal{L}_K(\hat{\theta}_j)$ is bounded from above);
(iv) the function $\mathcal{Q}(\theta, \hat{\theta}_j)$ is continuous in both θ and $\hat{\theta}_j$.

Then, by [6, Theorem 2], the limit of the sequence of EM estimates $\{\hat{\theta}_j\}$ has a stationary point $\bar{\theta}$ of $\mathcal{L}_K$. Also, $\{\mathcal{L}_K(\hat{\theta}_j)\}$ converges monotonically to $\bar{\mathcal{L}}_t = \mathcal{L}_t(\bar{\theta})$ for some stationary point $\bar{\theta}$. To make sure that $\bar{\mathcal{L}}_t$ is a maximum value of the likelihood, it is necessary to try different initial values $\hat{\theta}_0$.

Here, we wish to update the parameters from $\widetilde{\theta} = \{\widetilde{d}_i^c,\ i, c = 1, \ldots, G\}$ to a set $\theta(n) = \{d_i^c(n),\ i, c = 1, \ldots, G\}$.

Let $D = \{d_i^c\}$, $c, i = 1, \ldots, G$, be a $G \times G$ nonsingular matrix and $\delta_n = (\delta_n^1, \ldots, \delta_n^G)$.

Definition 6.1. Given two (column) vectors X and Y, the tensor or Kronecker product $X \otimes Y$ is the (column) vector obtained by stacking the rows of the matrix XY', where $'$ is the transpose, with entries obtained by multiplying the ith entry of X by the jth entry of Y.

For instance, if $\{e_1, e_2\}$ is the standard basis of $\mathbb{R}^2$,

$$\begin{aligned} e_1 \otimes e_1 &= (1,0,0,0)', & e_1 \otimes e_2 &= (0,1,0,0)', \\ e_2 \otimes e_1 &= (0,0,1,0)', & e_2 \otimes e_2 &= (0,0,0,1)'. \end{aligned} \tag{6.4}$$

Maximum likelihood estimation of the parameters via the EM algorithm requires computation of the filtered estimates of quantities such as

$$\mathcal{T}_n^{(1)} = \sum_{m=1}^{n} \delta_m \otimes \delta_{m-1}, \qquad \mathcal{T}_n^{(2)} = \sum_{m=1}^{n} \delta_{m-1} \otimes \delta_{m-1}. \tag{6.5}$$

Let $f_i, f_j \in \mathbb{R}^G$ denote unit vectors with 1 in the ith and jth positions, respectively.

For $i,j\in\{1,\dots,G\}$,

$$\mathcal{T}_n^{ij(1)}=\sum_{m=1}^{n}\langle\delta_m,f_i\rangle\langle\delta_{m-1},f_j\rangle,\qquad \mathcal{T}_n^{ij(2)}=\sum_{m=1}^{n}\langle\delta_{m-1},f_i\rangle\langle\delta_{m-1},f_j\rangle;\tag{6.6}$$

here, $\langle\cdot,\cdot\rangle$ denotes the scalar product.

Note that $\mathcal{T}_n^{ij(1)}$ and $\mathcal{T}_n^{ij(2)}$ are merely the elements of the matrices $\mathcal{T}_n^{(1)}$ and $\mathcal{T}_n^{(2)}$, respectively.

Now the expression for $\mathcal{Q}(\theta,\widetilde{\theta})$ is derived.

To update the set of parameters from $\widetilde{\theta}$ to θ, we introduce the density $dP_\theta/dP_{\widetilde{\theta}}|_{\mathcal{G}_n}=\prod_{m=0}^{n}\gamma_m$, where $\gamma_0=1$ and, for $m\geq 1$, $\gamma_m=\psi(\sigma_n^{-1}(\delta_m-D\delta_{m-1}))/\psi(\sigma_n^{-1}(\delta_m-\widetilde{D}\delta_{m-1}))$. Here, $\psi(\cdot)$ is the standard multivariate normal distribution $N(0_{1\times G},I_{G\times G})$ and σ_n is a $G\times G$ diagonal matrix with diagonal entries $\sigma_n^1,\dots,\sigma_n^G$, where σ_n^c is given in (1.4). Now

$$\begin{aligned}E_{\widetilde{\theta}}\left[\log\frac{dP_\theta}{dP_{\widetilde{\theta}}}\bigg|_{\mathcal{G}_n}\mid\mathcal{Y}_n\right]&=\frac{1}{2}E_{\widetilde{\theta}}\left[\sum_{m=1}^{n}(\delta_m-D\delta_{m-1})'\sigma_n^{-2}(\delta_m-D\delta_{m-1})\mid\mathcal{Y}_n\right]+R(\widetilde{\theta})\\&=\mathcal{Q}(\theta,\widetilde{\theta}),\end{aligned}\tag{6.7}$$

where $R(\widetilde{\theta})$ does not involve θ.

To implement the M-step, set the derivatives $\partial\mathcal{Q}/\partial\theta=0$. This yields

$$D(n)=E_{\widetilde{\theta}}\left[\sum_{m=1}^{n}\delta_m\otimes\delta_{m-1}\mid\mathcal{Y}_n\right]\left(E_{\widetilde{\theta}}\left[\sum_{m=1}^{n}\delta_{m-1}\otimes\delta_{m-1}\mid\mathcal{Y}_n\right]\right)^{-1}.\tag{6.8}$$

Define the measure-valued processes

$$\begin{aligned}\beta_n^{ij(1)}(x)&=E^{\dagger}\left[\Lambda_n\mathcal{T}_n^{ij(1)}I(\delta_n\in dx)\mid\mathcal{Y}_n\right],\\\beta_n^{ij(2)}(x)&=E^{\dagger}\left[\Lambda_n\mathcal{T}_n^{ij(2)}I(\delta_n\in dx)\mid\mathcal{Y}_n\right].\end{aligned}\tag{6.9}$$

Then, for any "test" function $g:\mathbb{R}^G\to\mathbb{R}$, write

$$\begin{aligned}E^{\dagger}\left[\Lambda_n\mathcal{T}_n^{ij(1)}g(\delta_n)\mid\mathcal{Y}_n\right]&=\int_{\mathbb{R}^G}\beta_n^{ij(1)}(x)g(x)dx,\\E^{\dagger}\left[\Lambda_n\mathcal{T}_n^{ij(2)}g(\delta_n)\mid\mathcal{Y}_n\right]&=\int_{\mathbb{R}^G}\beta_n^{ij(2)}(x)g(x)dx.\end{aligned}\tag{6.10}$$

Theorem 6.2. *Denote by* $\beta_0^{ij(1)}(x)$, $\beta_0^{ij(2)}(x)$ *the initial probability density functions of* $\mathcal{T}^{ij(1)}$ *and* $\mathcal{T}^{ij(1)}$, *respectively. The unnormalized probability densities* $\beta_n^{ij(i)}(x) \in \mathbb{R}_+$, $i = 1, 2$, *satisfy the recursions*

$$\begin{aligned}
&\beta_n^{ij(1)}(x_1,\dots,x_G)\\
&\quad = \prod_{c=1}^{G} \frac{\exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c}\exp\{1-x_c\}(x_c)^{y_n^c}}{\sigma_n^c\sqrt{2\pi}}\\
&\qquad\times\Bigg[\int_{\mathbb{R}_+^G}\exp\Bigg\{-\frac{1}{2}\sum_{c=1}^{G}\bigg(\frac{x_c-\sum_{i=1}^G d_i^c u^i}{\sigma_n^c}\bigg)^2\Bigg\}\beta_{n-1}^{ij(1)}(u)du\\
&\qquad\quad+\langle x, f_j\rangle\int_{\mathbb{R}_+^G}\exp\Bigg\{-\frac{1}{2}\sum_{c=1}^{G}\bigg(\frac{x_c-\sum_{i=1}^G d_i^c u^i}{\sigma_n^c}\bigg)^2\Bigg\}\langle u, f_i\rangle g_{n-1}(u)du\Bigg],
\end{aligned} \tag{6.11}$$

$$\begin{aligned}
&\beta_n^{ij(2)}(x_1,\dots,x_G)\\
&\quad = \prod_{c=1}^{G} \frac{\exp\{1-\mu_n^c\}(\mu_n^c)^{N_n^c}\exp\{1-x_c\}(x_c)^{y_n^c}}{\sigma_n^c\sqrt{2\pi}}\\
&\qquad\times\Bigg[\int_{\mathbb{R}_+^G}\exp\Bigg\{-\frac{1}{2}\sum_{c=1}^{G}\bigg(\frac{x_c-\sum_{i=1}^G d_i^c u^i}{\sigma_n^c}\bigg)^2\Bigg\}\beta_{n-1}^{ij(2)}(u)du\\
&\qquad\quad+\int_{\mathbb{R}_+^G}\exp\Bigg\{-\frac{1}{2}\sum_{c=1}^{G}\bigg(\frac{x_c-\sum_{i=1}^G d_i^c u^i}{\sigma_n^c}\bigg)^2\Bigg\}\langle u, f_i\rangle\langle u, f_j\rangle g_{n-1}(u)du\Bigg].
\end{aligned}$$

Proof. The proof is similar to that of Theorem 4.1. □

To update the parameters in (1.6), let $A(X_n)$ be a $G\times G$ diagonal matrix with diagonal entries $a^1(X_n),\dots,a^G(X_n)$, that is, on the event $[X_n = e_j]$, $A(e_j) = \operatorname{diag}(a_j^1,\dots,a_j^G)$. We assume that $a_j^c \neq 0$ for all c and all j.

Write $\overline{S}_n = (\overline{S}_n^1,\dots,\overline{S}_n^G)$ and $\Sigma = \operatorname{diag}(\Sigma^1,\dots,\Sigma^G)$.

To update the set of parameters from $\widetilde{\theta} = \{\widetilde{A}^j,\ j = 1,\dots,K,\widetilde{\Sigma}\}$ to $\theta(n) = \{A^j(n),\ j = 1,\dots,K,\Sigma(n)\}$, introduce $dP_\theta/dP_{\widetilde{\theta}}|_{\mathcal{G}_n} = \prod_{m=0}^k \epsilon_m$, where $\epsilon_0 = 1$ and

$$\epsilon_m = \prod_{j=1}^{K}\left[\frac{|\widetilde{\Sigma}^j|\,\psi\big((\Sigma^j)^{-1}(\overline{S}_m - A^j\overline{S}_{m-1})\big)}{|\Sigma^j|\,\psi\big((\widetilde{\Sigma}^j)^{-1}(\overline{S}_m - \widetilde{A}^j\overline{S}_{m-1})\big)}\right]^{\langle X_m, e_j\rangle}. \tag{6.12}$$

Here, $\psi(\cdot)$ is the standard multivariate normal distribution $N(0_{1\times G}, I_{G\times G})$.

Now recall that $\mathcal{S}_n = \sigma\{\overline{S}_k^c, c = 1,\dots,G, k \le n\}$; therefore

$$
\begin{aligned}
E_{\tilde{\theta}}\left[\log \frac{dP_\theta}{dP_{\tilde{\theta}}}\bigg|_{\mathcal{G}_n} \mid \mathcal{S}_n\right]
&= -\sum_{i=1}^{K} E_{\tilde{\theta}}\left[\sum_{m=1}^{n}\langle X_m, e_j\rangle \mid \mathcal{S}_n\right]\log|\tilde{\Sigma}^j| \\
&\quad + \frac{1}{2}\sum_{i=1}^{K} E_{\tilde{\theta}}\left[\sum_{m=1}^{n}\langle X_m, e_j\rangle \mid \mathcal{S}_n\right](\overline{S}_m - A^j\overline{S}_{m-1})'(\Sigma^j)^{-2}(\overline{S}_m - A^j\overline{S}_{m-1}) \\
&\quad + R(\tilde{\theta}) = \mathcal{Q}(\theta, \tilde{\theta}),
\end{aligned}
\tag{6.13}
$$

where $R(\tilde{\theta})$ does not involve θ.

To implement the M-step, set the derivatives $\partial\mathcal{Q}/\partial\theta = 0$. This yields

$$
\begin{aligned}
A^j(n) &= \frac{E_{\tilde{\theta}}\left[\sum_{m=1}^{n}\langle X_m, e_j\rangle \mid \mathcal{S}_n\right]\overline{S}_m \otimes \overline{S}_{m-1}}{E_{\tilde{\theta}}\left[\sum_{m=1}^{n}\langle X_m, e_j\rangle \mid \mathcal{S}_n\right]\overline{S}_{m-1} \otimes \overline{S}_{m-1}}, \\
(\Sigma^j)^2(n) &= \frac{E_{\tilde{\theta}}\left[\sum_{m=1}^{n}\langle X_m, e_j\rangle \mid \mathcal{S}_n\right](\overline{S}_m - A^j\overline{S}_{m-1}) \otimes (\overline{S}_m - A^j\overline{S}_{m-1})}{E_{\tilde{\theta}}\left[\sum_{m=1}^{n}\langle X_m, e_j\rangle \mid \mathcal{S}_n\right]}.
\end{aligned}
\tag{6.14}
$$

Let $\mathbf{T}_n^j = \sum_{m=1}^{n}\langle X_m, e_j\rangle$ and define the process $\Upsilon_n^j = \overline{E}[\Lambda_n \mathbf{T}_n^j \mid \mathcal{S}_n]$.

Remark 6.3. The following recursive filters are derived under $\overline{P}$ which is defined in Section 4. A closed-form finite-dimensional recursion is only possible for the conditional joint distributions of $\mathbf{T}_n^j$ and X_n. That is, we will consider recursive filters for $\overline{E}[\Gamma_n \mathbf{T}_n^j X_n \mid \mathcal{S}_n] \triangleq \varepsilon_n^i$. However, $\Upsilon_n^j = \sum_\ell \langle \varepsilon_n^j, e_\ell\rangle$.

THEOREM 6.4. *Let ε_0^j be the initial joint density function of $\mathbf{T}_0^j$, X_0 and, for $n \ge 1$,*

$$
\begin{aligned}
\varepsilon_n^j = \prod_{c=1}^{G}\frac{1}{\Sigma_n^c}\exp\left\{-\frac{1}{2}\left(\frac{\overline{S}_n^c - a_j^c\overline{S}_{n-1}^c}{\Sigma_n^c}\right)^2 + \frac{1}{2}(\overline{S}_n^c)^2\right\} \\
\times \sum_{t,\ell=1}^{K} p_{\ell t} e_\ell \langle \varepsilon_{n-1}^j, e_t\rangle + e_j\zeta_n(j).
\end{aligned}
\tag{6.15}
$$

Proof. Note that $\mathbf{T}_n^j = \mathbf{T}_{n-1}^j + \langle X_n, e_j\rangle$. Hence

$$
\overline{E}[\Gamma_n \mathbf{T}_n^j X_n \mid \mathcal{S}_n] = \overline{E}[\Gamma_n \mathbf{T}_{n-1}^j X_n \mid \mathcal{S}_n] + \overline{E}[\Gamma_n \langle X_n, e_j\rangle X_n \mid \mathcal{S}_n]. \tag{6.16}
$$

However, in view of (1.8), (2.1), and (2.2),

$$\begin{aligned}
&\overline{E}[\Gamma_n \mathbf{T}^j_{n-1} X_n \mid \mathcal{S}_n] \\
&\quad= \sum_{\ell=1}^{K} e_\ell \overline{E}[\Gamma_n \mathbf{T}^j_{n-1} \langle X_n, e_\ell\rangle \mid \mathcal{S}_n] \\
&\quad= \prod_{c=1}^{G} \frac{\phi^c_m(\overline{S}^c_n, \overline{S}^c_{n-1}, j)}{\psi^c(\overline{S}^c_n)} \sum_{\ell=1}^{K} e_\ell \overline{E}[\Lambda_{n-1} \mathbf{T}^j_{n-1} \langle PX_{n-1}, e_\ell\rangle \mid \mathcal{S}^c_n] \\
&\quad= \prod_{c=1}^{G} \frac{\phi^c_m(\overline{S}^c_n, \overline{S}^c_{n-1}, j)}{\psi^c(\overline{S}^c_n)} \sum_{t,\ell=1}^{K} p_{\ell t} e_\ell \overline{E}[\Lambda_{n-1} \mathbf{T}^j_{n-1} \langle X_{n-1}, e_t\rangle \mid \mathcal{S}^c_{n-1}] \\
&\quad= \prod_{c=1}^{G} \frac{\phi^c_m(\overline{S}^c_n, \overline{S}^c_{n-1}, j)}{\psi^c(\overline{S}^c_n)} \sum_{t,\ell=1}^{K} p_{\ell t} e_\ell \langle \varepsilon^j_{n-1}, e_t\rangle, \\
&\overline{E}[\Gamma_n \langle X_n, e_j\rangle X_n \mid \mathcal{S}_n] = e_j \overline{E}[\Gamma_n \langle X_n, e_j\rangle \mid \mathcal{S}_n] = e_j \zeta_n(j),
\end{aligned} \tag{6.17}$$

where $\zeta_n(r)$ is given recursively in Theorem 4.1. This finishes the proof. □

To update the parameters in (1.9) from $\widetilde{\theta} = \{\widetilde{Z}^c, \widetilde{\Psi}^c\}$ to $\theta(n) = \{Z^c(n), \Psi^c(n)\}$, set $dP_\theta / dP_{\widetilde{\theta}}|_{\mathcal{G}_n} = \prod_{m=0}^{n} \kappa_m$, where $\kappa_0 = 1$ and

$$\kappa_m = \frac{\widetilde{\Psi}^c \psi\big((\widetilde{\Psi}^c)^{-1}(\overline{S}_m - Z^c \overline{S}_{m-1} - (1 - Z^c)\overline{S}_{m-2})\big)}{\Psi^c \psi\big((\widetilde{\Psi}^c)^{-1}(\overline{S}_m - \widetilde{Z}^c \overline{S}_{m-1} - (1 - \widetilde{Z}^c)\overline{S}_{m-2})\big)}. \tag{6.18}$$

Here, $\psi(\cdot)$ is the standard normal distribution $N(0_{1\times G}, I_{G\times G})$.

Now

$$\begin{aligned}
\log \frac{dP_\theta}{dP_{\widetilde{\theta}}} &= -n \log \widetilde{\Psi}^c + \frac{1}{2} \sum_{m=1}^{n} \big(\overline{S}^c_m - Z^c \overline{S}^c_{m-1} - (1 - Z^c)\overline{S}^c_{m-2}\big)^2 + R(\widetilde{\theta}) \\
&= \mathcal{Q}(\theta, \widetilde{\theta}),
\end{aligned} \tag{6.19}$$

where $R(\widetilde{\theta})$ does not involve θ.

To implement the M-step, set the derivatives $\partial \mathcal{Q} / \partial \theta = 0$. This yields

$$\begin{aligned}
Z^c(n) &= \frac{\sum_{m=1}^{n} (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})(\overline{S}^c_{m-2} - \overline{S}^c_m)}{\sum_{m=1}^{n} (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})^2}, \\
(\Psi^c)^2(n) &= \frac{1}{n} \sum_{m=1}^{n} \big(\overline{S}^c_m - Z^c \overline{S}^c_{m-1} - (1 - Z^c)\overline{S}^c_{m-2}\big)^2.
\end{aligned} \tag{6.20}$$

Remark 6.5. Since $Z \in (0,1)$, it is clear that $\overline{S}^c_m$ is between $\overline{S}^c_{m-1}$ and $\overline{S}^c_{m-2}$. Therefore, $\overline{S}^c_{m-2} - \overline{S}^c_{m-1}$ and $\overline{S}^c_{m-2} - \overline{S}^c_m$ have the same sign. So we may assume that they are both positive. Hence we can use the Cauchy-Schwartz inequality to see that

$$\begin{aligned} 0 < Z^c(n) &= \frac{\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})(\overline{S}^c_{m-2} - \overline{S}^c_m)}{\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})^2} \\ &\leq \frac{\left[\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})^2\right]^{1/2}\left[\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_m)^2\right]^{1/2}}{\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})^2} \\ &\leq \left[\frac{\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_m)^2}{\sum_{m=1}^n (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})^2}\right]^{1/2} < 1 \end{aligned} \tag{6.21}$$

because $(\overline{S}^c_{m-2} - \overline{S}^c_m)^2 < (\overline{S}^c_{m-2} - \overline{S}^c_{m-1})^2$.

To replace the parameters p_{ji} by $\hat{p}_{ji}(n)$ in the Markov chain X, we define

$$\mathcal{L}_n = \prod_{m=1}^n \prod_{i,j=1}^K \left(\frac{\hat{p}_{ji}(n)}{p_{ji}}\right)^{\langle X_m, e_j\rangle\langle X_{m-1}, e_i\rangle} \tag{6.22}$$

and set $dP_{\hat{\theta}}/dP_\theta|_{\mathcal{G}_n} = \mathcal{L}_n$.

Then one can show [3] that the new estimates of the parameter $\hat{p}_{ji}(n)$, given the observations up to time n, are given by

$$\hat{p}_{ji}(n) = \frac{\overline{E}[\Gamma_n \mathcal{J}^{ij}_n \mid \mathcal{Y}_n]}{\sum_{j=1}^K \overline{E}[\Gamma_n \mathcal{J}^{ij}_n \mid \mathcal{Y}_n]} \triangleq \frac{\gamma_n(\mathcal{J}^{ij}_n)}{\sum_{j=1}^K \gamma_n(\mathcal{J}^{ij}_n)}, \tag{6.23}$$

where $\mathcal{J}^{ij}_n = \sum_{m=1}^n \langle X_{m-1}, e_i\rangle\langle X_m, e_j\rangle$.

Remark 6.6. The following recursive filters are derived under $\overline{P}$ which is defined in Section 4. A closed-form finite-dimensional recursion is only possible for the conditional joint distributions of $\mathcal{J}^{ij}_n$ and X_n. That is, we will consider recursive filters for $\overline{E}[\Gamma_n \mathcal{J}^{ij}_n X_n \mid \mathcal{Y}_n] \triangleq \rho^{ij}(n)$. However, $\gamma_n(\mathcal{J}^{ij}_n) = \sum_\ell \langle \rho^{ij}(n), e_\ell\rangle$.

LEMMA 6.7. *Let ρ^{ij}_0 be the initial joint density function of $\mathcal{J}^{ij}_0$, X_0 and, for $n \geq 1$,*

$$\begin{aligned} \rho^{ij}_n = \prod_{c=1}^G \frac{1}{\Sigma^c_n} \exp&\left\{-\frac{1}{2}\left(\frac{\overline{S}^c_n - a^c_j \overline{S}^c_{n-1}}{\Sigma^c_n}\right)^2 + \frac{1}{2}(\overline{S}^c_n)^2\right\} \\ &\times \left[\sum_{t,\ell=1}^K p_{\ell t} e_\ell \langle \rho^{ij}_{n-1}, e_t\rangle + p_{ji} e_j \zeta_{n-1}(i)\right]. \end{aligned} \tag{6.24}$$

Proof. The proof is similar to that of Theorem 4.1. □

7. Conclusion

In this paper, using hidden Markov model techniques, recursive filters for various quantities of interest related to an insurance model were derived. Formulae to predict future claims were established. The EM algorithm was used to update the parameters of the discussed model.

References

[1] L. E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, Ann. Math. Statist. **37** (1966), 1554–1563.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. Roy. Statist. Soc. Ser. B **39** (1977), no. 1, 1–38.

[3] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models. Estimation and Control*, Applications of Mathematics, vol. 29, Springer-Verlag, New York, 1995.

[4] J. H. Hossak, J. H. Pollard, and B. Zehnwirth, *Introductory Statistics with Applications in General Insurance*, 2nd ed., Cambridge University Press, Cambridge, 1999.

[5] T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels, *Stochastic Processes for Insurance and Finance*, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, 1999.

[6] C.-F. J. Wu, *On the convergence properties of the EM algorithm*, Ann. Statist. **11** (1983), no. 1, 95–103.

Lakhdar Aggoun: Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, P.O. Box 36, Al-Khod 123, Sultanate of Oman

E-mail address: laggoun@squ.edu.om