POISSON HIDDEN MARKOV MODELS FOR TIME SERIES OF OVERDISPERSED INSURANCE COUNTS

ROBERTA PAROLI
Istituto di Statistica
Università Cattolica S.C. di Milano

GIOVANNA REDAELLI
Istituto di Econometria e Matematica per le decisioni economiche, finanziarie e assicurative
Università Cattolica S.C. di Milano

LUIGI SPEZIA
Dipartimento di Ingegneria
Università degli Studi di Bergamo

ABSTRACT

We suggest the use of Poisson hidden Markov models (PHMMs) in non life insurance. PHMMs are an extension of the well-known mixture models and we use them to model the dynamics of overdispersed data, in particular of the claim number. PHMMs allow us to explicitly consider unobserved factors influencing the dynamics of the claim number. This has an immediate impact on the value of the pure risk premium: the expected claim number is given by a weighted average of the intensity parameters of a PHMM. We show how the maximum likelihood estimators of the parameters of PHMMs may be suitably obtained using the EM algorithm and apply PHMMs to model the daily frequencies of injuries in the work place in Lombardia (Italy).

KEYWORDS

Poisson processes; overdispersion; Markov chains; mixture models; EM algorithm.

INTRODUCTION

In non life insurance, the Poisson distribution is commonly used to model the claim number distribution. Such a choice assumes, among others, mutual independence among the number of claims occurring in disjoint time intervals. Nevertheless, if we consider background factors such as economic or weather conditions affecting the claim-causing events, the claim intensity (known as risk propensity) may vary significantly. As long as these variations are deterministic, the Poisson distribution still applies. When, on the contrary, the intensity variations are random, the independence assumption holds no longer. Some generalizations of the Poisson distribution were proposed in the literature in order to overcome these difficulties. For example, in Daykin, Pentikäinen, Pesonen (1994) the mixed Poisson distribution is suggested; it depends on two parameters: one coinciding with the Poisson parameter of the claim number distribution when the independence assumption holds, the other one being a random variable (called mixing variable) with unit expected value, which multiply the Poisson parameter in the probability function.

Another approach was proposed by Consul (1990). He suggested the use of the so called Generalized Poisson distribution (GPD) as an alternative to the Poisson distribution to design the bonus-malus systems. The GPD differs from the Poisson distribution in that it depends on two parameters whose values were given an interesting interpretation: in the context of the automobile third party liability portfolios, one parameter would reflect road as well as traffic conditions while the other one would depend on the number of passengers in a car when an accident happens.

In any case, we see that all these generalizations aim to model the effect of some unobservable and non-perfectly predictable phenomena which reasonably influence the claim number distribution and do this by introducing an additional parameter in the model probability distribution. Furthermore, in these models the claim number distribution is not time-dependent.

When we want to explicitly consider the dynamics of the claim number, we can refer to counting processes. In non life insurance Poisson processes are commonly used (see Embrechts, Klüppelberg, Mikosh, 1997). A Poisson process is carachterized, among others, by a constant through time intensity; the increments of the Poisson process have mean equals to variance and this parameter is the intensity of the Poisson process.

On the contrary, when we have count data with variance greater than mean, i.e. overdispersion, we may assume the Poisson intensity no longer constant but having a given probability distribution. In this case, we may model the counting process using Poisson mixture models, assuming both independent observations and Markov dependent mixture models, i.e. Poisson hidden Markov models (PHMMs).

PHMMs were originally developed and applied in the biometric field (see Albert, 1991; Le, Leroux, Puterman, 1992; Leroux and Puterman, 1992).

In this paper, we suggest the use of PHMMs to model the dynamics of the claim number in non life insurance, dealing with problems of parameters estimation. We assume discrete time stochastic processes $\{(X_t;Y_t)\}_{t\in\mathbb{N}}$ where $\{X_t\}_{t\in\mathbb{N}}$ is an unobserved finite-state Markov chain and $\{Y_t\}_{t\in\mathbb{N}}$ is the sequence of day t claim number such that Y_t given a state of X_t is, for every t, a Poisson random variable, whose parameter depends on the state of X_t. The marginal distribution of each Y_t is then a finite mixture of Poisson distributions with expected value $E(Y_t)$ equals to the weighted average of the Poisson intensity parameters, with weights the marginal probabilities of the Markov chain. This expected value is relevant in non life insurance to determine the total claim amount distribution and to compute pure risk premiums.

Notice that the Poisson process, commonly used in non life insurance, is a special case of PHMMs, obtained when the Markov chain $\{X_t\}_{t\in\mathbb{N}}$ has only one state.

The paper is organized as follows: the basic PHMM is introduced in Section 1; then, in Section 2 we show how the maximum likelihood estimators of the unknown parameters of PHMMs may be suitably obtained using the EM algorithm: we obtain the likelihood function and describe the EM method; then we give explicit formulas for the parameters estimators. Finally, in Section 3 we apply the results obtained in the previous section to model the dynamics of claim number on a data set of frequencies of injuries in the work place in Lombardia, a northern region of Italy.

1. POISSON HIDDEN MARKOV MODELS

Poisson hidden Markov models are special hidden Markov models (HMMs), which are discrete time stochastic processes $\{(X_t;Y_t)\}_{t\in\mathbb{N}}$ such that $\{X_t\}_{t\in\mathbb{N}}$ is an unobservable finite state Markov chain and $\{Y_t\}_{t\in\mathbb{N}}$ is an observed sequence of random variables depending on $\{X_t\}_{t\in\mathbb{N}}$. This dependence is modelled assuming that the conditional distribution of each observed Y_t, given the sequence $\{X_t\}_{t\in\mathbb{N}}$, depends only on the contemporary unobservable X_t (*contemporary dependence condition*); furthermore, given $\{X_t\}_{t\in\mathbb{N}}$, $\{Y_t\}_{t\in\mathbb{N}}$ is a sequence of conditionally independent random variables (*conditional independence condition*). If we assume that, for every t, Y_t given a state of X_t is a Poisson random variable, we have the so-called Poisson hidden Markov models. In this case, X_t determines the Poisson parameter used to generate Y_t.

Let us introduce some notation and assumptions.

We assume the unobserved process $\{X_t\}_{t\in\mathbb{N}}$ is a discrete, homogeneous, aperiodic, irreducible Markov chain on a finite state-space $S_X = \{1,2,\ldots,m\}$ (for details on Markov chains, see, for example, Grimmett and Stirzaker, 1992, or Guttorp, 1995); we denote with $\gamma_{i,j}$ the transition probability from state i, at time $t-1$, to state j, at time t (for any state i,j and for any time t) , i.e.: $\gamma_{i,j} = P(X_t = j \mid X_{t-1} = i) = P(X_2 = j \mid X_1 = i)$. Let $\Gamma = [\gamma_{i,j}]$ be the $(m\times m)$ transition probabilities matrix, with $\sum_{j\in S_X} \gamma_{i,j} = 1$, for any $i \in S_X$. The marginal distribution of X_1 is the initial distribution denoted by $\delta = (\delta_1, \delta_2, \ldots, \delta_m)'$, with $\delta_i = P(X_1 = i)$, for any $i = 1,2,\ldots,m$, and $\sum_{i\in S_X} \delta_i = 1$; as an immediate consequence of the assumptions on the Markov chain $\{X_t\}_{t\in\mathbb{N}}$, δ is the stationary distribution and the equality $\delta' = \delta'\Gamma$ holds; i.e. δ is the left eigenvector of the matrix Γ, associated with the eigenvalue 1 which always exists since Γ is a stochastic matrix (see Guttorp, 1995, p. 19).

Let us consider now the observed sequence $\{Y_t\}_{t\in\mathbb{N}}$. In PHMMs, any observed variable Y_t conditioned on X_t is Poisson for any t; when X_t is in state i ($i \in S_X$; $t \in \mathbb{N}$), then the conditional distribution of Y_t is a Poisson random variable with parameter λ_i; for any $y \in \mathbb{N}$, the state-dependent probabilities are given by

$$\pi_{y,i} = P\left(Y_t = y | X_t = i\right) = e^{-\lambda_i}\frac{\lambda_i^y}{y!}$$

with $\sum_{y\in\mathbb{N}} \pi_{y,i} = 1$ for every $i \in S_X$. Since $\{X_t\}_{t\in\mathbb{N}}$ is a strongly stationary process also the observed process $\{Y_t\}$ is strongly stationary; therefore, Y_t, for every t, has the same marginal distribution:

$$\begin{aligned} P(Y_t = y) &= \sum_{i\in S_X} P(Y_t = y, X_t = i) = \sum_{i\in S_X} P(Y_t = y \mid X_t = i)P(X_t = i) \\ &= \sum_{i\in S_X} \delta_i \pi_{y,i} \end{aligned}$$

which is a finite mixture of Poisson distributions. Furthermore, it can be easily shown that the expected value of Y_t, for every t, is given by:

$$E(Y_t) = \sum_{i\in S_X} \delta_i\lambda_i.$$

Finally, we notice that the variables Y_t's are overdispersed, that is the variance is greater than the mean; in fact, it holds: $V(Y_t) = \lambda' D\lambda + \delta'\lambda - (\delta'\lambda)^2 > E(Y_t) = \delta'\lambda$, for any t, with $\lambda = (\lambda_1, \ldots, \lambda_m)'$ and $D = \text{diag}(\delta)$ (see MacDonald and Zucchini, 1997, p. 70).

2. PARAMETERS ESTIMATION

The PHMM described in the previous section depends on the following set of parameters: the initial stationary distribution $\delta = (\delta_1, \delta_2, \ldots, \delta_m)'$, the transition probabilities $\gamma_{i,j}$ $(i, j \in S_X)$ and the state-dependent probabilities $\pi_{y,i}$ $(y \in \mathbb{N};\ i \in S_X)$.
We now search for some estimators of these parameters. In particular, we search for the maximum likelihood estimators of the $m^2 - m$ transition probabilities $\gamma_{i,j}$ with $i \neq j$, i.e. the off-diagonal elements of the matrix Γ (the diagonal elements are obtained by difference, since each row of Γ sums to one: $\gamma_{i,i} = 1 - \sum_{\substack{j \in S_X \\ j \neq i}} \gamma_{i,j}$, for any $i \in S_X$) and the maximum likelihood estimators of the m Poisson parameters λ_i entering the state-dependent probabilities $\pi_{y,i}$. By using the estimated matrix Γ, we then get the estimator of the initial distribution δ from the equality $\delta' = \delta'\Gamma$ (being δ the stationary distribution).
Let us denote with ϕ the vector of the unknown parameters to be estimated with the maximum likelihood method,

$$\phi = (\gamma_{1,2}, \gamma_{1,3}, \ldots, \gamma_{m,m-1}, \lambda_1, \ldots, \lambda_m)',$$

and let Φ be the parameter space.
Let $y = (y_1, \ldots, y_T)'$ be the vector of the observed data, i.e. the sequence of T realizations of the stochastic process $\{Y_t\}_{t\in\mathbb{N}}$; the vector y is incomplete because the sequence of the states of the chain $\{X_t\}_{t\in\mathbb{N}}$ is missing. Let $x = (i_1, \ldots, i_T)'$ be the vector of the unobserved states of the chain $\{X_t\}_{t\in\mathbb{N}}$; hence $(i_1, y_1, \ldots, i_T, y_T)'$ is the vector of the complete data.
The likelihood function of the complete data $L_T^c(\phi)$ is defined as the joint probability of the T observations and the T unobserved states. Applying the Markov dependence, conditional independence and contemporary dependence conditions, we easily get:

$$L_T^c(\phi) = P(Y_1 = y_1, \ldots, Y_T = y_T, X_1 = i_1, \ldots, X_T = i_T) = \delta_{i_1} \pi_{y_1,i_1} \prod_{t=2}^{T} \gamma_{i_{t-1},i_t} \pi_{y_t,i_t};$$

summing over $i_1, \ldots, i_T$ both sides, we obtain the likelihood function of incomplete data:

$$L_T(\phi) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T) = \sum_{i_1 \in S_X} \sum_{i_2 \in S_X} \cdots \sum_{i_T \in S_X} \delta_{i_1} \pi_{y_1,i_1} \prod_{t=2}^{T} \gamma_{i_{t-1},i_t} \pi_{y_t,i_t},$$

where π_{y_t,i_t} is the state-dependent probability of y_t conditioned on the state i_t $(t = 1, \ldots, T)$:

$$\pi_{y_t,i_t} = e^{-\lambda_{i_t}} \frac{\lambda_{i_t}^{y_t}}{y_t!}. \tag{1}$$

In order to find the maximum likelihood estimator of ϕ we should solve the likelihood system but it is very hard to analitically find the solution, then we must use a numerical algorithm. Given that we are in a situation with incomplete data, we shall perform the EM algorithm (see McLachlan and Krishnan, 1997; Lange, 1999), which is based on an iterative procedure with two steps at each iteration: the first step, E step, provides the computation of an *Expectation*; the second one, M step, provides a *Maximization*.
Let $Q(\phi; \phi')$ the function defined at the E step:

$$Q(\phi; \phi') = E_{\phi'}(\ln L_T^c(\phi) \mid y),$$

for any given vector ϕ' belonging to the parameter space Φ.

In Dempster, Laird, Rubin (1977) it is proved that a sufficient condition for maximizing $\ln L_T(\phi)$ is to maximize $Q(\phi;\phi')$ with respect to ϕ. Without going into details, the iterative scheme of the EM algorithm is the following. Let $\phi^{(k)}$ be the vector of estimates obtained at the k^{th} iteration:

$$\phi^{(k)} = \left(\gamma_{1,2}^{(k)}, \gamma_{1,3}^{(k)}, \ldots, \gamma_{m,m-1}^{(k)}, \lambda_1^{(k)}, \ldots, \lambda_m^{(k)}\right)',$$

at the $(k+1)^{th}$ iteration, the E and M steps are defined as follows:

- E step - given $\phi^{(k)}$, compute

$$Q\left(\phi;\phi^{(k)}\right) = E_{\phi^{(k)}}\left(\ln L_T^c(\phi) \mid y\right);$$

- M step - search for that $\phi^{(k+1)}$ which maximize $Q\left(\phi;\phi^{(k)}\right)$, i.e. such that

$$Q\left(\phi^{(k+1)};\phi^{(k)}\right) \geq Q\left(\phi;\phi^{(k)}\right),$$

  for any $\phi \in \Phi$.

The E and M steps must be repeated in an alternating way until the sequence of log-likelihood values $\left\{\ln L_T\left(\phi^{(k)}\right)\right\}$ converges, i.e. until the difference

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

is less than or equal to a sufficiently small arbitrary value. When some regularity conditions on the parameter space Φ and on the functions $L_T(\phi)$ and $Q(\phi;\phi')$ are satisfied (see Wu, 1983, pp. 94-96) we can say that, if the algorithm converges at the $(k+1)^{th}$ iteration, then $\left(\phi^{(k+1)}; \ln L_T\left(\phi^{(k+1)}\right)\right)$ is a stationary point and $\phi^{(k+1)} = \left(\gamma_{1,2}^{(k+1)}, \gamma_{1,3}^{(k+1)}, \ldots, \gamma_{m,m-1}^{(k+1)}, \lambda_1^{(k+1)}, \ldots, \lambda_m^{(k+1)}\right)'$ is the maximum likelihood estimator of the unknown parameter ϕ. In PHMMs, a sufficient condition for Wu's conditions to hold is that the Poisson parameters λ_i $(i = 1, 2, \ldots, m)$ are strictly positive and bounded (see Appendix A).

For HMMs the log-likelihood surface is irregular and characterized by many local maxima or stationary points; then, the stationary point to which the EM algorithm converges may not be the global maximum. Hence, in order to identify the global maximum, the choice of the starting point is of primary importance.

Implementing the algorithm, the search for the estimators of the unknown parameters with the EM algorithm may be simplified using the *forward* and the *backward probabilities,* introduced by Baum *et al.* (1970). The forward probability, denoted by $\alpha_t(i)$, is the joint probability of the past and the present observations and the current state of the chain:

$$\alpha_t(i) = P(Y_1 = y_1, \ldots, Y_t = y_t, X_t = i);$$

while the backward probability, denoted by $\beta_t(i)$, is the probability of the future observations conditioned on the current state of the chain:

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

The probabilities $\alpha_t(i)$ and $\beta_t(i)$ may be obtained recursively as follows:

$$\begin{aligned} &\alpha_1(i) = \delta_i\ \pi_{y_1,i}, \text{ with } i = 1, 2, \ldots, m, \\ &\alpha_t(j) = \left(\sum_{i \in S_X} \alpha_{t-1}(i)\gamma_{i,j}\right)\pi_{y_t,j}, \text{ with } t = 2, \ldots, T, \text{ and } j = 1, 2, \ldots, m, \end{aligned} \tag{2}$$

for the forward probabilities and

$$\beta_T(i) = 1, \text{ with } i = 1, 2, \ldots, m,$$
$$\beta_t(i) = \sum_{j \in S_X} \pi_{y_{t+1},j} \; \beta_{t+1}(j) \; \gamma_{i,j}, \text{ with } t = T-1, \ldots, 1, \text{ and } i = 1, 2, \ldots, m, \tag{3}$$

for the backward probabilities (see MacDonald and Zucchini, 1997, p. 60).
Then, we obtain the following expression for the function $Q\left(\phi; \phi^{(k)}\right)$ at the E step of the $(k+1)^{th}$ iteration of the EM algorithm

$$\begin{aligned} Q\left(\phi; \phi^{(k)}\right) &= E_{\phi^{(k)}}\left(\ln L_T^c(\phi) \mid y\right) = \\ &= \sum_{i \in S_X} \frac{\alpha_1^{(k)}(i)\, \beta_1^{(k)}(i)}{\sum_{l \in S_X} \alpha_t^{(k)}(l)\, \beta_t^{(k)}(l)} \ln \delta_i + \sum_{i \in S_X} \sum_{j \in S_X} \frac{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i)\, \gamma_{i,j}^{(k)}\, \pi_{y_{t+1},j}^{(k)}\, \beta_{t+1}^{(k)}(j)}{\sum_{l \in S_X} \alpha_t^{(k)}(l)\, \beta_t^{(k)}(l)} \ln \gamma_{i,j} + \\ &+ \sum_{i \in S_X} \frac{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\, \beta_t^{(k)}(i)}{\sum_{l \in S_X} \alpha_t^{(k)}(l)\, \beta_t^{(k)}(l)} \ln \pi_{y_t,i} \end{aligned} \tag{4}$$

(see Spezia, 1999, pp. 70-75), where $\pi_{y_t,i}^{(k)}$, $\alpha_t^{(k)}(i)$ and $\beta_t^{(k)}(i)$ are computed according to formulas (1), (2) and (3), respectively, using the values of the parameter $\phi^{(k)}$, obtained at the k^{th} iteration; while $\delta^{(k)}$ is computed as $\delta'^{(k)} = \delta'^{(k)}\Gamma^{(k)}$.
It should be noticed that δ, by the stationarity assumption, contains informations about the transition probability matrix Γ, since $\delta_j = \sum_{i \in S_X} \delta_i \gamma_{i,j}$, for any $j \in S_X$. Nevertheless, for large T, the effect of δ is negligible (see Basawa and Prakasa Rao, 1980, pp. 53-54). Therefore, at the M step of the $(k+1)^{th}$ iteration, to obtain $\phi^{(k+1)}$, we may ignore the first addendum in (4) when maximizing $Q\left(\phi; \phi^{(k)}\right)$ with respect to the $m^2 - m$ parameters $\gamma_{i,j}$'s.
The expression for the maximum likelihood estimator of $\gamma_{i,j}$ obtained at the $(k+1)^{th}$ iteration of the EM algorithm is given by (see Spezia, 1999, pp. 64-66):

$$\gamma_{i,j}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i)\; \gamma_{i,j}^{(k)}\; \pi_{y_{t+1},j}^{(k)}\; \beta_{t+1}^{(k)}(j)}{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i)\; \beta_t^{(k)}(i)}, \tag{5}$$

for any state i and any state j, $j \neq i$, of the Markov chain $\{X_t\}$. The maximum likelihood estimator of λ_i obtained at the $(k+1)^{th}$ iteration of the EM algorithm, is given by[1]:

$$\lambda_i^{(k+1)} = \frac{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\; \beta_t^{(k)}(i)\; y_t}{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\; \beta_t^{(k)}(i)}, \tag{6}$$

for any state i of the Markov chain $\{X_t\}$.
Leroux (1992) and Bickel, Ritov, Rydén (1998) proved that the estimators in (5) and in (6) are consistent and asymptotically normal.

[1]The formula for $\lambda_i^{(k+1)}$ is easily obtained deriving $Q\left(\phi; \phi^{(k)}\right)$ in (4) with respect to λ_i and setting this derivative equal to 0.

3. APPLICATION TO A DATA SET OF INJURIES IN THE WORK PLACE

The EM algorithm introduced in Section 2 is now applied to compute the parameters of the PHMMs used to describe the dynamics of the daily frequencies of injuries in the work place in the first four months 1998 in each of the 11 provinces in Lombardia, Italy (source INAIL: *Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro*[2], private communication).
The iterative procedure of the algorithm is implemented in a GAUSS code. The use of formulas (5) and (6) simplifies the optimization problem, because it allows us to solve the M-step exactly, without using a numerical maximization algorithm, such as the Newton-Raphson method. Hence the procedure is more stable and converges faster in the neighborhood of the maximum.
In order to identify the global maximum, the code repeats the iterative procedure more than once, starting from several different initial points, randomly chosen in the parameter space Φ. Then, we compare the stationary points obtained at each run of the algorithm and choose the one with the largest likelihood value. The corresponding vector of parameters value is the vector of the maximum likelihood estimators we search for.
Since we do not know the dimension m of the state-space of the Markov chain, we estimate it following Leroux and Puterman (1992): we use two maximum-penalized-likelihood methods, i.e. we search for m^* which maximizes the difference $\ln L_T^{(m)}(\phi) - a_{m,T}$, where $\ln L_T^{(m)}(\phi)$ is the log-likelihood function maximized over a PHMM with an m-states Markov chain, while $a_{m,T}$ is a penalty term depending on the number m of states and the length T of the observed sequence. If $a_{m,T} = d_m$, where d_m is the dimension of the model, that is the number of the parameters estimated with the EM algorithm (m^2), we have the *Akaike Information Criterion* (AIC); if $a_{m,T} = (\ln T)d_m/2$ we have the *Bayesian Information Criterion* (BIC).
In Appendix B it is reported the table of estimates of the parameters of the PHHMs for each of the 11 provinces in Lombardia.
For each province we, first, give the estimated number m^* of states of the Markov chain for both the AIC and the BIC; we see that in many cases, the two selection criteria have given two different values for the optimal number m^*. Nevertheless, it is worth noticing that in all cases where the empirical data were overdispersed (as confirmed by the sample mean less than the sample variance) we get $m^* > 1$ according to both the AIC and the BIC criteria. Similarly, in almost all the provinces where the data were not overdispersed, we get $m^* = 1$.
For each province, we report the number $(k+1)$ of iterations after which the algorithm has converged, given $m = m^*$; then, the estimated transition probabilities matrix is $\Gamma^{(k+1)}$ whose elements are the estimated transition probabilities from state i to state j; for example, looking at Como (where $m^* = 2$ for both the criteria) we have that the algortithm has converged at the 50^{th} iteration giving the matrix of the estimated transition probabilities

$$\Gamma^{(50)} = \begin{bmatrix} 0.5990 & 0.4010 \\ 0.2957 & 0.7043 \end{bmatrix}$$

i.e. we have a probability 0.5990 of staying in state 1, a probability 0.4010 of visiting state 2 from state 1; similarly, the probability of visiting state 1 from state 2 is equal to 0.2957 while that of staying in state 2 is given by 0.7043.

[2] *National board for the insurance against injuries in the work place*

The vector $\delta^{(k+1)}$ gives the estimated initial stationary distribution obtained by the equality $\delta'^{(k+1)} = \delta'^{(k+1)}\Gamma^{(k+1)}$; in the case of Como, we get the estimated initial probability of state 1 equal to 0.4244 and the estimated initial probability of state 2 is equal to 0.5576. The vector $\lambda^{(k+1)}$ of estimates of the Poisson parameters are reported in the fifth row of the table; in the case of Como, when the Markov chain is in state 1, the estimated Poisson parameter is 0.2969; when the Markov chain is in state 2 the corresponding estimated Poisson parameter is 2.1963. Notice it is not necessary to give the Markov chain a real interpretation: we use it only for inferential aims.

Whenever $m^* = 1$, the daily frequencies of injuries in the work place constitutes a sequence of indipendent and identical distributed Poisson random variables with parameter $\lambda^{(k+1)}$. In this case, $\lambda^{(k+1)}$ coincides with the sample mean of the observations, which is also the maximum likelihood estimator $\widehat{\lambda}$ (reported in the sixth row of the table) of the Poisson parameter, when we assume a Poisson distribution to model our data.

In order to examine the impact of the PHMMs on the insurance premium calculation, we also report the expected number of injuries per day, given by $E(Y_t) = \delta'^{(k+1)}\lambda^{(k+1)}$. We notice that we get $E(Y_t) > \widehat{\lambda}$ in all those cases with $m^* > 1$. This fact has obvious implications on the value of the pure risk premium, which is given by the product of the expected number of claims and the expected amount of each claim (see Daykin, Pentikäinen, Pesonen, 1994).

CONCLUSIONS AND EXTENSIONS

We suggested PHMMs as a more general approach than Poisson distribution and Poisson process to model claim number in non life insurances. PHMMs allows to model overdispersion in count data and to explain variability, by switching the Poisson parameter according to an unobserved Markov chain.

The main consequence in non life insurance applications is the way the expected claim number is computed; it is a weighted average of the state-dependent intensities λ_i with weights the marginal distribution δ of the unobserved variables affecting the dynamics of the claim number. As we said at the end of the application, this fact may have strong implications on the value of the pure risk premium.

In this application, the dimension m of the Markov chain state-space has been estimated by the *Akaike Information Criterion* (AIC) and the *Bayes Information Criterion* (BIC). The way to estimate m is yet an open question, because the consistency of AIC and BIC has not been formally established; other criteria have been proposed, but they did not join the optimum (see Rydén, 1999, and the references therein).

Also the study of a suitable criterion for model validation is an open question, because in HMMs residual analysis can not be performed, given that residuals can not be computed, being unobserved the Markov chain (Rydén, 1999).

As we can see, there is a lot to study in the field of HMMs, because they are a recent topic; consider that the main asymptotical results have been obtained in the last years: consistency (Leroux, 1992), asymptotic normality (Bickel, Ritov, Rydén, 1998) and likelihood ratio test (Giudici, Rydén, Vandekerkhove, 1998).

Furthermore, we are interested in the computation of the information matrix, according to Oakes (1999), implementing it in our GAUSS code.

Finally, since we may consider HMMs with other specified distributions, both discrete and continuous (Bernoulli, binomial, negative binomial, gaussian), yet implemented in the authors' codes, we think insurance is an interesting field of applications of HMMs.

APPENDIX A - Wu's conditions for PHMMs

Let $\phi = (\gamma_{1,2}, \gamma_{1,3}, \ldots, \gamma_{m,m-1}, \lambda_1, \ldots, \lambda_m)'$ denote the vector of the m^2 parameters to be estimated with the EM algorithm, Φ the set of admissible estimates (i.e., $\phi \in \Phi$) and $L_T(\phi)$ the likelihood function. The following result holds.

Proposition 1 *Let $\lambda_i \in [\varepsilon; 1/\varepsilon]$ for every i (ε arbitrary small). Then the following conditions hold:*

1. *Φ is a bounded subset of $\mathbb{R}^{m^2}$;*
2. *$L_T(\cdot)$ is continuous in Φ and differentiable in the interior of Φ;*
3. *$\Phi_{\phi_0} = \{\phi \in \Phi : L_T(\phi) \geq L_T(\phi_0)\}$ is compact for any $L_T(\phi_0) > -\infty$;*
4. *$Q\left(\phi; \phi^{(k)}\right)$ is continuous in both ϕ and $\phi^{(k)}$.*

Proof.

1. It is $\gamma_{i,j} \in [0; 1]$ for every i, j since $\gamma_{i,j} = P(X_t = j | X_{t-1} = i)$ and $\lambda_i \in [\varepsilon; 1/\varepsilon]$ (ε arbitrary small) by assumption. Therefore, we get $\Phi = [0; 1]^{m^2-m} \times [\varepsilon; 1/\varepsilon]^m$ which is a bounded subset of $\mathbb{R}^{m^2}$.

2. Condition 2 holds since $L_T(\cdot)$ is obtained by summing up products of continuous (in Φ) and differentiable (in the interior of Φ) functions.

3. Let $\phi_0 \in \Phi$ be given. The set Φ_{ϕ_0} is bounded since $\Phi_{\phi_0} \subset \Phi$. We now prove Φ_{ϕ_0} is also closed, then compact. The proof is by contradiction. Let $\{\phi_n\}_{n \geq 1}$ be a sequence in Φ_{ϕ_0}, i.e. $L_T(\phi_n) \geq L_T(\phi_0)$ for every n , such that $\phi_n \to \phi^*$. Let us suppose $\phi^* \notin \Phi_{\phi_0}$, i.e. $L_T(\phi^*) < L_T(\phi_0)$. Let $\varepsilon = L_T(\phi_0) - L_T(\phi^*)$; since $L_T(\cdot)$ is continuous in Φ, it is: $\lim_{n\to\infty} L_T(\phi_n) = L_T(\phi^*)$, i.e. there exists n^* such that for every $n \geq n^*$ it holds: $L_T(\phi_n) \leq L_T(\phi^*) + \varepsilon/2 < L_T(\phi_0)$ which is a contradiction since $\phi_n \in \Phi_{\phi_0}$ by assumption. Therefore, Φ_{ϕ_0} is a compact set in $\mathbb{R}^{m^2}$ since it is bounded and closed in $\mathbb{R}^{m^2}$.

4. In order to prove the continuity of $Q\left(\phi; \phi^{(k)}\right)$ with respect to both its arguments we refer to expression (4). The components of vector ϕ appear only in the arguments of the log terms in expression (4); since these are continuous functions (where defined) of the parameters $\gamma_{1,2}, \gamma_{1,3}, \ldots, \gamma_{m,m-1}, \lambda_1, \ldots, \lambda_m$, the continuity of $Q\left(\phi; \phi^{(k)}\right)$ with respect to ϕ immediately follows. On the other hand, the components of vector $\phi^{(k)} = \left(\gamma_{1,2}^{(k)}, \gamma_{1,3}^{(k)}, \ldots, \gamma_{m,m-1}^{(k)}, \lambda_1^{(k)}, \ldots, \lambda_m^{(k)}\right)'$ determine the $\alpha_t^{(k)}(i)$ and $\beta_t^{(k)}(i)$ terms ($t = 1, 2 \ldots, T$ and $i \in S_X$) in expression (4), which can be defined recursively according to formulas (2) and (3), respectively. From those definitions and noticing that $\pi_{y,i}$ is continuous in λ_i, we get the continuity of $Q\left(\phi; \phi^{(k)}\right)$ with respect to $\phi^{(k)}$.$\square$

Conditions 1 to 4 of the previous theorem correspond to Wu's regularity conditions (see Wu, 1983, conditions (5), (6), (7) p.96 and (10) p.98), implying that the point $\left(\phi^{(k+1)}; \ln L_T\left(\phi^{(k+1)}\right)\right)$ obtained with the EM algorithm is a stationary point and $\phi^{(k+1)} = \left(\gamma_{1,2}^{(k+1)}, \gamma_{1,3}^{(k+1)}, \ldots, \gamma_{m,m-1}^{(k+1)}, \lambda_1^{(k+1)}, \ldots, \lambda_m^{(k+1)}\right)'$ is the maximum likelihood estimator of the unknown parameter ϕ, whenever the algorithm converges at the $(k+1)^{th}$ iteration.

APPENDIX B

Table of the parameters estimates of the PHHMs for each of the 11 provinces in Lombardia: Bergamo, Brescia, Como, Cremona, Lecco, Lodi, Mantova, Milano, Pavia, Sondrio, Varese.

	AIC	BIC
Bergamo	$m^*=3$	$m^*=2$
$k+1$	57	54
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.4117 & 0 & 0.5883 \\ 0.2724 & 0.7276 & 0 \\ 0 & 0.4058 & 0.5942 \end{bmatrix}$	$\begin{bmatrix} 0.3760 & 0.62400 \\ 0.1495 & 0.8505 \end{bmatrix}$
$\delta^{(k+1)}$	$\begin{pmatrix} 0.2170 & 0.4685 & 0.3145 \end{pmatrix}'$	$\begin{pmatrix} 0.1932 & 0.8068 \end{pmatrix}'$
$\lambda^{(k+1)}$	$\begin{pmatrix} 0.5004 & 3.1951 & 5.3213 \end{pmatrix}'$	$\begin{pmatrix} 0.4295 & 3.9965 \end{pmatrix}'$
$\hat{\lambda}$	3.275	3.275
$E(Y_t)$	3.279	3.307
Brescia	$m^*=3$	$m^*=2$
$k+1$	80	174
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.4309 & 0.5691 & 0 \\ 0 & 0.6228 & 0.3372 \\ 0.3301 & 0 & 0.6699 \end{bmatrix}$	$\begin{bmatrix} 0.4479 & 0.5521 \\ 0.1426 & 0.8574 \end{bmatrix}$
$\delta^{(k+1)}$	$\begin{pmatrix} 0.2267 & 0.3825 & 0.3908 \end{pmatrix}'$	$\begin{pmatrix} 0.2053 & 0.7947 \end{pmatrix}'$
$\lambda^{(k+1)}$	$\begin{pmatrix} 0 & 3.1415 & 2.1796 \end{pmatrix}'$	$\begin{pmatrix} 0.0047 & 2.6006 \end{pmatrix}'$
$\hat{\lambda}$	2.0417	2.0417
$E(Y_t)$	2.0535	2.0677
Como	$m^*=2$	$m^*=2$
$k+1$	50	50
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.5990 & 0.4010 \\ 0.2957 & 0.7043 \end{bmatrix}$	$\begin{bmatrix} 0.5990 & 0.4010 \\ 0.2957 & 0.7043 \end{bmatrix}$
$\delta^{(k+1)}$	$\begin{pmatrix} 0.4244 & 0.5756 \end{pmatrix}'$	$\begin{pmatrix} 0.4244 & 0.5756 \end{pmatrix}'$
$\lambda^{(k+1)}$	$\begin{pmatrix} 0.2969 & 2.1963 \end{pmatrix}'$	$\begin{pmatrix} 0.2969 & 2.1963 \end{pmatrix}'$
$\hat{\lambda}$	1.375	1.375
$E(Y_t)$	1.3902	1.3902
Cremona	$m^*=1$	$m^*=1$
$k+1$	3	3
$\lambda^{(k+1)}=\hat{\lambda}=E(Y_t)$	0.4833	0.4833
Lecco	$m^*=2$	$m^*=1$
$k+1$	135	3
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.5286 & 0.4714 \\ 0.1732 & 0.8268 \end{bmatrix}$	
$\delta^{(k+1)}$	$\begin{pmatrix} 0.2687 & 0.7313 \end{pmatrix}'$	
$\lambda^{(k+1)}$	$\begin{pmatrix} 0 & 0.9112 \end{pmatrix}'$	0.6583
$\hat{\lambda}$	0.6583	0.6583
$E(Y_t)$	0.6664	0.6583
Lodi	$m^*=1$	$m^*=1$
$k+1$	3	3
$\lambda^{(k+1)}=\hat{\lambda}=E(Y_t)$	0.6500	0.6500

Mantova	m^*=1	m^*=1
$k+1$	3	3
$\lambda^{(k+1)} = \hat{\lambda} = E(Y_t)$	0.8167	0.8167
Milano	m^*=5	m^*=3
$k+1$	61	96
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.1206 & 0 & 0.1067 & 0.7727 & 0 \\ 0 & 0.2621 & 0 & 0 & 0.7379 \\ 0 & 0.6528 & 0.3472 & 0 & 0 \\ 0.0958 & 0 & 0.9042 & 0 & 0 \\ 0.4835 & 0 & 0 & 0 & 0.5165 \end{bmatrix}$	$\begin{bmatrix} 0.8003 & 0 & 0.1997 \\ 0.9131 & 0.0014 & 0.0855 \\ 0.1010 & 0.7809 & 0.1091 \end{bmatrix}$
$\delta^{(k+1)}$	$\begin{pmatrix} 0.1735 & 0.1894 & 0.2140 & 0.1341 & 0.2890 \end{pmatrix}'$	$\begin{pmatrix} 0.6984 & 0.1324 & 0.1692 \end{pmatrix}'$
$\lambda^{(k+1)}$	$\begin{pmatrix} 1.7839 & 12.0161 & 8.2727 & 0.5308 & 7.8651 \end{pmatrix}'$	$\begin{pmatrix} 9.1281 & 0.4999 & 1.8417 \end{pmatrix}'$
$\hat{\lambda}$	6.7	6.7
$E(Y_t)$	6.7	6.7530
Pavia	m^*=2	m^*=1
$k+1$	311	3
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.3303 & 0.6697 \\ 0.3511 & 0.6489 \end{bmatrix}$	
$\delta^{(k+1)}$	$\begin{pmatrix} 0.3439 & 0.6561 \end{pmatrix}'$	
$\lambda^{(k+1)}$	$\begin{pmatrix} 0.0169 & 0.8978 \end{pmatrix}'$	0.5917
$\hat{\lambda}$	0.5917	0.5917
$E(Y_t)$	0.5948	0.5917
Sondrio	m^*=1	m^*=1
$k+1$	3	3
$\lambda^{(k+1)} = \hat{\lambda} = E(Y_t)$	0.1083	0.1083
Varese	m^*=3	m^*=2
$k+1$	68	43
$\Gamma^{(k+1)}$	$\begin{bmatrix} 0.4720 & 0 & 0.5280 \\ 0.4669 & 0.5331 & 0 \\ 0 & 0.2978 & 0.7022 \end{bmatrix}$	$\begin{bmatrix} 0.5580 & 0.4420 \\ 0.2403 & 0.7597 \end{bmatrix}$
$\delta^{(k+1)}$	$\begin{pmatrix} 0.2562 & 0.2897 & 0.4541 \end{pmatrix}'$	$\begin{pmatrix} 0.3522 & 0.6478 \end{pmatrix}'$
$\lambda^{(k+1)}$	$\begin{pmatrix} 0.3541 & 2.6110 & 5.2137 \end{pmatrix}'$	$\begin{pmatrix} 0.6496 & 4.6308 \end{pmatrix}'$
$\hat{\lambda}$	3.1917	3.1917
$E(Y_t)$	3.2146	3.2286

REFERENCES

Albert, P. S. (1991). A Two-State Markov Mixture Model for a Time Series of Epileptic Seizure Counts. *Biometrics*, 47, 1371-1381.

Basawa I. V. and Prakasa Rao B. L. S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

Baum, L. E., Petrie, T., Soules, G., Weiss, N. (1970). A maximization technique occuring in the statistical estimation for probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41, 164-171.

Bickel, P. J., Ritov Y., Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics*, 26, pp. 1614-1635.

Consul P.C. (1980). A model for distribution of injuries in autoaccidents. *Swiss Association of Actuaries: Bulletin* 1990, 1, 161-168.

Daykin C.D., Pentikäinen T., Pesonen M. (1994). *Practical Risk Theory for Actuaries.* Chapman & Hall, London.

Dempster A. P., Laird N. M., Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Embrechts P., Klüppelberg C., Mikosh T. (1997). *Modeling Extremal Events for Insurance and Finance.* Springer Verlag, Berlin.

Giudici P., Rydén T., Vandekerkhove P. (1998). *Likelihood ratio tests for hidden Markov models.* Technical report 1998:19, Lund University, Sweden.

Grimmett G. R. and Stirzaker D. R. (1992). *Probability and Random Processes*, second edition. Oxford University Press, Oxford.

Guttorp P. (1995). *Stochastic Modeling for Scientific Data.* Chapman & Hall, London.

Lange, K. (1999). *Numerical Analysis for Statisticians.* Springer Verlag, New York.

Le N.D., Leroux B. G., Puterman M. L. (1992). Reader Reaction: Exact Likelihood Evaluation in a Markov Mixture Model for Time Series of Seizure Counts. *Biometrics*, 48, 317-323.

Leroux B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications*, 40, 127-143.

Leroux B. G. and Puterman M. L. (1992). Maximum-Penalized-Likelihood Estimation for Independent and Markov-Dependent Mixture Models. *Biometrics*, 48, 545-558.

MacDonald I. L. and Zucchini W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman & Hall, London.

McLachlan, G. J. and Krishnan T. (1997). *The EM algorithm and extensions.* John Wiley & Sons, New York.

Oakes D. (1999). Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 61, 479-482.

Rydén, T. (1999). Likelihood inference in ergodic hidden Markov models: a unified approach, implications and future directions. *Abstracts of Second European Conference on Highly Structured Stochastic System.* http:// www. unipv. it /hsss99 /abstracts/tobias.ps.

Spezia L. (1999). *Stima dei parametri di un modello markoviano binomiale negativo parzialmente osservato.* Tesi di dottorato, Università degli Studi di Trento.

Wu C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11, 95-103.

ROBERTA PAROLI
ISTITUTO DI STATISTICA
UNIVERSITÀ CATTOLICA S.C.
VIA NECCHI 9
20123 MILANO
e-mail: rparoli@mi.unicatt.it

GIOVANNA REDAELLI
ISTITUTO DI ECONOMETRIA E MATEMATICA
PER LE DECISIONI ECONOMICHE,
FINANZIARIE E ASSICURATIVE
UNIVERSITÀ CATTOLICA S.C.
VIA NECCHI 9
20123 MILANO
e-mail: redaelli@mi.unicatt.it

LUIGI SPEZIA
DIPARTIMENTO DI INGEGNERIA
UNIVERSITÀ DEGLI STUDI DI BERGAMO
VIA MARCONI 5
24044 DALMINE (BG)
e-mail: spezia@unibg.it