Recursive Estimation of Parameters in Markov-Modulated Poisson Processes

Georg Lindgren and Ulla Holst

Abstract—A hidden Markov regime is a Markov process that governs the time or space dependent distributions of an observed stochastic process. Recursive algorithms can be used to estimate parameters in mixed distributions governed by a Markov regime. Here we derive a recursive algorithm for estimation of parameters in a Markov-modulated Poisson process also called a Cox point process. By this we mean a doubly stochastic Poisson process with a time dependent intensity that can take on a finite number of different values. The intensity switches randomly between the possible values according to a Markov process. We consider two different ways to observe the Markov-modulated Poisson process: in the first model the observations consist of the observed time intervals between events, and in the second model we use the total number of events in successive intervals of fixed length. We derive an algorithm for recursive estimation of the Poisson intensities and the switch intensities between the two states and illustrate the algorithm in a simulation study. The estimates of the switch intensities are based on the observed conditional switch probabilities.

I. INTRODUCTION

MIXTURE MODELS AND RECURSIVE ESTIMATION

STOCHASTIC processes whose statistical characteristics vary randomly over time arise in many applications, such as automatic control, telecommunication, econometrics, image and speech technology, etc. In a mixture model with time sequenced data, observations \( x(1), x(2), \ldots, x(n) \) are generated with distributions which are determined by some underlying regime or state variable \( z(k) \). In the standard mixture model the \( z(k) \)-variables are independent and the \( x(k) \)-variables are conditionally independent, given the \( z(k) \), but in practice it is often more realistic to assume some sort of statistical dependence between the regime variables. One such model is the Markov model—often called hidden Markov model because the \( z(k) \) are unobservable. For an introduction to mixture models, including hidden Markov models, see the monograph by Titterington et al. [1], the survey papers [2] and [3] and the paper by Rabiner [4] for a tutorial in speech technology.

Recursive estimation techniques give approximate maximum likelihood estimates in mixture models with independent data (see [1]) and have recently been shown to work amazingly well also in mixture models with Markov dependent regime; see [5] for estimates in models with conditional independence, and also [6] for models with dependence of autoregressive type between successive observations.

In this paper we shall consider a mixture model in continuous time where the observations consist of observations from a Markov-modulated Poisson process (MMPP). Such models are widely used, e.g., in medical imaging systems as models for photon counts, and in neurophysiology as models for neural firing. They are also often used in telecommunication theory as models for varying loads on a queuing system. For discussions on the properties of the Markov-modulated Poisson processes, see [7]–[9].

Nonrecursive estimation procedures for Markov-modulated Poisson processes have been proposed by several authors; see, e.g., [10] and [7] in which the relationship with the EM-algorithm is discussed. Based on a technique developed by Leroux [11] it is proven in [12] that the ML-estimator is strongly consistent; see also [13] for an alternative technique. In [12] also different practical methods to calculate the MLE’s are discussed; among them a nonrecursive generalized EM-algorithm. Use of classification algorithms for the hidden regime followed by ML-estimation will in general result in biased estimates because of the (small) probability of misclassification; c.f., [12]. Asymptotic normality of a modified ML-estimator in the MMPP is proven in [14].

Reconstruction algorithms based on continuous observation of a Cox process can sometimes be given in recursive form; see the monograph by Snyder and Miller, [15, ch. 7], and references therein, in particular [16]. In models where the intensity is generated by some dynamic (recursive) Markov model with a known structure and partly known parameters, the recursive reconstruction algorithms can also be used for estimation of some intensity specifying parameters. We shall here derive a recursive estimation algorithm based on a discrete sequence of observations of the MMPP. The procedure differs from those in [15] in that our procedure assumes the underlying Markov process to have a finite (and known) number of states, but otherwise to be completely unknown.

A. Recursive Estimation in Markov Mixtures

The recursive estimation procedure for mixture models with Markov regime presented in [5] was based on the assumption that the observations are conditionally independent with a distribution that depends only on the present regime variable. The method can easily be extended to the case where observations depend on a finite number of state variables, provided these still possess the Markov structure. We first give a brief version of the method for the simplest case, and then
give a hint to the proof in the more general case where the observed distribution depends on two of the previous regime or state variables. The reason for choosing just the last two state variables is that the natural observation schemes in the MMP give rise to a two step dependent model.

Suppose \{z(t)\} is a Markov chain with transition probabilities
\[ p_{jk} = P(z(t+1) = k \mid z(t) = j) \]
and stationary distribution
\[ p_j = P(z(t) = j). \]

Further assume that \{x(t)\} is a sequence of random variables with marginal density of the mixed type,
\[ f(x) = \sum_{j=1}^{r} p_j f_j(x) \]  
(1)
where \( f_j(x) \) is the conditional density of \( x(t) \) given that \( z(t) = j \).

Finally assume that \( x(t) \) is conditionally independent of all previous \( x(s) \), \( s \leq t-1 \), given all the \( z(s) \), \( s \leq t \). It is then convenient to start the \( x(t) \)-sequence at \( t = 1 \) and to introduce an extra regime variable \( z(0) \) at the beginning of the \( x(t) \)-sequence. We write \( x^{(n)} = (x(1), \ldots, x(n)) \), \( z^{(n)} = (z(0), z(1), \ldots, z(n)) \), and \( j^{(n)} = (j_0, j_1, \ldots, j_n) \). Then the conditional distribution of \( x^{(n)} \) given \( z^{(n)} = j^{(n)} \), has the density
\[ \prod_{t=1}^{n} f_{j_{t-1}j_t}(x(t)) \]
where \( f_{j_{t-1}j_t}(x) \) is the conditional density of \( x(t) \) given \( z^{(t)} \).

**Remark:** In the simplest case, the conditional distribution of the \( x(t) \) depends only on the present \( z(t) \), and then \( f_{j_{t-1}j_t}(x(t)) = f_{j_t}(x(t)) \) where the \( f_{j_t}(x) \) are the densities in the marginal mixture distribution (1). We call this the simple dependence case, in which
\[ f_{j_{t-1}j_t}(x(t)) = f_{j_t}(x(t)). \]  
(2)

In the MMP to be studied in the next section, the dependence of \( x(t) \) on previous regime variables \( z(s) \) extends two step back, i.e.,
\[ f_{j_{t-1}j_t}(x(t)) = f_{j_{t-1}j_{t+1}}(x(t)). \]  
(3)

We call this the two-step dependence case.

Now, let the densities \( f_{j_{t-1}j_t} \) depend on some unknown parameters \( \psi_i \) to be estimated from an observed sequence \( x(1), \ldots, x(n) \). Also the transition probabilities \( p_{jk} \) may depend on unknown parameters, and we combine all unknown parameters into one parameter vector \( \Psi = (\psi_i, i = 1, \ldots, s)^T \). When needed for clarity, we indicate the dependence of \( p_{jk} \) and other quantities on \( \Psi \) by the notation \( p_{jk}(\Psi) \).

A recursive procedure for estimation of the \( \Psi \)-vector has the form
\[ \Psi^{(n+1)} = \Psi^{(n)} + (n + 1) \left[ H - h(x(n+1); \Psi^{(n)}) \right] \]
(4)
where \( \Psi^{(n)} \) is the estimate after \( n \) observations, \( H_n \) is a suitably chosen adaptive matrix and \( h(x; \Psi_i) = (h(x(j_i)); i = 1, \ldots, s)^T \) is some score vector function to be chosen. For a tutorial on recursive estimation, see e.g. the book by Ljung and Söderström [17].

If one takes
\[ h(x; \Psi) = \frac{\partial \log f(x; \Psi)}{\partial \Psi}, \]
the procedure (4) yields the recursive maximum likelihood estimator in case of independent observations. It was shown in [5] that in the simple dependence case, the score function for recursive estimation in a Markov mixture model can be formulated by means of two functions, \( g(x, i, j; \Psi), x \in R \), and \( \pi_{ij}(n), n = 1, 2, \ldots \), defined as follows.

In [5] the first function was defined as the combined conditional probability for a transition from \( z(n-1) = i \) to \( z(n) = j \) and the conditional likelihood of \( x(n) \) given that \( z(n) = j \), i.e., \( P(z(n) = j \mid z(n-1) = i) f_{ij}(z(n)) f_j(x(n)) = \pi_{ij}(n) f_j(x; \Psi) \).

To cover the two-step dependence case we now extend this definition to include the conditional distribution of \( x(n) \) given the last two regime variables, and define
\[ g(x, i, j, \Psi) = P(z(n) = j \mid z(n-1) = i) \]
\[ f_{ij}(z(n-1) = i, z(n) = j) f_j(x(n)) = \pi_{ij}(n) f_j(x; \Psi). \]
(5)

The second function \( \pi_{ij}(n) \) is the conditional probability, given the \( x(t) \)-observations, that a transition actually takes place from \( z(n-1) = i \) to \( z(n) = j \), i.e.,
\[ \pi_{ij}(n) = P(z(n-1) = i, z(n) = j \mid z(1), \ldots, x(n)). \]  
(6)

We also need the state occupancy probabilities, i.e., the conditional probabilities
\[ \pi_j(n) = P(z(n) = j \mid x(1), \ldots, x(n)). \]

Note that \( \pi_{ij}(n) \) is the conditional expectation of the indicator function
\[ 1\{z(n-1) = i, z(n) = j\}, \]
i.e., \( \pi_{ij}(n) = E\{1\{z(n-1) = i, z(n) = j\} \mid x(1), \ldots, x(n)\} \).

Also write \( 1\{z(0) = i\} \) with \( E\{1\{z(0) = i\} \mid x(1), \ldots, x(n)\} = \pi_i \).

The probabilities \( \pi_{ij} \) and \( \pi_j \) are called the filtering probabilities. They can be calculated recursively from the Markov structure, as in the simple dependence case; see [5] and the following Theorem 1. The normalizing operator \( N_{op} \) appearing in (8) is such that if \( \alpha_{ij} \) are nonnegative real numbers with finite sum, then
\[ \alpha_{jk} N_{op} = \frac{\alpha_{jk}}{\sum \alpha_{il}}. \]

The recursive procedure in [5] for simple dependence works also for two-step dependence as will now be shown.
Theorem 1: a) In a Markov mixture model, with two-step dependence of \( z(t) \) on previous \( z(s) \), a score function for a new observation \( x(n) \) in the recursive procedure (4) is

\[
h(x(n); \Psi) = \sum_{i,j} \pi_{ij}(n) \frac{\partial \log g(x(n); i, j; \Psi)}{\partial \Psi}. \tag{7}
\]

b) The filtering probabilities in a mixture model with Markov regime can be calculated recursively as

\[
\pi_{jk}(n) = \pi_j(n-1) p_{j,k} \phi_{jk}(x(n)) N_{op}, \tag{8}
\]

\[
\pi_j(n) = \sum_i \pi_{ij}(n). \tag{9}
\]

Proof: a) If the \( z(t) \)-variables had been observable, and not only the \( s(t) \)-variables, then the complete likelihood function, given the observations \( \{x(n)\} \), had been

\[
L_n(x(n); z(n); \Psi) = \prod_{j=1}^r \prod_{i=1}^n \left( \int p_{ij}(\Psi) f_{ij}(x(t); \Psi) \right) \left( h(x(n); i, j; \Psi) \right)^{\pi_{ij}(n)},
\]

or, written in additive form,

\[
\log L_n = \log L_{n-1} + s_n, \tag{10}
\]

where

\[
s_n = \sum_{i,j} \pi_{ij}(n) \left( \log p_{ij}(\Psi) + \log f_{ij}(x(n)) \right).
\]

From (10), we get the conditional score function as

\[
h(x(n); \Psi) = E \left\{ \frac{\partial s_n}{\partial \Psi} \mid x(1), \ldots, x(n) \right\},
\]

which is equal to (7). The proof of part (b) is the same as in the simple dependence case; see [5].

Remark: Note that the (8) and (9) for the filtering probabilities are well-known recursive updating formulas, see, e.g., [18].

B. The Adaptive Matrix \( H_n \)

The choice of the adaptive matrix \( H_n \) in the recursion (4) is important for the properties of the algorithm. For independent observations with marginal density \( f(x; \Psi) \) the score function is

\[
h(x; \Psi) = \left( \partial \ln f(x; \Psi) \right)/\partial \Psi, \quad i = 1, \ldots, r \square^2,
\]

and then the optimal choice of \( H_n \) is the inverse of the information matrix, i.e., \( H_n^{-1} = I(\Psi(n)) \), where

\[
I(\Psi) = E \left( h(x(1); \Psi) \cdot h^T(x(1); \Psi) \right).
\]

Recursive estimators with this \( H_n \) were treated by Fabian [19] and proved to be asymptotically normal under suitable regularity conditions, i.e.,

\[
\sqrt{n}(\Psi(n) - \Psi) \rightarrow N(0, I(\Psi)^{-1}).
\]

To avoid the numerical integrations needed to calculate \( H_n^{-1} = I(\Psi(n)) \) we will instead use the inverse of the observed information matrix, i.e.,

\[
H_n^{-1} = n^{-1} \sum_{k=1}^n h(x(k); \Psi(k-1)) \cdot h^T(x(k); \Psi(k-1)).
\]

The matrix \( H_n \) can then be computed recursively by means of the matrix inversion lemma; see e.g. [20, Sect. 1.6]. Writing \( h_n = h(x(n); \Psi(n-1)) \) one has

\[
H_n = \frac{n}{n-1} \left( H_{n-1} + \frac{(n-1)^2}{n} h_n h_n^T \right)^{-1} = \frac{n}{n-1} \left( H_{n-1} - \frac{H_{n-1} h_n h_n^T H_{n-1}}{n} \right). \tag{11}
\]

II. MIXTURES GENERATED BY A MARKOV-MODULATED POISSON PROCESS

A. The Markov-Modulated Poisson Process

Let \( N \) be a Cox point process on the positive real line with random intensity \( \Lambda \). This means that the conditional distribution of \( N \), given that \( \{A(t), t \geq 0\} \), is that of an inhomogeneous Poisson process with intensity function \( \Lambda(t) \). We further assume that the intensity process \( \Lambda(t) \) is a Markov process in continuous time with finitely many states \( \lambda_1, \ldots, \lambda_r \). Then \( N \) is a Markov-modulated Poisson process. The theory presented here, works for any finite number of states \( r \), but we shall give the results explicitly only for the case with two states, \( \lambda_1 \) and \( \lambda_2 \). In that case we denote the jump intensity from state \( i \) by \( \mu_i \), i.e.,

\[
\mu_i = \lim_{h \to 0} h^{-1} P(\Lambda(t+h) = \lambda_j \mid \Lambda(t) = \lambda_i), \quad i \neq j.
\]

We shall define two different random sequences of mixture type where the regime process is a hidden Markov chain generated by the random intensity process. The sequences are obtained by two different ways to observe the MMPP, namely to observe the time intervals between events, or to observe the number of events in fixed time length intervals.

B. Two Observation Models in the MMPP

Model I: For the first model, let the events in \( N \) occur at \( \tau_k \geq 0, k = 0, 1, \ldots \), and let the observations be

\[
x^l_k = \tau_k - \tau_{k-1},
\]

the intervals between occurring events. Then the distribution of the sequence \( \{x^l_k\} \) will depend on a hidden Markov chain \( \{z^l_k\} \), defined as

\[
x^l_k = j \quad \text{if} \quad \Lambda(\tau_k) = \lambda_j,
\]

i.e., the value of the intensity function at the \( k \)th event. Note that \( x^l_k \) is really a Markov chain and that \( \{z^l_k, x^l_k\} \) defines a Markov renewal process.

The dependence between the \( z^l \)- and \( x^l \)-sequences is of the two-step type. For a given value of the regime \( z^l_k \), there is a dependence between the next regime \( z^l_{k+1} \) and the observation \( x^l_k \), but the conditional distribution of \( x^l_k \) given \( z^l_k = k \) and \( z^l_{k+1} \) does not depend on older \( z^l \)-values.
Model II: In the second model we consider a fixed observation interval, without loss of generality taken to have length 1, and let the observed variables be the number of events in the $k$th interval,

$$x^{II}(k) = N(k) - N(k - 1).$$

The hidden Markov chain is then

$$z^{II}(k) = j \text{ if } \Lambda(k) = \lambda_j.$$

Also in this model, the conditional distribution of $z^{II}(k)$ depends on $z^{II}(k - 1)$ and $z^{II}(k)$, but not on older $z^{II}$-values, which means that the dependence is of two-step type.

In both observation models, the pair $(x(k), z(k))$ is the complete data, where $x, z$ can be either of the two processes, and the distribution of the observed data $x(k)$ depends on the unobserved data $z(j)$, for $j = k - 1, k$. These can be calculated analytically or numerically, as we shall now show.

C. Conditional Distributions in the MMPP

In the recursive procedures we need the conditional distribution of the observed data $x(n)$ and the unobserved regime variable $z(n)$ given the previous regime $z(n - 1)$,

$$g(x, i, j; \Psi) = P(z(n) = j | z(n - 1) = i).$$

The formulas below for Model I are known and could be found in, e.g., [7] or [9]. They are presented here for readability and notation.

Model I: Due to stationarity we may consider $n = 1$, assume $\tau_0 = 0$ and write $\tau_1 = \tau$, so that $x(1) = \tau$. Hence we shall calculate the conditional distribution of $\tau = x(1)$ given $\Lambda(0) = i$, and $\Lambda(\tau) = j$. Let $-\delta_k$, $k = 1, 2$ be the different and real roots of the equation

$$(s + \lambda_1 + \mu_1)(s + \lambda_2 + \mu_2) - \mu_1\mu_2 = 0,$$

i.e.

$$s_1, s_2 = \frac{\lambda_1 + \lambda_2 + \mu_1 + \mu_2}{2} \pm \sqrt{\frac{(\lambda_1 - \lambda_2 + \mu_1 - \mu_2)^2}{4} + \mu_1\mu_2}. \quad (12)$$

Further, for $j = 1, 2$ and with $i \neq j$, define

$$A_j^{(1)}, A_j^{(2)} = \frac{\lambda_j}{2}\left\{ \begin{array}{l} \frac{\lambda_j + \mu_j - (\lambda_i + \mu_i)}{2\sqrt{(\lambda_i - \lambda_j + \mu_i - \mu_j)^2 + \mu_1\mu_2}} \\ \frac{\lambda_i + \mu_i - (\lambda_i + \mu_i)}{2\sqrt{(\lambda_i + \lambda_j + \mu_i - \mu_j)^2 + \mu_1\mu_2}} \end{array} \right\} \quad (13)$$

Theorem 2: Model I: The conditional bivariate distribution of $x(n), z(n)$ given $z(n - 1)$ is given by

$$g(x, j; \Psi) = A_j^{(1)} e^{-s_1 x} + A_j^{(2)} e^{-s_2 x}, \quad (14)$$

$$g(x, i, j; \Psi) = \frac{\lambda_j \mu_j}{s_2 - s_1} \{ e^{-s_1 x} - e^{-s_2 x} \} \quad (15)$$

for $j = 1, 2$, and for $i \neq j$, respectively. The filtering probabilities in (8) follow from

$$\pi_{ja}(n) = \pi_{ja}(n - 1)g(x(n), j; k; \Psi)N_{0j},$$
$$\pi_{j}(n) = \sum_i \pi_{ij}(n).$$

Fig. 1. $z^{II}(k - 1)$ influences $z^{II}(k)$ via $x^{II}(k)$.

Fig. 2. $z^{II}(k - 1)$ influences $z^{II}(k)$ and $z^{II}(k)$.

Proof: For shorter notation, write

$$g_{ij}(t) = g(t, i, j; \Psi) = \frac{d}{dt} P(\tau \leq t, \Lambda(\tau) = j | \Lambda(0) = i).$$

To be specific, assume the regime starts at $\Lambda(0) = 1$ and write $s$ for the time of the first transition to $\Lambda(s) = 2$, so $x(1) = \tau$ is the time for the first event in a MMPP with initial intensity $\lambda_1$. Then $g_{11}(t)$ and $g_{21}(t)$ satisfy the following pair of integral equations:

$$g_{11}(t) = \lambda_1 e^{-\lambda_1 t} e^{-\mu_1 t} + \int_0^t \lambda_1 e^{-\lambda_1 s} \mu_1 e^{-\mu_1 s} g_{21}(t - s) ds, \quad (16)$$
$$g_{21}(t) = \int_0^t \lambda_2 e^{-\lambda_2 s} \mu_2 e^{-\mu_2 s} g_{11}(t - s) ds. \quad (17)$$

These equations are of a type common in renewal theory; one may think, e.g., on the first term in (16) as the probability of one event in the Poisson process (with intensity $\lambda_1$) at time $t$ and no state switch, while the integral in the second term represents the integration of the probability of no Poisson event before time $s$, combined with a state switch at time $s$, and finally the first Poisson event (now with intensity $\lambda_2$) after additionally a time $t - s$.) These equations have the solutions (14) and (15) which can be seen, e.g., by taking Laplace transforms.

Remark: The joint density function matrix

$$g(t) = (g_{ij}(t))$$

can be expressed as $g(t) = e^{(Q - \Lambda) t} \Lambda$, where $\Lambda = \text{diag}(\lambda_j)$ and $Q$ is the intensity generator for the hidden Markov process, $Q_{ij} = \mu_{ij}$ for $i \neq j$. The transition probabilities $p_{ij} = P(x(n) = j | z(n - 1) = i)$ are given by, see [9],

$$P = (p_{ij}) = \int_0^\infty e^{(Q - \Lambda) s} A ds = (\Lambda - Q)^{-1} \Lambda. \quad (18)$$

Remark: The state occupancy probabilities $\pi_i(n)$ after the $n$th jump in the process can be calculated by means of the following differential equation derived by Rudemo, see ([16]). If $\overline{\pi}_i(x)$ denotes the occupancy probability for state $i$ at time $x$,

$$\frac{d\overline{\pi}_i(x)}{dx} = -\mu_i \overline{\pi}_i(x) + \mu_j \overline{\pi}_j(x) - \overline{\pi}_i(x) \{\lambda_i - \overline{\lambda}(x)\},$$
$$\pi_i(n) = \overline{\pi}(\tau_i^-) \lambda_i \overline{\lambda}(x),$$
$$\overline{\lambda}(x) = \lambda_1 \overline{\pi}_1(x) + \lambda_2 \overline{\pi}_2(x).$$
where the first equation is valid between the jumps at $\tau_{n-1}$ and $\tau_n$, and the second is the updating equation at the jump time $\tau_n$.

**Model II:** In this model, the observations are $x(n) = N(n) - N(n-1)$, with regime $z(n) = j$ if $A(n) = \lambda_j$. To find and approximate the density $g(x, j, k)$, we shall use the following two random variables:

$$U^{(n)}(t) = \sum_{n-1}^{n-1+t} 1_{\{z(s)=1\}}ds,$$

$$V^{(n)}(t) = \sum_{n-1}^{n-1+t} |dz(s)|,$$

which are equal to the random time between $n-1$ and $n-1+t$ the regime spends in state 1, and the number of switches in the same interval, respectively. Obviously, if $U^{(n)}(t) = u$ then $z(n)$ is a Poisson variable with mean

$$\lambda(u) = \lambda_1 u + \lambda_2 (1-u).$$

Due to stationarity we take $n = 1$ and write $U(t) = U^{(1)}(t)$, $V(t) = V^{(1)}(t)$. Further, define

$$g_i(t, u, v) = \left\{ \begin{array}{ll} \text{conditional density of } (U(t), V(t)), & \text{evaluated at } (u, v), \\
\text{given that } z(0) = i, & \end{array} \right.$$ 

and let $\delta_i(u)$ denote the delta density at the point $t$.

**Lemma 3:** a) The densities $g_i(t; u, v)$ satisfy the following recursions,

$$g_1(t; u, 0) = \delta_t(u)e^{-\mu_1 t},$$

$$g_2(t; u, 0) = \delta_0(u)e^{-\mu_2 t},$$

$$g_1(t; u, 1) = \mu_1 e^{-\mu_1 u} e^{-\mu_2 (t-u)},$$

$$g_2(t; u, 1) = \mu_2 e^{-\mu_2 u} e^{-\mu_1 (t-u)},$$

and, for $v = 1, 2, \ldots$,

$$g_1(t; u, v+1) = \int_0^t \mu_1 e^{-\mu_1 s} g_2(t-s; u-s, v)ds,$$

$$g_2(t; u, v+1) = \int_0^t \mu_2 e^{-\mu_2 s} g_1(t-s; u-s, v)ds.$$

b) The densities $g(x, i, j; \Psi)$ are given by

$$g(x, j, j; \Psi) = \sum_{v=0}^{\infty} g_1(1; u, v)e^{-\lambda(u)} \frac{\lambda(u)^v}{v!} du,$$

$$g(x, i, j; \Psi) = \sum_{v=0}^{\infty} g_2(1; u, v)e^{-\lambda(u)} \frac{\lambda(u)^v}{v!} du,$$

for $j = 1, 2$, and for $i \neq j$, respectively.

**Proof:** For a), just consider the instance $s$ of first switch of $z(s)$. Part b) is just the conditional Poisson distribution of $x(n)$ given $U^{(n)}(1)$.

**Remark:** The transition probabilities $p_{ij} = P(z(n) = j \mid x(n-1) = i), j \neq i$, are given by

$$p_{ii} = \frac{\mu_i}{\mu_1 + \mu_2},$$

$$p_{ij} = \frac{\mu_j}{\mu_1 + \mu_2} e^{-\mu_1 + \mu_2},$$

with $p_{ij} = 1 - p_{ii}$.

**Remark:** The Markov-modulated Poisson process with observation model II is practically useful as a mixture model only when the switching rates $\mu_1$ and $\mu_2$ are small compared to the sampling interval. With large switching rates, the time $U^{(n)}$ spent in state 1 during each observation interval, will tend to be concentrated around its average value $\bar{u} = \mu_2/(\mu_1 + \mu_2)$, and hence the observations $x(n)$ will have a distribution rather like a Poisson distribution with mean $\lambda(\bar{u})$. No information can then be extracted concerning the individual $\lambda_i$- and $\mu_i$-values.

Even if $\mu_1$ and $\mu_2$ are both small, it might still not be possible to estimate $\lambda_i$ by a recursive procedure, since the discriminating influence of each new observation $x(n)(= 0$ or $1)$, might be too small. One could then use a larger sampling interval, but still small compared to the switching rates.

When the switching rates are small, one can approximate the densities $g(x, j, k)$ by the first term in (20) and (21), representing no switch in (20), and only one switch in (21).

**Theorem 4:** Model II) For small switching rates, the conditional distribution of $x(n)$, given $z(n-1)$, can be approximated by

$$g(x, j, ; \Psi) = e^{-\lambda_1 + \lambda_2} \frac{\lambda_j^x}{x!},$$

$$g(x, i, j; \Psi) = \mu_i \int_0^1 e^{-\lambda_1 u + \lambda_2 (1-u)} \{\lambda_1 u + \lambda_2 (1-u)\}^x ds,$$

for $j = 1, 2$, and for $i \neq j$, respectively. The densities (24) will be small compared to those in (23).

The filtering probabilities in (8) can be approximated by

$$\pi_{ij} = \sum_i \pi_{ij}(n),$$

$$\pi_i = \sum_j \pi_{ij}(n).$$

**III. RECURSIVE ESTIMATION**

**A. General Structure**

We shall now use a recursive procedure to estimate the parameters in the two observation models of the MMPP. The most general recursive procedure (4),

$$\Psi^{(n+1)} = \Psi^{(n)} + (n + 1)^{-1} H_n h(x(n+1); \Psi^{(n)})$$

updates estimates of $\Psi = (\lambda_1, \lambda_2, \mu_1, \mu_2)^T$, by means of the conditional score function (7) in Theorem 1,

$$h(x(n); \Psi) = \sum_{i,j} \pi_{ij}(n) \frac{\partial \log g(x(n), i, j; \Psi)}{\partial \Psi}.$$ 

It will turn out that the recursive procedure works well for the Poisson intensities ($\lambda_1, \lambda_2$) and that it converges to the true parameter values even when started far away from the true values. Estimation of the switch intensities ($\mu_1, \mu_2$) seems...
to be more difficult, in particular when the \( \lambda_j \) estimates are started at a bad value. Hence we have also tried to estimate the switch intensities from recursive moment estimates of the transition probabilities; see Section III-B.

The score function in (7) can be calculated analytically in terms of derivatives of the functions \( g(x, i; j; \Psi) \) according to Theorem 2 or of the approximating integrals \( \tilde{g}(x, i; j; \Psi) \) according to Theorem 4. However, the expressions are complex and to save computer time we have approximated the derivatives by the corresponding symmetric difference ratio. Further, as remarked, recursive estimation based on observations according to Model II is possible only for small switching rates, and in that case the approximate expressions in Theorem 4 can be used with sufficient accuracy.

**Remark:** Figs. 3 and 4 illustrate the effect of the approximations (23) and (24). Fig. 3 shows simulations of the average score, i.e., the long run average of the four elements of the score function \( h(x(n); \Psi) \) for one of the MMPP's used in the simulation study in Section IV, parameter set (c). We simulated 200 000 observations and calculated the long run average of the observed scores for \( \lambda_1, \lambda_2 \) and \( \mu_1, \mu_2 \) as functions of \( \mu_1 \) and \( \mu_2 \). The true parameter set used in the simulations was \( \lambda_1 = 2, \lambda_2 = 10, \mu_1 = 0.1, \) and \( \mu_2 = 0.1 \).

The lines show average values of \( h(x(n))_{\mu_i} \) and \( h(x(n))_{\lambda_j} \), as functions of the varying parameter. As seen from the figure the observed average score is zero for the correct parameter value \( \mu_1 = 0.1 \) for all the four quantities estimated. Further, the score for \( \mu_2 \) has the same sign as the score for \( \mu_1 \) which means that if the algorithm has underestimated \( \mu_1 \) it will tend to increase the estimate of \( \mu_1 \) and decrease the estimate of \( \mu_2 \) and vice versa.

Fig. 4 shows the corresponding results for Model II. Here the average score for \( \mu_1 \) is zero for a \( \mu_1 \)-value slightly smaller than the true value, which will lead to a systematic underestimation of \( \mu_1 \).

It can also be seen from the figures that the average scores for \( \lambda_1, \lambda_2 = 1,2 \) and \( \mu_2 \) are small when \( \mu_1 \) varies, except in Model I. This means that the corresponding estimates are rather independent of the estimate of \( \mu_1 \).

In the simulation experiment reported in the next section we found that the fully recursive procedure (4) worked well for the Poisson intensities, while the switch intensities were, as expected, more difficult to estimate. We therefore choose to estimate them by first estimating the state transition probabilities \( p_{ij} \) and then use the relations (18) and (22) between the transition probabilities and the switch intensities, i.e.

- Model I: \( P = (p_{ij}) = (\Lambda - Q)^{-1} \Lambda \)
- Model II: \( p_{ij} = \frac{\mu_j}{\mu_1 + \mu_2} + \frac{\mu_i}{\mu_1 + \mu_2} e^{-(\mu_1 + \mu_2)} \)

Solving these equations for \( \mu_i \) we get, for Model I, \( Q = \Lambda - \Lambda P^{-1} \), i.e. (with \( i \neq j \)),

\[
\mu_i = \frac{\lambda_i p_{ij}}{p_{11} p_{22} - p_{12} p_{21}} = \frac{\lambda_i p_{ij}}{p_{11} - p_{21}}
\]

if \( p_{11} \neq p_{21} \), and for Model II,

\[
\mu_i = \frac{p_{jj} - p_{ii}(p_{11} - p_{21})}{1 - (p_{11} - p_{21})^2} (-\log(p_{11} - p_{21})).
\]

The probabilities \( p_{ij} \) can be estimated from the observed conditional state transition probabilities, and state occupancy probabilities, respectively, \( \pi_{ij}(n) \) and \( \pi_j(n) \). To estimate \( p_{ij} \) we shall use that,

\[
p_{ij} = \lim_{N \to \infty} \frac{\sum_{n=1}^{N} \pi_{ij}(n)}{\sum_{n=1}^{N} \pi_j(n)}.
\]

Since the \( \pi_{ij}(n) \) and \( \pi_j(n) \) are calculated in each observation step one can just form the sums

\[
S_{ij}(N) = \sum_{n=1}^{N} \pi_{ij}(n),
\]

\[
S_i(N) = \sum_{n=1}^{N} \pi_i(n)
\]

and estimate \( p_{ij} \) by \( S_{ij}(N)/S_i(N) \).
TABLE I

<table>
<thead>
<tr>
<th>Case</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2</td>
<td>10</td>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>10</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td>10</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>d</td>
<td>5</td>
<td>10</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>e</td>
<td>2</td>
<td>10</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>f</td>
<td>5</td>
<td>100</td>
<td>0.02</td>
<td>0.1</td>
</tr>
</tbody>
</table>

IV. SOME SIMULATED EXAMPLES

We tried the recursive procedures on a number of parameter combinations in the Markov-modulated Poisson process, and with both observation models. In Model II the observation interval was always taken to have length 1 and instead of making experiments with different interval lengths we adjusted the Poisson and switch intensities, to the same effect. The different parameter combinations are listed in Table I.

<table>
<thead>
<tr>
<th>Parameter Combination in the Simulated MMPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>c</td>
</tr>
<tr>
<td>d</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>f</td>
</tr>
</tbody>
</table>

A. Choice of Initial Values and Other Complications

As all recursive estimation procedures, the recursive algorithm (4) needs to be kept under strict control in particular at the early stages. One needs at least some prior information of the range of possible parameter values, and further one needs initially to modify the optimal choice for the update formula. But provided some care is taken, the recursive likelihood based procedure should be consistent even if this is very hard to prove. Before a recursive likelihood procedure can be implemented in a mixture model, one therefore needs to design a surveillance procedure, perhaps ad hoc, to guarantee convergence to the true values.

First, the recursive procedure (4) needs starting values both for the parameter estimates $\Psi^{(0)}$ and for the adaptive matrix $H_0$. We approximated the adaptive matrix $H_0$ by the inverse of the information matrix $I(\Psi^{(0)})$ obtained by simulation in the model with the initial parameter values $\Psi^{(0)}$. We simulated $N_0 = 100000$ observations in the MMPP and estimated the information matrix by

$$I_N(\Psi^{(0)}) = \frac{1}{N_0} \sum_{k=1}^{N_0} h(x(k); \Psi^{(0)}) \cdot h(x(k); \Psi^{(0)})^T.$$ 

Further, to keep the algorithm under control at the early stages we modified the procedure in the following ways.

1. The denominator in (4) was increased to $n_0 + n_0$ where $n_0 = 100$ was kept fixed.

2. During the first $n_{rec}$ steps in the recursion we used the original adaptive matrix $H_0$. The adaptive matrix was updated according to formula (11) but not used in (4) until after $n_{rec}$ observations. We tried both $n_{rec} = 10$ and $n_{rec} = 100$ in the simulations. The value $n_{rec} = 100$ gave consistently smaller fluctuations of the estimates; these are the values reported in the tables.

3. The estimates were always kept within 0.1 and 10 times to original values

$$0.1 \cdot \psi_j^{(0)} \leq \psi_j^{(n)} \leq 10 \cdot \psi_j^{(0)}.$$ 

B. Simulation Results

We simulated the six cases in Table I with different starting values. The observation intervals in Model II was taken to 1 and we simulated the process over $N_1 = 10000$ observation intervals. In Model I we observed an equivalent number of individual events, i.e. we took approximately

$$N_1 = \frac{\mu_2}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2},$$

observations in each combination, conveniently rounded. We made 100 replicates of each parameter combination.

The results are presented in Tables II-III, which give observed means and standard deviations of the final estimates of the Poisson intensities, and the switch intensities (estimated through the transition estimates in Section III-B). Table II shows the results when the algorithm is started at the true parameter values, while Table III contains results for various incorrect starting values.
As seen, the estimates are stable around the correct values when the algorithm is started at the correct parameter point. Similar experiments with the constant $n_{rec} = 10$ behaved less stable, and had a tendency to occasionally diverge at an early stage of the simulation. This tendency still exists in the reported study, but to a much lesser extent. The final estimates followed a normal distribution reasonably well.

The estimates in Table III are the results when the algorithm is started at quite some distance from the true point. To gain experience we tried different off-sets for the six different combinations. Not unexpectedly, the Poisson parameters were easy to estimate and the estimates of $\lambda_j$ were held fixed than when they were estimated, even if their starting point was correct, (these results are not shown in the table). This was true also when the $\mu_j$ were held fixed at twice the correct value.

The $\mu_j$-estimates were much more sensitive to the starting values. They seemed to converge, but at a very slow rate with bad starting values. This also had consequences in the situation when the $\lambda_j$-estimates got bad starting values. Then the $\lambda_j$-estimates converged, but the $\mu_j$-estimates drifted away even when started correctly. They still ultimately converged, but very slowly. This indicates that one should try a procedure where the $\mu_j$-estimates are given an alternative to move more freely after that the $\lambda_j$-estimates have stabilized.

### REFERENCES


Ulla Holst received the M.Sc. degree in engineering physics in 1970 and the Ph.D. degree in Mathematical statistics in 1982 from Lund Institute of Technology, Lund, Sweden. She is associate professor in mathematical statistics at the Lund Institute of Technology, Lund, Sweden. Her current research interests include recursive estimation, estimation in hidden Markov models, nonparametric curve estimation, and environmental statistics.