International Mathematical Forum, 1, 2006, no. 26, 1255 - 1272

An Insurance Model with Bonus-Malus System[1]

Lakhdar Aggoun

Department of Mathematics and Statistics
Sultan Qaboos University
P. O. Box 36, Al-Khod 123, Sultanate of Oman
aggounl@squ.edu.om

Lakdere Benkherouf

Department of Statistics and Operations Research
College of Science, Kuwait University
P.O.Box 5969, Safat 13060 Kuwait
lakdereb@kuc01.kuniv.edu.kw

Abstract. This paper considers a Markovian insurance model in which a policy-holder belongs to one of a finite number of classes (tariff groups). The class of a policy-holder in a given year is determined on the basis of the class of the previous year and on the number of claims reported in that year. If no claims have been reported, then the policy-holder gets a bonus expressed in moving to a class with lower premium rates. Else, the policy-holder may stay in the same class or gets maluses (penalized) by being shifted to a class with higher premium rates. The movement between classes is modeled as a discrete time Markov chain which is dependent on the claim process which is also assumed to be Markovian. However, the transition matrix of the claim process is assumed unknown and random. In this paper, we propose recursive filters for a related transition matrix. We also provide a formula to predict claims m years into the future. Finally, parameters estimation of the transition matrix governing the movement between classes are proposed.

Mathematics Subject Classification: 60G35, 62M05, 62M20, 91B30, 91B70

Keywords: Insurance, Policy-holder, Bonus-Malus, Markov chains, Change of Measure, EM algorithm

[1]This paper is in final form and no version of it will be submitted for publication elsewhere.

1. Introduction

All processes are defined on a measurable space $(\Omega, \mathcal{F})$, with probability measure P.

Consider a set of L policy-holders of automobile insurance. Each policy-holder belongs to one of a finite number C of classes (tariff groups) sorted by order of merit; class 1 being the one with lowest premiums etc. That is, each premium depends on the class to which a policy-holder belongs. Each year the class of a policy-holder is determined on the basis of the class of the previous year and on the number of claims reported during that year. If no claim has been reported, then the policy-holder gets a bonus expressed in the lowering to a class with a possibly lower premium. Otherwise the policy-holder may stay in the same class or gets maluses (penalized) by being shifted to a higher class with possibly higher premium. New policy-holders are assigned to a certain class.

The Bonus-Malus system is popular in Europe and is well established in the motor insurance markets. For more details on the Bonus-Malus system see, for instance, [5], [6] and [7].

Let x_m refers to the number of policy holders assigned to tariff group m, $m = 1, ..., C$ and let $X_n \in S_X = \{(x_1, \ldots, x_C)\}$ be the states of these groups during the $n-th$ year.

Let the nonnegative half of the real line be partitioned into a convenient set of disjoint intervals $I_1, \ldots, I_K$ and let Z_n^1 be the number of claims with sizes falling in I_1, ..., Z_n^K be the number of claims with sizes falling in I_K and $Z_n = (Z_n^1, \ldots, Z_n^K) \in S_Z$, where S_Z is a (finite) subset of $\mathbb{Z}_+^K$.

We shall be using the following notation and assumptions:

a. The processes X is a Markov chain which, for technical reasons (that will become apparent later) and without loss of generality, lives on the standard basis $\{e_1, \ldots, e_{|S_X|}\}$ of $\mathbb{R}^{|S_X|}$, where the vector e_i has 1 in the i-th position and 0's elsewhere and $|S_X|$ is the cardinal or size of the set S_X. The probability transition matrix of X is given below in representation (1.2).

b. The process Z is a Markov chain on the set of unit vectors $\{f_1, \ldots, f_{|S_Z|}\}$ of $\mathbb{R}^{|S_Z|}$, where the vector f_k has 1 in the k-th position and 0's elsewhere and $|S_Z|$ is the cardinal or size of the set S_Z. (See representation (1.3) below). However, the transition probability matrix of Z is time-dependent and not known a priori. In this paper, it is assumed that these matrices form a stochastic process. (See representation (1.4) below).

Write $\mathcal{I}_n = \sigma\{X_k, Z_{k-1}, k \leq n\}$ for the complete filtration generated by X and Z up to the $n-th$ year. We assume that:

$$P[X_n = e_j \mid \mathcal{I}_{n-1}, Z_{n-1}] = P[X_n = e_j \mid X_{n-1}, Z_{n-1}].$$

Conditional on the event $[X_{n-1} = e_i, Z_{n-1} = f_k]$ expression (**??**) may be rewritten:

$$P[X_n = e_j \mid X_{n-1} = e_i, Z_{n-1} = f_k] \overset{\triangle}{=} p_{j,ik}. \tag{1.1}$$

Write $P = \{p_{j,ik}\}$, $i, j = 1, \ldots, |S_X|$, $k = 1, \ldots, |S_Z|$. Then $\sum_{j=1}^{|S_X|} p_{j,ik} = 1$ and P is an $|S_X| \times |S_X|\ |S_Z|$ stochastic matrix of tensor mapping $\mathbb{R}^{|S_X||S_Z|}$ into $\mathbb{R}^{|S_X|}$ and has the form

$$P = \begin{bmatrix} p_{1,11} & p_{1,12} & \cdots & p_{1,|S_X||S_Z|} \\ \vdots & \vdots & \ldots & \vdots \\ p_{|S_X|,11} & p_{|S_X|,12} & \cdots & p_{|S_X|,|S_X||S_Z|} \end{bmatrix}.$$

Lemma 1. *The process X has the following semimartingale representation:*

$$X_n = PX_{n-1} \otimes Z_{n-1} + V_n, \tag{1.2}$$

where V_n is an $\mathcal{I}_n$ - martingale increment and $\otimes$ denotes the tensor product.

Proof. We must show that V is a sequence of martingale increments.

$$\begin{aligned} &E[V_n \mid \mathcal{I}_{n-1}] \\ &\quad = E[E[X_n - PX_{n-1} \otimes Z_{n-1} \mid \mathcal{I}_{n-1}, Z_{n-1}] \mid \mathcal{I}_{n-1}] \\ &\quad = E[E[X_n \mid \mathcal{I}_{n-1}, Z_{n-1}] \mid \mathcal{I}_{n-1}] - E[PX_{n-1} \otimes Z_{n-1} \mid \mathcal{I}_{n-1}] \\ &\quad = E[PX_{n-1} \otimes Z_{n-1} - PX_{n-1} \otimes Z_{n-1} \mid \mathcal{I}_{n-1}] \\ &\quad = 0, \end{aligned}$$

and the proof is finished. □

Set

$$q_n(s, r, h) \triangleq P[Z_n = f_s | Z_{n-1} = f_r, X_{n-1} = e_h],$$

and let $Q_n = \{q_n(s, r, h)\}$, $r, s = 1, \ldots, |S_Z|$, $h = 1, \ldots,\ |S_X|$ and $\sum_{s=1}^{|S_Z|} q_n(s, r, h) = 1$. That is, Q_n is an $|S_Z| \times |S_Z||S_X|$ time-dependent, stochastic matrix or tensor mapping $\mathbb{R}^{|S_Z||S_X|}$ into $\mathbb{R}^{|S_Z|}$. The same analysis as for the Markov chain X shows that the Markov chain Z has representation

$$Z_n = Q_n Z_{n-1} \otimes X_{n-1} + W_n. \tag{1.3}$$

Here $\{W_n\}$ is a sequence of martingale increments with respect to $\mathcal{I}_n$.

In standard models of Bonus-Malus systems (see for instance [6]) it is generally assumed that for a policy holder the number of reported claims are i.i.d. random variables. However, policy holders may choose not to report small claims to avoid being penalized. The presence of X_{n-1} expresses this fact.

d. Consider the simplex $U = \{(u_1, \ldots, u_{|S_Z|}) : u_i \geq 0 \text{ and } \sum_{i=1}^{|S_Z|} u_i = 1\}$. By definition each column of Q_n is a point in U. Let $U_1, \ldots, U_M$ be a partition of the set U, that is $U = U_1 \cup U_2 \cup \cdots \cup U_M$ and $U_i \cap U_j = \emptyset$, $i \neq j$. Let $\Theta = (U_1, \ldots, U_M)^{|S_X|\ |S_Z|}$, that is Θ is the Cartesian product of $|S_X||S_Z|$ copies of the ordered set $(U_1, \ldots, U_M)$. We define the following Markov chain $\widetilde{Q}_n$ on the set Θ as follows.

If the first column of Q_n is a point in U_{m_1}, the second column of Q_n is a point in U_{m_2}, ..., the last column of Q_n is a point in $U_{m_{|S_X||S_Z|}}$, then $\widetilde{Q}_n =$

$(U_{m_1}, U_{m_2}, \ldots, U_{m_{|S_X||S_Z|}})$, that is, $\widetilde{Q}_n$ keeps track only of the location of the columns of Q_n in $U_1, \ldots, U_M$.

Write $\Theta = \{\theta_1, \theta_2, \ldots, \theta_{M|S_X||S_Z|}\}$. We shall identify the ordered set Θ with the standard basis $\{b_1, b_2, \ldots, b_{M|S_X||S_Z|}\}$ of $\mathbb{R}^{M|S_X|\ |S_Z|}$. We assume that

$$P[\widetilde{Q}_n = b_q \mid \widetilde{Q}_{n-1} = b_p] \stackrel{\triangle}{=} a_{qp}.$$

Write $A = \{a_{qp}\}$, $p, q = 1, \ldots, M|S_X||S_Z|$ and $\sum_{q=1}^{M\ |S_X||S_Z|} a_{qp} = 1$. That is, A is an $M|S_X||S_Z| \times M\ |S_X||S_Z|$ stochastic matrix. Again using an analogous argument to that used above shows that we have the respresentation

$$\widetilde{Q}_n = A\widetilde{Q}_{n-1} + R_n, \tag{1.4}$$

where, as before, $\{R_n\}$ is a sequence of martingale increments with respect to the $\sigma-$field $\sigma\{\widetilde{Q}_k,\ k \leq n\}$.

The model discussed in this paper can be seen as a generalization of existing models in the literature. It allows for some dependence between claims which are usually assumed i.i.d. Another perhaps more important aspect of this paper is the exposition of the so-called measure change techniques (or Hidden Markov Models) which have become very popular in stochastic modeling.

The objective of this paper is to predict the number of claims m years into the future. However these estimates depend on an unknown, time-dependent, transition matrix $\widetilde{Q}_n$ for which we provide recursive estimates.

Initially we start with a guess on the initial probability distribution of $\widetilde{Q}_n$. As time goes information about the evolution of the insurance model is gathered in $\mathcal{I}_n$. Based on this information we set to update our knowledge about $\widetilde{Q}_n$. This is done in the next section where a 'reference' probability measure, under which the processes X and Z become independent and identically distributed, is introduced. In Section 3 we provide a formula to predict future claims m years into the future. Again all calculations are performed under the 'reference' probability measure introduced in Section 2. In Section 4 we propose the so-called Expectation Maximization (EM) to estimate the transition probabilities of the Markov chain X. Finally in Section 5 alternative estimates for this transition matrix are proposed. At each time n these estimates are expressed as the sum of an estimate at time $(n-1)$ plus a correction based on new information.

2. Reference Probability

In order to estimate recursively the states of the 'hidden' process $\widetilde{Q}$ we choose a measure $P^{\dagger}$, on the measurable space $(\Omega, \mathcal{F})$, under which processes X and Z are two sequences of statistically independent and identically distributed random variable. The probability measure P is referred to as the 'real world' measure, that is, under this measure

$$P \qquad \begin{cases} X_n = PX_{n-1} \otimes Z_{n-1} + V_n, \\ Z_n = Q_n Z_{n-1} \otimes X_{n-1} + W_n, \\ \widetilde{Q}_n = A\widetilde{Q}_{n-1} + R_n. \end{cases} \tag{2.1}$$

Suppose that under the measure $P^\dagger$ processes X, Z are i.i.d with the following distributions

$$\begin{aligned} P^\dagger[X_n = e_j \mid \mathcal{I}_{n-1}, Z_{n-1}] &= \frac{1}{|S_X|} \\ P^\dagger[Z_n = f_k \mid \mathcal{I}_n] &= \frac{1}{|S_Z|}. \end{aligned}$$

Further, under the measure $P^\dagger$, the dynamics of $\widetilde{Q}_n$ are unchanged.

Define

$$\Lambda_n = \prod_{m=1}^{n} \lambda_m, \tag{2.2}$$

where $\lambda_0 = 1$ and

$$\begin{aligned} \lambda_m &= \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_m, e_j\rangle \langle X_{m-1}, e_i\rangle \langle Z_{m-1}, f_k\rangle} \\ &\quad \times \prod_{r,s=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (|S_Z| q_m(s,r,h))^{\langle Z_m, f_s\rangle \langle Z_{m-1}, f_r\rangle \langle X_{m-1}, e_h\rangle}, \end{aligned} \tag{2.3}$$

where $\langle .,. \rangle$ refers to the usual scalar product. It can be shown ([4, 1]) that Λ is a positive martingale with mean 1. Hence one can define (via Kolmogorov Extension Theorem) the 'real world' measure P in terms of $P^\dagger$, by setting

$$\left.\frac{dP}{dP^\dagger}\right|_{\mathcal{G}_n} \triangleq \Lambda_n. \tag{2.4}$$

Here $\mathcal{G}_n$ is the complete filtration generated by $\{X_k, Z_{k-1}, \widetilde{Q}_k, k \leq n\}$.

Lemma 2. *Under probability measure P, as defined from $P^\dagger$ via (2.4), the dynamics in (2.1) hold.*

Proof. We give the proof only for the process X.

Using a version of Bayes' Theorem [4]

$$\begin{aligned} & P(X_n = e_j \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k) \\ = \; & \frac{E^\dagger[\langle X_n, e_j\rangle \Lambda_n \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k]}{E^\dagger[\Lambda_n \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k]} \\ = \; & E^\dagger[\langle X_n, e_j\rangle \lambda_n \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k], \end{aligned}$$

In view of (2.3) we see that

$$
\begin{aligned}
& E^\dagger[\langle X_n, e_j\rangle \lambda_n \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k] \\
= \; & |S_X| p_{jik} |S_Z| \sum_{s=1}^{|S_Z|} E^\dagger[\langle X_n, e_j\rangle q_n(s,k,i) \langle Z_n, f_s\rangle \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k] \\
= \; & p_{jik} E^\dagger[\sum_{s=1}^{|S_Z|} q_n(s,k,i) \mid \mathcal{G}_{n-1}, X_{n-1} = e_i, Z_{n-1} = f_k] \\
= \; & p_{jik}
\end{aligned}
$$

where we used the distribution of Z_n and X_n under $P^\dagger$ and the fact that $\sum_{s=1}^{|S_Z|} q_n(s,k,i) = 1$. This finishes the proof. □

Define the measure-valued process

$$
g_n(w) = E^\dagger[\Lambda_n \langle \widetilde{Q}_n, b_w\rangle \mid \mathcal{I}_n]. \tag{2.5}
$$

Remark 1. *By a generalized version of Bayes' Theorem* [4]

$$
P(\widetilde{Q}_n = b_w \mid \mathcal{I}_n) = \frac{g_n(w)}{\displaystyle\sum_{u=1}^{M|S_X||S_Z|} g_n(u)}.
$$

Theorem 1. *Denote by $g_0(w)$, the initial probability of $\widetilde{Q}_n$. The unnormalised probability $g_n(w) \in \mathbb{R}_+$, satisfies the recursion*

$$
\begin{aligned}
& g_n(w) \\
& = |S_X| \langle P X_{n-1} \otimes Z_{n-1}, X_n\rangle \sum_{s=1}^{|S_Z|} \langle \widetilde{Q}_w Z_{n-1} \otimes X_{n-1}, f_s\rangle \sum_{v=1}^{M|S_X||S_Z|} a_{wv} g_{n-1}(v)
\end{aligned}
$$

Proof. In view of the special form of the martingale Λ as defined in (2.2) and (2.3) we see that (2.5) is equal to

$$
\begin{aligned}
& E^{\dagger}[\Lambda_n\langle \widetilde{Q}_n, b_w\rangle \mid \mathcal{I}_n] \\
= \; & E^{\dagger}[\Lambda_{n-1}\langle \widetilde{Q}_n, b_w\rangle \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
& \times \prod_{r,s=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_n(s,r,h))^{\langle Z_n,f_s\rangle\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \mid \mathcal{I}_n] \\
= \; & \sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\
& \times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
& \times E^{\dagger}[\Lambda_{n-1}\langle \widetilde{Q}_n, b_w\rangle\langle Z_n, f_s\rangle \mid \mathcal{I}_n]
\end{aligned}
$$

Here, we use the distribution of Z_n under $P^{\dagger}$

$$
\begin{aligned}
= \; & \frac{1}{|S_Z|}\sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\
& \times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
& \times E^{\dagger}[\Lambda_{n-1}\langle \widetilde{Q}_n, b_w\rangle \mid \mathcal{I}_n].
\end{aligned}
$$

Now recall the dynamics form (1.4) of $\widetilde{Q}_n$.

$$
\begin{aligned}
& E^{\dagger}[\Lambda_n\langle \widetilde{Q}_n, b_w\rangle \mid \mathcal{I}_n] \\
& = \frac{1}{|S_Z|}\sum_{v=1}^{M|S_X||S_Z|} a_{wv}\sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\
& \times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
& \times E^{\dagger}[\Lambda_{n-1}\langle \widetilde{Q}_{n-1}, b_v\rangle \mid \mathcal{I}_{n-1}]
\end{aligned}
$$

By using the notation in (2.5), the above is equal to

$$
\begin{aligned}
&\sum_{v=1}^{M|S_X||S_Z|} a_{wv} \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle \langle X_{n-1}, e_h\rangle} \\
&\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle \langle X_{n-1}, e_i\rangle \langle Z_{n-1}, f_k\rangle} g_{n-1}(v),
\end{aligned}
$$

which finishes the proof. □

3. Predicting The Number an Size of future Claims

It is important to insurance companies to be able to predict the total amount to be claimed by policy-holders m years into the future. As a first step we shall derive a formula to predict the amount (or more precisely the interval which will contain the total amount) to be claimed by a tariff group of policy-holders at the end of the current year. That is, we wish to compute the conditional probability of Z_n given the history up to the beginning of the current year.

By Bayes' Theorem

$$
P(Z_n = f_s \mid \mathcal{I}_n) = \frac{E^\dagger[\langle Z_n, f_s\rangle \Lambda_n \mid \mathcal{I}_n]}{E^\dagger[\Lambda_n \mid \mathcal{I}_n]},
$$

where the denominator is a constant.

Lemma 3. *The (unnormalised) conditional probability distribution of the amount to be claimed by a given risk group of policy-holders at the end of the current year given the history $\mathcal{I}_n$ up to the beginning of this same year is*

$$
\begin{aligned}
&E^\dagger[Z_n \Lambda_n \mid \mathcal{I}_n] \\
&= |S_X| \langle P X_{n-1} \otimes Z_{n-1}, X_n\rangle \sum_{s=1}^{|S_Z|} \sum_{w=1}^{M|S_X||S_Z|} \langle \widetilde{Q}_w Z_{n-1} \otimes X_{n-1}, f_s\rangle f_s \, g_n(w).
\end{aligned}
$$

Proof. We have

$$
\begin{aligned}
&E^\dagger[Z_n \Lambda_n \mid \mathcal{I}_n] \\
&\sum_{s=1}^{|S_Z|} f_s E^\dagger[\langle Z_n, f_s\rangle \Lambda_n \mid \mathcal{I}_n] \\
&= \sum_{s=1}^{|S_Z|} f_s \sum_{w=1}^{M|S_X||S_Z|} E^\dagger[\langle \widetilde{Q}_n, b_w\rangle \langle Z_n, f_s\rangle \Lambda_n \mid \mathcal{I}_n].
\end{aligned}
$$

Now, the proof of Theorem 1 implies that the above is equal to:

$$\prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle}$$
$$\times\sum_{s=1}^{|S_Z|}f_s\sum_{w=1}^{M|S_X||S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle}g_n(w)$$
$$=|S_X|\langle PX_{n-1}\otimes Z_{n-1},X_n\rangle\sum_{s=1}^{|S_Z|}\sum_{w=1}^{M|S_X||S_Z|}\langle\widetilde{Q}_wZ_{n-1}\otimes X_{n-1},f_s\rangle f_sg_n(w).$$

□

Now we wish to derive the joint conditional probability distribution of $\langle Z_n,f_{s_0}\rangle$ and $\langle Z_{n+1},f_{s_1}\rangle$ given the history up to the beginning of the n-th year. Again we consider the unnormalised version:

$$E^\dagger[\langle Z_n,f_{s_0}\rangle\langle Z_{n+1},f_{s_1}\rangle\Lambda_{n+1}\mid\mathcal{I}_n]$$
$$=\sum_{w_0,w_1=1}^{M|S_X||S_Z|}\sum_{j_1=1}^{|S_X|}E^\dagger[\langle Z_n,f_{s_0}\rangle\langle Z_{n+1},f_{s_1}\rangle\langle X_{n+1},e_{j_1}\rangle\langle\widetilde{Q}_n,b_{w_0}\rangle\langle\widetilde{Q}_{n+1},b_{w_1}\rangle\Lambda_{n+1}\mid\mathcal{I}_n]$$
$$=\sum_{w_0,w_1=1}^{M|S_X||S_Z|}\sum_{j_1=1}^{|S_X|}\prod_{i=1}^{|S_X|}(|S_X|p_{j_1,is_0})^{\langle X_n,e_i\rangle}\prod_{h=1}^{|S_X|}(|S_Z|q_{w_1}(s_1,s_0,h))^{\langle X_n,e_h\rangle}$$
$$\times E^\dagger[\langle X_{n+1},e_{j_1}\rangle\langle\widetilde{Q}_n,b_{w_0}\rangle\langle Z_n,f_{s_0}\rangle\langle A\widetilde{Q}_n,b_{w_1}\rangle\langle Z_{n+1},f_{s_1}\rangle\Lambda_n\mid\mathcal{I}_n].$$

Note that Z_n, X_{n+1} and Z_{n+1} are not in $\mathcal{I}_n$, therefore we use their distributions under $P^\dagger$. This means that the above is equal to

$$\sum_{w_0,w_1=1}^{M|S_X||S_Z|}\sum_{j_1=1}^{|S_X|}\prod_{i=1}^{|S_X|}(p_{j_1,is_0})^{\langle X_n,e_i\rangle}\frac{1}{K^2}\prod_{h=1}^{|S_X|}(|S_Z|q_{w_1}(s_1,s_0,h))^{\langle X_n,e_h\rangle}$$
$$\times E^\dagger[\langle\widetilde{Q}_n,b_{w_0}\rangle\langle Ab_{w_0},b_{w_1}\rangle\Lambda_n\mid\mathcal{I}_n].$$

Noting that $\langle Ab_{w_0},b_{w_1}\rangle=a_{w_1w_0}$ and using (2.5) we get

$$\sum_{w_0,w_1=1}^{M|S_X||S_Z|}\sum_{j_1=1}^{|S_X|}\prod_{i=1}^{|S_X|}(p_{j_1,is_0})^{\langle X_n,e_i\rangle}\frac{1}{|S_Z|}\prod_{h=1}^{|S_X|}(q_{w_1}(s_1,s_0,h))^{\langle X_n,e_h\rangle}$$
$$\times a_{w_1w_0}g_n(w_0)$$
$$=\sum_{w_0,w_1=1}^{M|S_X||S_Z|}\sum_{j_1=1}^{|S_X|}\frac{1}{|S_Z|}\prod_{i=1}^{|S_X|}(p_{j_1,is_0}q_{w_1}(s_1,s_0,i))^{\langle X_n,e_i\rangle}a_{w_1w_0}g_n(w_0).$$

The same argument leads to the m-th step predictor

Theorem 2. *We have*

$$
\begin{aligned}
&E^{\dagger}[\Lambda_{n+m}\prod_{t=0}^{m}\langle Z_{n+t}, f_{s_t}\rangle \mid \mathcal{I}_n] \\
&= \sum_{w_0,\ldots,w_m=1}^{M|S_X||S_Z|}\ \sum_{j_1,\ldots,j_m=1}^{|S_X|} \frac{1}{|S_Z|}\prod_{i=1}^{|S_X|}(p_{j_1,is_0}q_{w_1}(s_1,s_0,i))^{\langle X_n,e_i\rangle} \\
&\times \prod_{t=1}^{m} p_{j_{t+1},j_t s_t} q_{w_{t+1}}(s_{t+1},s_t,j_t)a_{w_t w_{t-1}} g_n(w_0).
\end{aligned}
$$

4. Parameter Re-estimation

In this section we show how, using the Expectation Maximization (EM) algorithm, the parameters of the model can be estimated. In fact it is a conditional pseudo *log-likelihood* which is maximized, and the new parameters are expressed in terms of the recursive estimates obtained in Section 6. We begin by describing the EM algorithm.

The basic idea behind the EM algorithm is as follows ([2]). Let $\{P_\theta, \theta \in \Theta\}$ be a family of probability measures on a measurable space $(\Omega, \mathcal{F})$ all absolutely continuous with respect to a fixed probability measure P_0 and let $\mathcal{I} \subset \mathcal{F}$. The likelihood function for computing an estimate of the parameter θ based on the information available in $\mathcal{I}$ is

$$L(\theta) = E_0\left[\frac{dP_\theta}{dP_0} \mid \mathcal{I}\right],$$

and the maximum likelihood estimate (MLE) is defined by

$$\hat{\theta} \in \text{argmax }_{\theta\in\Theta} L(\theta).$$

In general, the MLE is difficult to compute directly, and the EM algorithm provides an iterative approximation method:

Step 1. Set $p = 1$ and choose $\hat{\theta}_0$.

Step 2. (E-step) Set $\theta* = \hat{\theta}_p$ and compute $Q(\cdot, \theta*)$, where

$$Q(\theta, \theta*) = E_{\theta*}\left[\log\frac{dP_\theta}{dP_{\theta*}} \mid \mathcal{I}\right].$$

Step 3. (M-step) Find

$$\hat{\theta}_{p+1} \in \text{argmax }_{\theta\in\Theta} Q(\theta, \theta*)$$

Step 4. Replace p by $p+1$ and repeat beginning with Step 2 , until a stopping criterion is satisfied.

The sequence generated $\{\hat{\theta}_p, p \geq 0\}$ gives non-decreasing values of the likelihood function: indeed, it follows from Jensen's inequality (see Elliott [3]) that

$$\log L(\hat{\theta}_{p+1}) - \log L(\hat{\theta}_p) \geq Q(\hat{\theta}_{p+1}, \hat{\theta}_p),$$

with equality if and only if $\hat{\theta}_{p+1} = \hat{\theta}_p$. We call $Q(\theta, \theta*)$ a conditional pseudo-log-likelihood.

Sufficient conditions for convergence of the EM algorithm are given in [8].

We wish to update the set of parameters

$$\theta := (a_{wv},\ 1 \le v, w \le M|S_X||S_Z|),$$

which is subject to the constraints $\sum_{w=1}^{M|S_X||S_Z|} a_{wv} = 1$. Suppose our model is determined by such a set θ and we wish to determine a new set

$$\widehat{\theta} = (\hat{a}_{wv}(k),\ 1 \le i, j \le M|S_X||S_Z|),$$

which maximizes the conditional pseudo-log-likelihoods defined below.

To replace the parameters a_{wv} by $\hat{a}_{wv}(n)$ in the Markov chain $\widetilde{Q}$ we define

$$\Gamma_n = \prod_{m=1}^{n} \prod_{v,w=1}^{M|S_X||S_Z|} \Big(\frac{\hat{a}_{wv}(n)}{a_{wv}}\Big)^{\langle \widetilde{Q}_m, b_w\rangle \langle \widetilde{Q}_{m-1}, b_v\rangle},$$

and set

$$\frac{dP_{\hat{\theta}}}{dP_\theta}\Big|_{\mathcal{F}_n} = \Gamma_n$$

Notation 1. *The number of jumps of the Markov chain $\widetilde{Q}$ from state b_v to state b_w in time n is given by $\mathcal{J}^{vw}(n) = \sum_{m=1}^{n} \langle \widetilde{Q}_{m-1}, b_v\rangle \langle \widetilde{Q}_m, b_w\rangle$.*

Theorem 3. *The new estimate of the parameter $\hat{a}_{wv}(n)$ given the observations up to time n are given by*

$$\hat{a}_{wv}(n) = \frac{E^\dagger[\Lambda_n \mathcal{J}^{vw}(n) \mid \mathcal{I}_n]}{\sum_{w=1}^{M|S_X||S_Z|} E^\dagger[\Lambda_n \mathcal{J}^{vw}(n) \mid \mathcal{I}_n]} \triangleq \frac{\gamma_n(\mathcal{J}^{vw}(n))}{\sum_{w=1}^{M|S_X|\ |S_Z|} \gamma_n(\mathcal{J}^{vw}(n))}.$$

Proof.

$$\begin{aligned}\log \Lambda_k &= \sum_{v,w=1}^{M|S_X||S_Z|} \sum_{m=1}^{n} \langle \widetilde{Q}_m, b_w\rangle \langle \widetilde{Q}_{m-1}, b_v\rangle (\log \hat{a}_{wv}(n) - \log a_{wv}) \\ &= \sum_{v,w=1}^{M|S_X||S_Z|} \mathcal{J}^{vw}(n) \log \hat{a}_{wv}(n) + R(a),\end{aligned}$$

where $R(a)$ is independent of $\hat{a}$. Therefore,

$$E[\log \Lambda_n \mid \mathcal{I}_n] = \sum_{v,w=1}^{M|S_X||S_Z|} E[\mathcal{J}^{vw}(n) \mid \mathcal{I}_n] \log \hat{a}_{wv}(n) + \hat{R}(a). \tag{4.1}$$

Now the $\hat{a}_{wv}(n)$ must also satisfy

$$\sum_{w=1}^{M|S_X||S_Z|} \hat{a}_{wv}(n) = 1. \tag{4.2}$$

We wish, therefore, to choose the $\hat{a}_{wv}(n)$ to maximize (4.1) subject to the constraint (4.2). Write λ for the Lagrange multiplier and put

$$L(\hat{a},\lambda) = \sum_{v,w=1}^{M|S_X||S_Z|} E[\mathcal{J}^{vw}(n) \mid \mathcal{I}_n] \log \hat{a}_{wv}(n) + \hat{R}(a) + \lambda\Big(\sum_{w=1}^{M|S_X||S_Z|} \hat{a}_{wv}(n) - 1\Big).$$

Differentiating in λ and $\hat{a}_{wv}(n)$, and equating the derivatives to 0 gives the result. □

Remark 2. *A closed form, finite-dimensional recursion is only possible for the conditional joint distributions of $\mathcal{J}^{vw}(n)$ and $\widetilde{Q}_n$. That is we shall consider recursive filters for $E^\dagger[\Lambda_n \mathcal{J}^{vw}(n)\widetilde{Q}_n \mid \mathcal{I}_n] \stackrel{\triangle}{=} \rho^{\ vw}(n)$. However $\gamma_n(\mathcal{J}^{vw}(n)) = \sum_u \langle \rho^{\ vw}(n), b_u\rangle$.*

Theorem 4. *Let $\rho^{\ vw}(0)$ be the initial distribution of $\mathcal{J}^{vw}(0)\widetilde{Q}_0$ and for $n \geq 1$*

$$\begin{aligned}
&\rho^{\ vw}(n)\\
&= \sum_{t,u=1}^{M|S_X||S_Z|} a_{ut} b_u \sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|} (q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle}\\
&\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|} (|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \langle \rho^{\ vw}(n-1), b_t\rangle\\
&+ b_w a_{wv} \sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|} (q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle}\\
&\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|} (|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} g_{n-1}(v).
\end{aligned}$$

Proof. First note that $\mathcal{J}^{vw}(n) = \mathcal{J}^{vw}(n-1) + \langle \widetilde{Q}_{n-1}, b_v\rangle\langle \widetilde{Q}_n, b_w\rangle$. Therefore

$$\begin{aligned}
&E^\dagger[\Lambda_n \mathcal{J}^{vw}(n)\widetilde{Q}_n \mid \mathcal{I}_n] = E^\dagger[\Lambda_n \mathcal{J}^{vw}(n-1)\widetilde{Q}_n \mid \mathcal{I}_n]\\
&+ E^\dagger[\Lambda_n \langle \widetilde{Q}_{n-1}, b_v\rangle\langle \widetilde{Q}_n, b_w\rangle \widetilde{Q}_n \mid \mathcal{I}_n].
\end{aligned}$$

However, using the fact that $\sum_{u=1}^{M|S_X||S_Z|}\langle \widetilde{Q}_n, b_u\rangle = 1$ and in view of (1.4), (2.2) and (2.3)

$$
\begin{aligned}
E^{\dagger}[\Lambda_n \mathcal{J}^{vw}(n-1)\widetilde{Q}_n \mid \mathcal{I}_n] &= \sum_{u=1}^{M|S_X||S_Z|} b_u E^{\dagger}[\Lambda_n \mathcal{J}^{vw}(n-1)\langle \widetilde{Q}_n, b_u\rangle \mid \mathcal{I}_n] \\
= \sum_{u=1}^{M|S_X||S_Z|} & b_u E^{\dagger}[\Lambda_{n-1}\mathcal{J}^{vw}(n-1) \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
\times \prod_{r,s=1}^{|S_Z|}\prod_{h=1}^{|S_X|} & (|S_Z|q_n(s,r,h))^{\langle Z_n,f_s\rangle\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle}\langle A\widetilde{Q}_{n-1}, b_u\rangle \mid \mathcal{I}_n] \\
= \sum_{u=1}^{M|S_X||S_Z|} & b_u \sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{k=1}^{|S_Z|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\
\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|} & (|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
\times E^{\dagger}[\Lambda_{n-1} & \langle \mathcal{J}^{vw}(n-1)\langle Z_n, f_s\rangle\langle A\widetilde{Q}_{n-1}, b_u\rangle \mid \mathcal{I}_n],
\end{aligned}
$$

where we used the fact that Z_{n-1}, X_{n-1}, X_n are all $\mathcal{I}_n$ measurable to get

$$
\begin{aligned}
&\frac{1}{|S_Z|}\sum_{u=1}^{M|S_X||S_Z|} b_u \sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\
&\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
&\times E^{\dagger}[\Lambda_{n-1}\mathcal{J}^{vw}(n-1)\langle A\widetilde{Q}_{n-1}, b_u\rangle \mid \mathcal{I}_{n-1}],
\end{aligned}
$$

after using the uniform distribution and independence of Z_n under $P^{\dagger}$. This leads to

$$
\begin{aligned}
&\frac{1}{|S_Z|}\sum_{t,u=1}^{M|S_X||S_Z|} a_{ut} b_u \sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\
&\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\
&\times E^{\dagger}[\Lambda_{n-1}\mathcal{J}^{vw}(n-1)\langle \widetilde{Q}_{n-1}, b_t\rangle \mid \mathcal{I}_{n-1}],
\end{aligned}
$$

because

$$\sum_{t=1}^{M|S_X||S_Z|} \langle \widetilde{Q}_{n-1}, b_t\rangle = 1)$$

$$= \sum_{t,u=1}^{M|S_X||S_Z|} a_{ut} b_u \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle\langle X_{n-1}, e_h\rangle}$$

$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle\langle X_{n-1}, e_i\rangle\langle Z_{n-1}, f_k\rangle} \langle \rho^{\,vw}(n-1), b_t\rangle,$$

using the notation in Remark 2 and, using the same assumptions,

$$E^\dagger[\Lambda_n \langle \widetilde{Q}_{n-1}, b_v\rangle \langle \widetilde{Q}_n, b_w\rangle \widetilde{Q}_n \mid \mathcal{I}_n] = b_w E^\dagger[\Lambda_n \langle \widetilde{Q}_{n-1}, b_v\rangle \langle \widetilde{Q}_n, b_w\rangle \mid \mathcal{I}_n]$$

$$= \frac{b_w}{|S_Z|} \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (|S_Z| q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle\langle X_{n-1}, e_h\rangle}$$

$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle\langle X_{n-1}, e_i\rangle\langle Z_{n-1}, f_k\rangle}$$

$$\times E^\dagger[\Lambda_{n-1} \langle \widetilde{Q}_{n-1}, b_v\rangle \langle A\widetilde{Q}_{n-1}, b_w\rangle \mid \mathcal{I}_{n-1}]$$

$$= \frac{b_w}{|S_Z|} \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (|S_Z| q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle\langle X_{n-1}, e_h\rangle}$$

$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle\langle X_{n-1}, e_i\rangle\langle Z_{n-1}, f_k\rangle}$$

$$\times E^\dagger[\Lambda_{n-1} \langle \widetilde{Q}_{n-1}, b_v\rangle \langle A b_w, b_w\rangle \mid \mathcal{I}_{n-1}]$$

$$= b_w a_{wv} \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle\langle X_{n-1}, e_h\rangle}$$

$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle\langle X_{n-1}, e_i\rangle\langle Z_{n-1}, f_k\rangle} g_{n-1}(v),$$

which finishes the proof. $\square$

5. Recursive Parameter Estimation

In Section 4 we proposed estimates for the matrix A via the EM algorithm. However, these are not expressed as the estimates at time $(n-1)$ plus a correction based on new information. This in fact if available should make computations simple. We shall next derive such estimates Write $A = A(\theta)$, and suppose that θ is not known a priori. Let us estimate θ, given the observations $\mathcal{I}_n$. We assume that θ will take values in some measure space (Γ, β, μ).

Let us now write $\mathcal{G}_n$ for the complete σ-field generated by knowledge of X, Z, $\widetilde{Q}$ together with θ. With this enlarged $\mathcal{G}_n$ the results of the previous sections still hold. We shall be working under the same probability measure $P^\dagger$ on $\left(\Omega, \bigvee_{m=1}^{\infty} \mathcal{G}_m\right)$ defined in Section 2. Write $\sigma_n(\theta)$ for the unnormalised, conditional density such that
$E^\dagger[\Lambda_n\langle\widetilde{Q}_n, b_w\rangle I(\theta \in d\theta) \mid \mathcal{I}_n] = \sigma_n(w,\theta)d\theta.$
Here, $I(A)$ is the indicator function of the set A, that is, the function which is 1 on A and 0 otherwise. The existence of $\sigma_n(w,\theta)$ will be discussed below. The normalized conditional density is simply
$$E[\langle\widetilde{Q}_n, b_w\rangle I(\theta \in d\theta) \mid \mathcal{Y}_k] = \frac{\sigma_n(w,\theta)}{\sum_{w=1}^{M|S_X||S_Z|}\int_\Gamma \sigma_n(w,\phi)d\mu(\phi)}.$$

A recursive expression for $\sigma_n(w,\theta)$ is now obtained:

Theorem 5. *The recursive estimates of the unnormalised joint conditional distribution of $\widetilde{Q}_n$ and θ are given by*

$$\begin{aligned}\sigma_n(w,\phi) &= |S_X|\langle PX_{n-1}\otimes Z_{n-1}, X_n\rangle \\ &\quad\times\sum_{s=1}^{|S_Z|}\langle\widetilde{Q}_w Z_{n-1}\otimes X_{n-1}, f_s\rangle \sum_{v=1}^{M|S_X||S_Z|} a_{wv}\sigma_{n-1}(v,\phi).\end{aligned} \tag{5.1}$$

Proof. Suppose g is any real valued Borel function on Γ. Then

$$E^\dagger[\langle\widetilde{Q}_n, b_w\rangle g(\theta)\Lambda_n \mid \mathcal{I}_n] = \int_\Gamma \sigma_n(w,\phi)g(\phi)d\mu(\phi). \tag{5.2}$$

However, in view of (1.4), (2.2) and (2.3) the left hand side of (5.2) is simply

$$\begin{aligned}&= E^\dagger[\Lambda_{n-1}\langle A\widetilde{Q}_{n-1}, b_w\rangle \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle} \\ &\times \prod_{r,s=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_n(s,r,h))^{\langle Z_n,f_s\rangle\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} g(\theta) \mid \mathcal{I}_n] \\ &= \frac{1}{|S_Z|}\sum_{v=1}^{M|S_X||S_Z|} a_{wv}\sum_{s=1}^{|S_Z|}\prod_{r=1}^{|S_Z|}\prod_{h=1}^{|S_X|}(|S_Z|q_w(s,r,h))^{\langle Z_{n-1},f_r\rangle\langle X_{n-1},e_h\rangle} \\ &\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_Z|}(|S_X|p_{jik})^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle Z_{n-1},f_k\rangle}\end{aligned}$$

$$\times E^{\dagger}[\Lambda_{n-1}\langle \widetilde{Q}_{n-1}, b_v\rangle g(\theta) \mid \mathcal{I}_{n-1}]$$
$$= \frac{1}{|S_Z|} \sum_{v=1}^{M|S_X||S_Z|} a_{wv} \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (|S_Z| q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle \langle X_{n-1}, e_h\rangle}$$
$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle \langle X_{n-1}, e_i\rangle \langle Z_{n-1}, f_k\rangle} \int_{\Gamma} \sigma_{n-1}(v,\phi) g(\phi) d\mu(\phi).$$

As g is arbitrary we see

$$\sigma_n(w,\phi) = \frac{1}{|S_Z|} \sum_{v=1}^{M|S_X||S_Z|} a_{wv} \sum_{s=1}^{|S_Z|} \prod_{r=1}^{|S_Z|} \prod_{h=1}^{|S_X|} (|S_Z| q_w(s,r,h))^{\langle Z_{n-1}, f_r\rangle \langle X_{n-1}, e_h\rangle}$$
$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_Z|} (|S_X| p_{jik})^{\langle X_n, e_j\rangle \langle X_{n-1}, e_i\rangle \langle Z_{n-1}, f_k\rangle} \sigma_{n-1}(v,\phi),$$

which is the result. □

Compared with Theorem 3 the new feature of Theorem 5 is that it up-dates recursively the estimate of the parameter.

Remark 3. *Suppose* $\pi = (\pi_1, \ldots, \pi_{M|S_X||S_Z|})$, *where* $\pi_w = P(\widetilde{Q}_0 = b_w)$, *is the initial distribution for* $\widetilde{Q}_0$, *and* $h(\theta)$, *is the prior density for* θ. *Then*

$$\sigma_0(w,\theta) = \pi_w h(\theta),$$

and the updated estimates are given by (5.1).

If the prior information about $\widetilde{Q}_0$ is that, say, $\widetilde{Q}_0 = b_w$, then the dynamics of $\widetilde{Q}$, (1.4) will move the state around and the estimate is given by (5.1). If the prior information about θ, is that θ takes a particular value then $h(\theta)$, (or a factor of h), is a delta function at this value. No noise or dynamics enters into θ, so the equations (5.1) just continue to give the delta function at this value. This is exactly to be expected. The prior distribution h taken for θ must represent the a priori information about θ; it is not an initial guess for the value of θ.

Finally, we note that the equations (5.1) are really just a family of equations parametrized by θ. In particular, if θ can take one of finitely many values $\theta_1, \theta_2, \ldots, \theta_p$ we obtain p equations (5.1) for each possible θ_i. The prior for θ is then just a distribution over $\theta_1, \ldots, \theta_p$.

Remark 4. *It could be of interest to look at a continuous-state version of (1.3) as given, for instance, by the following (conditional) probability distribution of* *the totality of the claims reported by a (risk) group.*

$$P(Z_n^1 \in dx \mid Z_{n-1}^1 = z, X_{n-1} = e_i)$$
$$= \frac{1}{\Psi_n^i \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{x - g_n^i z}{\Psi_n^i} \right)^2 \right\} dz.$$

Let $\lambda_n \in \mathbb{R}_+$ *be a stochastic process with transition density*

$$P(\lambda_n \in du \mid \lambda_{n-1} = \lambda, Z_{n-1}^2 = \ell, X_{n-1} = e_i)$$
$$= \frac{1}{\sigma_n^i \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{u - b_n^i \lambda}{\sigma_n^i} \right)^2 \right\} dz.$$

$$P(Z_n^2 = m \mid \lambda_n, X_{n-1} = e_i)$$
$$= \exp\{-\lambda_n\} \frac{\lambda_n^m}{m!},$$

In this case (1.1) takes the form

$$P[X_n = e_j \mid X_{n-1} = e_i, Z_{n-1}^1 = z, Z_{n-1}^2 = \ell] \stackrel{\triangle}{=} p_{j,i}(z, \ell),$$

that is, given the event $[Z_{n-1}^1 = z, Z_{n-1}^2 = \ell]$*,* X *is a Markov chain with probability transition matrix function of* z *and* ℓ*.*

Using the same techniques many quantities of interest could be derived.

6. Conclusion

In this paper, we considered a Markovian insurance model in which a policyholder belongs to one of a finite number of classes (tariff groups). The movement between classes is modeled as a discrete time Markov chain which is dependent on the claim process which is also assumed to be Markovian. However, the transition matrix of the claim process is assumed unknown and random. We proposed recursive filters for a related transition matrix. We also provide a formula to predict claims m years into the future. A probability transitions matrix is estimated via the EM algorithm and recursively.

References

[1] Aggoun L., Elliott R.J., (2004) *Measure Theory and Filtering: Introduction with Applications*. Cambridge Series In Statistical and Probabilistic Mathematics.

[2] Dempster, A.P., N.M. Laird and D.B. Rubin, *Maximum Likelihood from Incomplete Data via the EM algorithm*, Jour. of the Royal Statistical Society, B, 39, pp. 1-38, 1977.

[3] Elliott, R.J. 1982 *Stochastic Calculus and Applications*. Applications of Mathematics Vo l.18. Springer-Verlag, Berlin-Heidelberg-New York.

[4] Elliott, R.J. Aggoun, L. and Moore, J.B., *Hidden Markov Models:Estimation and Control*, Springer-Verlag, New-York, Applications of Mathematics No. 29, 1995.

[5] Hossak, J.H., J.H. Pollard and B. Zehnwirth *Introductory Statistics with Applications in General Insurance*, Cambridge University Press, Second Edition, 1999.

[6] Lemaire, J. *Bonus-Malus Systems in Automobile Insurance.*, Kluwer Academic Publishing, Boston, 1995.

[7] Rolski, T., Schmidli, H, Volker, S.and Teugels, J. *Stochastic Processes for Insurance and Finance*, Wiley Series in Probability and Statistics, 2000.

[8] Wu, C.F.J., *On the Convergence Properties of the EM algorithm*, The Annals of Statistics 11, pp. 95-103, 1983.

Received: October 25, 2005