An EM algorithm for Batch Markovian Arrival Processes and its comparison to a simpler estimation procedure

Lothar Breuer (breuer@info04.uni-trier.de)
University of Trier, Germany

Abstract. Although the concept of Batch Markovian Arrival Processes (BMAPs) has gained widespread use in stochastic modelling of communication systems and other application areas, there are few statistical methods of parameter estimation proposed yet. However, in order to practically use BMAPs for modelling, statistical model fitting from empirical time series is an essential task. The present paper contains a specification of the classical EM algorithm for MAPs and BMAPs as well as a performance comparison to the computationally simpler estimation procedure recently proposed by Breuer and Gilbert. Furthermore, it is shown how to adapt the latter to become an estimator for hidden Markov models.

Keywords: Markovian Arrival Process, EM algorithm, maximum likelihood estimation, hidden Markov models

AMS codes: 62M05, 60J75, 60J27, 60K25

1. Introduction

Markovian Arrival Processes (MAPs) and Batch Markovian Arrival Processes (BMAPs) have been introduced by Neuts (1979) and Lucantoni (1991) in order to provide input streams for queueing systems which are Markovian (and hence analytically more tractable) on the one hand but very versatile (even dense in the class of point processes, see Asmussen and Koole (1993)) on the other hand. This concept has proved very successful in queueing theory for more than twenty years now. For a bibliography demonstrating this, see Lucantoni (1993).

Although the concept of BMAPs has gained widespread use in stochastic modelling of communication systems and other application areas, there are few statistical methods of parameter estimation proposed yet. A survey of estimation methods is given in Asmussen (1997). His emphasis is on maximum likelihood estimation and its implementation via the EM algorithm. For the Markov Modulated Poisson Process (MMPP), an EM algorithm has been developed by Ryden (1993; 1994; 1996; 1997), whereas Asmussen et al. (1996) derived a fitting procedure for phase-type distributions via the EM algorithm. The single existing likelihood–oriented procedure that has been introduced for BMAPs up to now can be found in Breuer and Gilbert (2000) or Breuer (2000b) (see Breuer (2000a) for a special case).

However, in order to practically use BMAPs for modelling, statistical model fitting from empirical time series is an essential task. The present paper contains a specification of the classical EM algorithm (see Dempster et al. (1977)) for MAPs and BMAPs as well as a performance comparison to the computationally simpler estimation procedure proposed in Breuer and Gilbert (2000). In section 2, the EM algorithm will be specified for MAPs and BMAPs. Section 3 contains a short description of the estimation procedure for BMAPs introduced in Breuer and Gilbert (2000). In section 4 numerical results of the two procedures are compared. Finally, section 5 provides an adaptation of the procedure given in section 3 to the class of hidden Markov models. In order to understand the details of the estimation procedures to be introduced in the present paper, it is strongly recommended to be familiar with former EM specifications for PH distributions (see Asmussen (1996)) and for the MMPP (see Ryden (1996)), which is a special case of the MAP.

Assume that the empirical information observed from an input stream into a queueing system consists of the time instants $(T_1, \dots, T_N)$ of single arrivals (for MAPs) or the time instants $(T_1, \dots, T_N)$ of arrivals together with their batch sizes $(b_1, \dots, b_N)$, with N denoting the number of observed arrivals. The task is to find parameters D_0 and D_1 for a MAP (or $D_0, D_1, \dots, D_M$, with maximal batch size M, for a BMAP) that optimally fit this arrival stream.

In both estimation procedures that will be introduced, we fix the number of phases for the BMAP model to be a known integer $m \geq 2$. Procedures for estimating the number m of phases are discussed in Ryden (1997). A feasible method without a prior estimation of m is proposed in Jewell (1982) as follows. Denote the estimators (as given by the EM algorithm and the simpler procedure described below) for an assumed number m_k of phases by $(\hat{D}_0(k), \hat{D}_1(k), \dots, \hat{D}_M(k))$. Estimating the parameters by the methods given below for increasing m_k and stopping as soon as the likelihood ratio

$$\frac{f(z|\hat{D}_0(k+1), \hat{D}_1(k+1), \dots, \hat{D}_M(k+1))}{f(z|\hat{D}_0(k), \hat{D}_1(k), \dots, \hat{D}_M(k))}$$

is smaller than a threshold value $1 + \varepsilon$ leads to a reasonable model fitting. Since the adaptation of the model increases with the assumed number of phases, the likelihood gain is always positive. The threshold value reflects the limit of accuracy beyond which the gain in model adaptation is not worth the additional computation time.

This method has been applied to BMAP estimation via the procedure described in section 3 by Breuer and Gilbert (2000). Many numerical results can be found in Gilbert (2000). In particular, it can be seen there that the number of phases of the estimated BMAP representation does not always coincide with the number of phases of the input BMAP. Nevertheless, in these cases the likelihood of the arrival stream under the estimated representation is about the same as the likelihood under the original input BMAP.

2. An EM algorithm for MAPs and BMAPs

The typical property of observing time series derived from a MAP (resp. a BMAP) is that only the time instants of arrivals (and their batch sizes, resp.) but not the phases can be seen. If the phases were observable, then one could apply classical estimators for finite state Markov processes (see Albert (1962) or Basawa, Rao (1980)) and the problem would be solved. Thus we have a problem of estimation from incomplete data. For this type of statistical problems, the so–called EM algorithm has proven to be a good means of approximating the maximum likelihood estimator (see Dempster et al. (1977), McLachlan, Krishnan (1997) or Meng, Dyk (1997)). The name EM algorithm stems from the alternating application of an expectation step (for E) and a maximization step (for M) which yield successively higher likelihoods of the estimated parameters.

In this section we first describe the classical estimators which were applicable if we had the complete sample for a MAP (i.e. if phases were observable). Then we derive the specification of the EM algorithm for MAPs, assuming that phases are not observable (incomplete sample). Finally, this specification is extended to obtain an EM specification for BMAPs.

2.1. Complete sample case for MAPs

A sufficient statistic for the complete data sample would be the collection

$$\mathcal{S} = (J_n^k : 0 \leq n \leq M_k, 1 \leq k \leq N), (S_n^k : 0 \leq n \leq M_k - 1, 1 \leq k \leq N)$$

of random variables, where J_n^k denotes the phase immediately after the nth jump of the phase process in the kth interval between arrivals (i.e. in the interval $[T_{k-1}, T_k[$), M_k the number of jumps in this interval (not including the last jump accompanied by an arrival), and S_n^k the sojourn time after the nth jump of the phase process in the interval $[T_{k-1}, T_k[$.

Then the density of a complete sample x under parameters $D_0 = (D_{0;ij})$ and $D_1 = (D_{1;ij})$ would be given by

$$f(x|D_0, D_1) = \prod_{i=1}^{m} \exp\left(D_{0;ii} Z_i\right) \prod_{i=1}^{m} \prod_{j=1, j\neq i}^{m} D_{0;ij}^{N_{ij}} \prod_{i=1}^{m} \prod_{j=1}^{m} D_{1;ij}^{L_{ij}}$$

where Z_i denotes the total time spent in phase i, N_{ij} the number of jumps from phase i to phase j without accompanying arrival, and L_{ij} the number of jumps from phase i to phase j with accompanying arrival. These variables can be computed from the sufficient statistic by

$$Z_i = \sum_{k=1}^{N} \sum_{n=0}^{M_k - 1} 1_{(J_n^k = i)} S_n^k, \qquad N_{ij} = \sum_{k=1}^{N} \sum_{n=0}^{M_k - 1} 1_{(J_n^k = i)} 1_{(J_{n+1}^k = j)}$$

and

$$L_{ij} = \sum_{k=1}^{N-1} 1_{(J_{M_k}^k = i)} 1_{(J_0^{k+1} = j)}$$

for $1 \leq i \neq j \leq m$. Further define

$$Y_i = \sum_{k=1}^{N-1} \sum_{n=0}^{M_k - 1} 1_{(J_n^k = i)} S_n^k$$

for $1 \leq i \leq m$.

Acknowledging the relation $D_{0;ii} = -\left(\sum_{j=1}^m D_{1;ij} + \sum_{j=1, j\neq i}^m D_{0;ij}\right)$, the maximum likelihood estimators $\hat{D}_0$ and $\hat{D}_1$ for the matrices D_0 and D_1 would be

$$\hat{D}_{0;ij} = \frac{N_{ij}}{Z_i}, \qquad \hat{D}_{1;ij} = \frac{L_{ij}}{Y_i}, \tag{1}$$

$$\hat{D}_{0;ii} = -\left(\sum_{j=1}^m \hat{D}_{1;ij} + \sum_{j=1, j\neq i}^m \hat{D}_{0;ij}\right) \tag{2}$$

for $1 \leq i, j \leq m$, as given in Albert (1962).

2.2. EM FOR MAPs

In the case of observing only an incomplete sample z, the EM algorithm provides an iteration of alternating expectation (E) and maximization (M) steps that lead to a reevaluation of the estimators increasing their likelihoods in every cycle of E- and M–step. In our case, the incomplete sample y consists only of the sequence $(T_0 = 0, T_1, \ldots, T_N)$ of inter–arrival times that are observable. Keeping in mind $T_0 = 0$, we will not lose information by setting $z = (z_1, \ldots, z_N) := (T_1, T_2 - T_1, \ldots, T_n - T_{N-1})$.

Given the parameters D_0 and D_1 as well as an initial phase distribution π, the likelihood of the incomplete sample z is

$$f(z|\pi, D_0, D_1) = \pi \left(\prod_{n=1}^{N-1} \exp(D_0 z_n) D_1\right) \exp(D_0 z_N) \eta \tag{3}$$

with $\eta := D_1 1_m$, denoting by 1_m the m–dimensional column vector with all entries being 1.

Assume that the estimates after the kth EM iteration are given by the matrices $(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})$. Then in the first step of the $k+1$st cycle, the conditional expectations of the variables Z_i, N_{ij} and L_{ij} given the incomplete observation y and the current estimates $(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})$ are computed.

In order to simplify notations, define

$$\eta_N(i) := \sum_{j=1}^{m} \hat{D}_1^{(k)}(i,j) \qquad \text{and} \qquad \eta_{n-1}(i) := \hat{D}_1^{(k)} \exp(\hat{D}_0^{(k)} z_n)\eta_n \quad (4)$$

for $2 \leq n \leq N$ and $i = 1, \ldots, m$.

Since the empirical time series is observed in a stationary regime, we can set the phase distribution π_0 at time $T_0 = 0$ to be the phase equilibrium, i.e. satisfying $\pi_0(\hat{D}_0^{(k)} + \hat{D}_1^{(k)}) = 0$. Thus π_0 is a deterministic function of $(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})$ and in particular $f(z|\pi_0, \hat{D}_0^{(k)}, \hat{D}_1^{(k)}) = f(z|\hat{D}_0^{(k)}, \hat{D}_1^{(k)})$ holds. Next we define iteratively

$$\pi_{n+1} := \pi_n \exp(\hat{D}_0^{(k)} z_{n+1}) \; \hat{D}_1^{(k)} \tag{5}$$

for $0 \leq n \leq N-2$, interpreting the π_n as row vectors.

We can continue the E–step with

$$Z_i^{(k+1)} := E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(Z_i|z) = \sum_{n=1}^{N} E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(Z_i^n|z)$$

and

$$Y_i^{(k+1)} := E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(Y_i|z) = \sum_{n=1}^{N-1} E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(Z_i^n|z)$$

where Z_i^n denotes the random variable of the total amount of time within $[T_{n-1}, T_n[$ that is spent in phase i. This is given by

$$E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(Z_i^n|z) = \frac{c_n(i, i|z, \hat{D}_0^{(k)}, \hat{D}_1^{(k)})}{f(z|\hat{D}_0^{(k)}, \hat{D}_1^{(k)})} \tag{6}$$

for all $1 \leq l \leq N$, with c_n given as in definition (7) below. The derivation of (6) is completely analogous to the one in Asmussen et al. (1996), p.439. Likewise, the E–step for

$$N_{ij}^{(k+1)} := E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(N_{ij}|z) = \sum_{n=1}^{N} E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(N_{ij}^n|z)$$

with

$$E_{(\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}(N_{ij}^n|z) = \frac{\hat{D}_{0;ij}^{(k)} c_n(i, j|z, \hat{D}_0^{(k)}, \hat{D}_1^{(k)})}{f(z|\hat{D}_0^{(k)}, \hat{D}_1^{(k)})}$$

for $1 \leq n \leq N$ is derived using completely the same arguments as in Asmussen et al. (1996), p.440. Here, the matrix functions c_n are defined as

$$c_n(i,j|z,\hat{D}_0^{(k)},\hat{D}_1^{(k)}) := \int_0^{z_n} \pi_{n-1} \exp{(D_0^{(k)} u)} e_i \cdot \\ \cdot e_j^T \exp{(D_0^{(k)}(z-u))} \eta_n \, du \quad (7)$$

for $1 \leq n \leq N$ and $1 \leq i,j \leq m$.

The E–step is completed by

$$L_{ij}^{(k+1)} := E_{(\hat{D}_0^{(k)},\hat{D}_1^{(k)})}(L_{ij}|z) = \sum_{n=1}^{N-1} E_{(\hat{D}_0^{(k)},\hat{D}_1^{(k)})}(L_{ij}^n|z)$$

with

$$\begin{aligned} E_{(\hat{D}_0^{(k)},\hat{D}_1^{(k)})}(L_{ij}^n|z) &= P(J_{M_n}^n = i, J_0^{n+1} = j | z_1, \ldots, z_n, z_{n+1}, \ldots, z_N) \\ &= \frac{1}{f(z)} f(J_{M_n}^n = i, z_1, \ldots, z_n) P(J_0^{n+1} = j | J_{M_n}^n = i) \cdot \\ &\qquad \cdot f(z_{n+1}, \ldots, z_N | J_0^{n+1} = j) \\ &= \frac{\left(\pi_{n-1} \exp(D_0^{(k)} z_n)\right)_i \hat{D}_1^{(k)}(i,j) \left(\exp(D_0^{(k)} z_{n+1}) \eta_{n+1}\right)_j}{f(z|\hat{D}_0^{(k)},\hat{D}_1^{(k)})} \end{aligned}$$

for $1 \leq n \leq N-1$.

The second step of the $k+1$st cycle consists of the computation of maximum likelihood estimates given the new (conditional but complete) statistic computed in the E–step. This can be done by simply replacing the variables in equations (1) and (2) by the conditional expectations computed above. This leads to reevaluated estimates

$$\hat{D}_{0;ij}^{(k+1)} = \frac{N_{ij}^{(k+1)}}{Z_i^{(k+1)}}, \qquad \hat{D}_{1;ij}^{(k+1)} = \frac{L_{ij}^{(k+1)}}{Y_i^{(k+1)}},$$

and

$$\hat{D}_{0;ii}^{(k+1)} = -\left(\sum_{j=1}^{m} \hat{D}_{1;ij}^{(k+1)} + \sum_{j=1, j\neq i}^{m} \hat{D}_{0;ij}^{(k+1)}\right)$$

for $1 \leq i,j \leq m$.

Using these, one can compute the likelihood $f(z|\hat{D}_0^{(k+1)},\hat{D}_1^{(k+1)})$ of the empirical time series under the new estimates according to equation (3). If the likelihood ratio

$$\rho = \frac{f(z|\hat{D}_0^{(k+1)},\hat{D}_1^{(k+1)})}{f(z|\hat{D}_0^{(k)},\hat{D}_1^{(k)})}$$

remains smaller than a threshold $1+\varepsilon$, then the EM iteration process can be stopped, and the latest estimates may be adopted. The threshold value reflects the limit of accuracy beyond which the gain in model adaptation is considered not to be worth the additional computation time.

2.3. EM FOR BMAPs

Once we have obtained the above EM specification for MAPs, an extension to BMAPs is straightforward. In the BMAP case we not only know the times $(T_1,\ldots,T_N)$ of arrivals but also the arrival sizes $(b_1,\ldots,b_N)$, meaning that at time instant T_n there was a batch arrival of size b_n, with $n=1,\ldots,N$, $b_n \in I\!N$. In order to obtain reasonable estimates, we need an upper bound M such that $D_n = 0$ for all $n \geq M$. Furthermore, we need enough arrivals to ensure that there are reasonably many arrival events of every size in the time series.

Starting from the kth estimates $(\hat{D}_0^{(k)},\hat{D}_1^{(k)},\ldots,\hat{D}_M^{(k)})$ for $(D_0,\ldots,D_M)$, the $k+1$st EM iteration proceeds completely analogous as for MAPs, with obvious adaptations in equations (3), (4) and (5). The only more substantial difference is that for BMAPs we need to compute $L_{1;ij}^{(k+1)},\ldots,L_{M;ij}^{(k+1)}$ instead of only $L_{ij}^{(k+1)}$. These are given similarly by

$$L_{s;ij}^{(k+1)} := E_{(\hat{D}_0^{(k)},\hat{D}_1^{(k)})}(L_{s;ij}|z) = \sum_{n=1,b_n=s}^{N-1} E_{(\hat{D}_0^{(k)},\hat{D}_1^{(k)})}(L_{s;ij}^n|z)$$

for $s=1,\ldots,M$, with

$$E_{(\hat{D}_0^{(k)},\hat{D}_1^{(k)})}(L_{s;ij}^n|z) = \\ = \frac{\left(\pi_{n-1}\exp(D_0^{(k)}z_n)\right)_i \hat{D}_s^{(k)}(i,j) \left(\exp(D_0^{(k)}z_{n+1})\eta_{n+1}\right)_j}{f(z|\hat{D}_0^{(k)},\hat{D}_1^{(k)},\ldots,\hat{D}_M^{(k)})}$$

for $1 \leq n \leq N-1$.

These expectations lead to new estimates

$$\hat{D}_{s;ij}^{(k+1)} = \frac{L_{s;ij}^{(k+1)}}{Y_i^{(k+1)}}$$

for $s=1,\ldots,M$.

3. A simpler estimation procedure

In Breuer, Gilbert (2000), a computationally much lighter estimation procedure has been introduced. It works in three steps, each of which uses classical statistical methods. In the first step (section 3.1), the empirical interarrival times are used to estimate the matrix D_0, neglecting the correlations between consecutive interarrival times. Then those can be interpreted as a sample of phase–type distributions and hence the entries of D_0 can be estimated by an EM–algorithm for phase–type distributions (see Asmussen et al. (1996)). In the second step (section 3.2), for every empirical arrival instant the probability distribution of being in a certain phase immediately before resp. after this instant is estimated using discriminant analysis (see Titterington et al. (1985)). In the last step (section 3.3), the derived estimators of the first two steps are used in order to calculate the empirical estimator for the matrices D_n, $n \geq 1$. This is done according to standard estimators for Markov chains (see Anderson, Goodman (1957)).

3.1. Estimating the Matrix D_0

As in section 2, let $(z_n := T_n - T_{n-1} : n \in \{1, \ldots, N\})$ denote the empirical interarrival times. Denote the number of phases by m. According to Baum (1996), p.42, the interarrival times of a BMAP are distributed phase–type with generator D_0. Hence the $(z_n : n \in \{1, \ldots, N\})$ are a sample of a phase–type distribution with density

$$z(t) = \pi e^{D_0 t} \eta$$

for $t \in I\!R_+$. Here, $\pi = (\pi_1, \ldots, \pi_m)$ is the steady–state distribution of the phase process at arrival instants and $\eta := -D_0 1_m$ is the so–called exit vector of the phase–type distribution with representation (π, D_0).

If m is known, there is a maximum likelihood estimator for π and D_0. The solution of the estimating equations can be approximated iteratively by an EM algorithm (cf. Dempster et al. (1977) or McLachlan, Krishnan (1997)), which was derived for this special case by Asmussen et al. (1996) and proceeds as follows.

Starting from an intuitive first estimation $(\pi^{(1)}, D_0^{(1)})$ of the representation of the phase–type distribution, the recursions

$$\pi_i^{(k+1)} = \frac{1}{N} \sum_{n=1}^{N} \frac{\pi_i^{(k)} b_i^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)}$$

$$D_{0;ij}^{(k+1)} = \sum_{n=1}^{N} \frac{D_{0;ij}^{(k)} c_{ij}^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)} \Bigg/ \sum_{i=1}^{N} \frac{c_{ii}^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)}$$

for $i \neq j$ and

$$\eta_i^{(k+1)} = \sum_{n=1}^{N} \frac{\eta_i^{(k)} a_i^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)} \Big/ \sum_{i=1}^{N} \frac{c_{ii}^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)}$$

along with the relation

$$D_{0;ii}^{(k+1)} = -\eta_i^{(k+1)} - \sum_{j=1, j\neq i}^{m} D_{0;ij}^{(k+1)}$$

and the definitions

$$\begin{aligned} a^{(k)}(z_n) &:= \pi^{(k)} e^{D_0^{(k)} z_n} \\ b^{(k)}(z_n) &:= e^{D_0^{(k)} z_n} \eta^{(k)} \\ c_{ij}^{(k)}(z_n) &:= \int_0^{z_n} \pi^{(k)} e^{D_0^{(k)} u} e_i e_j^T e^{D_0^{(k)}(z_n - u)} \eta^{(k)} \, du \end{aligned}$$

for $i, j \in \{1, \ldots, m\}$ und $k \in I\!N$ lead to monotonically increasing likelihoods.

In Asmussen et al. (1996), it is proposed to compute the values of $a^{(k)}(z_n)$, $b^{(k)}(z_n)$ and $c_{ij}^{(k)}(z_n)$ numerically as the solution to a linear system of homogeneous differential equations.

3.2. Phases at Arrival Instants

Using the estimator $(\hat{\pi}, \hat{D}_0)$ from the last section, the distribution of the non–observable phases at times $(T_n : n \in I\!N)$ can be estimated using discriminant analysis in a standard way (cf. Titterington et al. (1985), pp.168f).

For a given empirical arrival instant T_n, let $Z_n := T_n - T_{n-1}$ and R_{n-1} denote the random variables of the last interarrival time and of the phase immediately after the last arrival instant T_{n-1}, respectively. Let $P(Z_n, R_{n-1})$ denote their common distribution and $P(Z_n|R_{n-1})$, $P(R_{n-1}|Z_n)$ the conditional distributions.

Since $(Z_n : n \in \{1, \ldots, N\})$ is given empirically by the time series $(T_n : n \in \{0, \ldots, N\})$, it suffices to estimate the distribution $P(R_{n-1}|Z_n = z_n)$ for every $n \in \{1, \ldots, N\}$. This distribution is discrete, because by assumption there are only finitely many phases.

Bayes' formula yields for $j \in \{1, \ldots, m\}$ and $n \in \{1, \ldots, N\}$

$$P(R_{n-1} = j | Z_n = z_n) = \frac{P(Z_n = z_n | R_{n-1} = j) \cdot P(R_{n-1} = j)}{\sum_{i=1}^{m} P(Z_n = z_n | R_{n-1} = i) \cdot P(R_{n-1} = i)}$$

The expressions $P(Z_n|R_{n-1})$ exist as conditional densities (with respect to the Lebesgue measure on $I\!R$), since the interarrival times Z_n are distributed phase–type.

The expressions on the right hand can be estimated via the estimated parameters $(\hat{\pi}, \hat{D}_0)$ and the resulting vector $\hat{\eta}$. Hence, for every arrival instant T_{n-1} the estimator for the conditional distribution of the phase at T_{n-1} given the interarrival time $Z_n = T_n - T_{n-1}$ is given by

$$\hat{P}(R_{n-1} = i | Z_n = z_n) = \frac{e_i^T e^{\hat{D}_0 z_n} \hat{\eta} \cdot \hat{\pi}_i}{\sum_{j=1}^m e_j^T e^{\hat{D}_0 z_n} \hat{\eta} \cdot \hat{\pi}_j} = \frac{\hat{\pi}_i \cdot e_i^T e^{\hat{D}_0 z_n} \hat{\eta}}{\hat{\pi} e^{\hat{D}_0 z_n} \hat{\eta}}$$

for every $i \in \{1, \ldots, m\}$ and $n \in \{1, \ldots, N\}$.

Furthermore, it will be necessary to estimate the phase immediately before an arrival instant. This can be done by the same method. Denote the respective conditional distribution by $P(R_{n-} | Z_n)$. Then we get the estimation

$$\hat{P}(R_{n-} = i | Z_n = z_n) = \frac{\hat{\pi} e^{\hat{D}_0 z_n} e_i \cdot \hat{\eta}_i}{\sum_{j=1}^m \hat{\pi} e^{\hat{D}_0 z_n} e_j \cdot \hat{\eta}_j} = \frac{\hat{\pi} e^{\hat{D}_0 z_n} e_i \cdot \hat{\eta}_i}{\hat{\pi} e^{\hat{D}_0 z_n} \hat{\eta}}$$

for every $i \in \{1, \ldots, m\}$ and $n \in \{1, \ldots, N\}$.

3.3. Estimating the Matrices D_n for $n \geq 1$

The matrices $(D_n : n \in \mathbb{N}_0)$, which are the blocks of the generator matrix Q of a BMAP of order m, have dimension $(m \times m)$ and the form

$$D_n(i, j) = -D_0(i, i) \cdot p_i(n, j)$$

for every $i, j \in \{1, \ldots, m\}$.

In order to complete the estimation of the generator matrix, it suffices to estimate the parameters $(p_i(n, j) : n \in \mathbb{N}, i, j \in \{1, \ldots, m\})$, which are the transition probabilities of the embedded Markov chain at arrival instants. Without any further assumptions regarding these, use of the empirical estimator is standard. This is given in Anderson, Goodman (1957). Since in the present statistical model the phase process is hidden, the phase change at each arrival instant cannot be observed but must be estimated. For this, the results of the last section are used.

For every $k \in \{0, \ldots, N-1\}$, let R_k and R_{k-} denote the random variable of the (non-observable) phase at the empirical arrival instant T_k and immediately before it, respectively. In the last section, estimators for the conditional distributions of the R_k and R_{k-} were given. Let δ denote the Kronecker function and remember that b_k denotes the size of the kth batch arrival. Define

$$n_i(n, j) := \sum_{k=1}^{N-1} \hat{P}(R_{k-} = i | Z_k = z_k) \cdot \delta_{b_k, n} \cdot \hat{P}(R_k = j | Z_{k+1} = z_{k+1})$$

and

$$n_i := \sum_{k=1}^{N-1} \hat{P}(R_{k-} = i | Z_k = z_k) = \sum_{n=1}^{\infty} \sum_{j=1}^{m} n_i(n, j)$$

for $n \in I\!N$ and $i, j \in \{1, \dots, m\}$. In a BMAP, the random variables R_{k-} and R_k are dependent in any non–trivial case. Since in the present model the phases are non–observable, this dependency can only be reflected by conditioning on the consecutive empirical interarrival times z_k and z_{k+1}.

Because of

$$p_i(n, j) = \frac{P(R_{k-} = i, b_k = n, R_k = j)}{P(R_{k-} = i, b_k \geq 1)} \cdot \frac{P(R_{k-} = i, b_k \geq 1)}{P(R_{k-} = i)}$$

for all $k \in \{1, \dots, N-1\}$, the empirical estimator for $p_i(n, j)$ is given by

$$\hat{p}_i(n, j) := \frac{n_i(n, j)}{n_i} \cdot \frac{\hat{\eta}_i}{-\hat{D}_{0,ii}}$$

for every $n \in I\!N$ and $i, j \in \{1, \dots, m\}$.

4. Numerical results and comparison of the two procedures

The following computations were done on a 500 MHz processor using the programming language Octave (a free MatLab version) under Linux. Tables 1 through 3 are results from time series of 100 arrivals, while tables 4 through 6 give estimates from 500 arrivals. The numerical examples were only computed for MAPs, since the estimator for BMAPs is essentially the same but BMAPs would require much larger time series and thus longer run times. The run times for the EM algorithm are further the main reason of not testing time series with more arrivals.

At first glance it can be observed that higher likelihoods do not necessarily mean that the estimated parameters are closer to the original parameters. As a maximum likelihood estimator, the EM algorithm simply tries to find any set of parameters which yields the highest likelihood of the observed time series. The numerical results show that the simpler estimator achieves almost the same likelihood as the EM algorithm while performing much faster and requiring less storage. Since phase–type distributions (and thus BMAPs) in general have no unique representation (see e.g. Botta et al. (1987)), it is not reasonable to search for such one. The aim of model fitting is simply to find parameters for a BMAP that make the observed time series under the BMAP model at least as likely as under the original parameters. In all numerical examples with time series derived from MAPs, this happened to be the case for both procedures.

Since MAPs are extensively used as models for arrival streams in queueing systems, a more reasonable measure for the goodness-of-fit of the tested estimation procedures might be a collection of performance measures for queues with the respective inputs. One such performance measure is the so–called caudal characteristic η (see (Neuts, 1986) or (Latouche and Ramaswami, 1999)). The value of η is a good description of the tail behaviour of the stationary distribution of the queue process.

In the following numerical examples, for every arrival process (as well as for the estimators) the caudal characteristic η of the respective $MAP/M/1$ queue has been computed. This has been computed as the largest eigenvalue of the rate matrix R of the QBD Markov chain embedded at jump times of the queue (including phase changes). The rate of the exponential service time distribution has been set to equal the highest arrival rate (i.e. the absolute value of the lowest entry of the input MAP's D_0). The numerical results show that the caudal characteristics computed from the two estimators are very close and (except in tables 1 and 4) approximate the caudal characteristic of the input MAP well enough.

The EM algorithm yields consistently higher likelihoods than the simpler procedure. Yet, comparing the gain in likelihood per arrival and the total run time, the question remains whether an EM procedure is worth the effort. Although the likelihoods of the estimated parameters are satisfying for both procedures, none of them can approach the original parameters in terms of a maximal distance norm. This can be seen best from the estimated values for D_1. However, in terms of the caudal characteristic for the respective queues with exponential service time distribution, the goodness-of-fit is much better. The EM algorithm uses all the available information in all iterations. Any improvement of the procedure introduced in section 3 will perform at most as good as the EM algorithm.

Table 1	Input MAP		Estimation by EM		Simple estimation	
$D_0 =$	-10.0	5.0	-10.5	5.4	-15.3	10.1
	50.0	-100.0	30.4	-78.2	59.4	-104.2
$D_1 =$	4.0	1.0	2.8	2.2	2.6	2.7
	10.0	40.0	26.7	21.1	22.1	22.7
run time:			6m 27s		1m 4s	
likelihood:	3.58 E58		1.62 E59		9.86 E58	
η	0.2445		0.1777		0.1752	

Table 2	Input MAP		Estimation by EM		Simple estimation	
$D_0 =$	-5.0	2.0	-3.5	1.4	-5.0	2.5
	5.0	-10.0	8.0	-21.4	5.1	-10.3
$D_1 =$	2.0	1.0	1.8	0.4	1.7	0.9
	3.0	2.0	10.1	3.3	3.4	1.8
run time:			1m 9s		8s	
likelihood:	9.08 E05		5.74 E06		1.51 E06	
η	0.3611		0.3583		0.3432	

Tab 3	Input MAP			Estimation by EM			Simple estimation		
$D_0 =$	-5.0	1.0	1.0	-7.8	2.7	2.6	-11.9	4.0	4.0
	2.0	-10.0	1.0	4.6	-14.7	4.3	6.2	-18.0	6.0
	10.0	2.0	-20.0	6.7	6.3	-23.7	8.7	8.4	-25.5
$D_1 =$	2.0	1.0	0.0	0.9	0.8	0.8	1.7	1.2	0.9
	2.0	4.0	1.0	2.0	1.8	2.0	2.6	1.8	1.4
	1.0	5.0	2.0	3.6	3.3	3.7	3.8	2.6	2.0
time:				27m 56s			14s		
likel.:	1.51 E26			3.43 E26			1.28 E26		
η	0.2627			0.2744			0.2684		

Table 4	Input MAP		Estimation by EM		Simple estimation	
$D_0 =$	-10.0	5.0	-9.5	4.9	-11.2	7.0
	50.0	-100.0	55.9	-119.9	67.2	-123.6
$D_1 =$	4.0	1.0	3.3	1.3	2.2	2.1
	10.0	40.0	28.3	35.6	28.4	28.0
run time:			27m 12s		4m 14s	
likelihood:	1.10 E269		1.17 E270		2.85 E269	
η	0.2445		0.2096		0.1816	

Table 5	Input MAP		Estimation by EM		Simple estimation	
$D_0 =$	-5.0	2.0	-5.4	2.6	-5.6	2.7
	5.0	-10.0	4.8	-10.2	5.0	-10.4
$D_1 =$	2.0	1.0	1.8	0.9	1.9	1.0
	3.0	2.0	3.5	1.8	3.5	1.8
run time:			11m 43s		26s	
likelihood:	3.59 E59		3.98 E59		3.91 E59	
η	0.3611		0.3456		0.3639	

Tab 6	Input MAP			Estimation by EM			Simple estimation		
$D_0 =$	-5.0	1.0	1.0	-5.7	1.8	1.9	-7.0	2.4	2.4
	2.0	-10.0	1.0	4.8	-17.8	5.1	7.0	-21.2	6.1
	10.0	2.0	-20.0	6.9	6.8	-24.6	10.3	8.7	-30.7
$D_1 =$	2.0	1.0	0.0	0.7	0.8	0.6	0.9	0.8	0.6
	2.0	4.0	1.0	2.8	3.0	2.2	3.2	2.9	2.1
	1.0	5.0	2.0	4.0	4.1	2.9	4.6	4.2	3.0
time:				51m 42s			2m 36s		
likel.:	8.53 E117			4.90 E118			1.35 E118		
η	0.2627			0.2869			0.2845		

5. Estimating hidden Markov models

There is a close relationship between MAPs and a very general class of Markov chains called hidden Markov models (see e.g. Elliott et al. (1995)). On the one hand, using the construction in Çinlar (1969), p.384f, one can represent BMAPs as special hidden Markov models. On the other hand, the estimation procedure described in section 3 can be adapted to general hidden Markov models in a straightforward manner. This shall be shown in the present section.

A hidden Markov model is a two–dimensional Markov chain $Z = (X, Y) = ((X_n, Y_n) : n \in I\!N)$ with the following properties:

1. $P(X_{n+1}|X_1, \ldots, X_n) = P(X_{n+1}|X_n)$, i.e. the marginal chain X is a Markov chain with finite state space $\{1, \ldots, m\}$. It is called the underlying Markov chain or the regime of Z. Let $P^0 = (p^0_{ij})_{1 \leq i,j \leq m}$ denote the true transition matrix of the underlying Markov chain X. The space of all Markov transition matrices of dimension m shall be denoted by $\mathcal{M}_1$.

2. $P(Y_n|X_1, Y_1, \ldots, X_{n-1}, Y_{n-1}, X_n) = P(Y_n|X_n)$, i.e. given the regime X, the distributions of Y_n, $n \in I\!N$, are independent. Assume that the distributions $P(Y_n|X_n = i)$, $i = 1, \ldots, m$ are all dominated by some measure μ and denote the μ–density of $P(Y_n|X_n = i)$ by $f(.; \theta_i)$, with $\theta_1, \ldots, \theta_m \in \Theta$, calling Θ^m the parameter space for the marginal chain Y.

According to the construction in Çinlar (1969), p.384f, and given this definition of a hidden Markov model, the sample of a MAP would be represented by the observable inter–arrival times $y_1, \ldots, y_N$ and the hidden variables $X_1, \ldots, X_N$ with X_n denoting the phases immediately after the adjacent arrival instants. Thus $X_n = (i, j)$ means that immediately after arrival instant T_{n-1}, the phase process was in state i, and immediately after arrival instant T_n, the phase process was in state j. Then we have

$$P(X_{n+1} = (l, k)|X_n = (i, j)) = 0 \qquad \text{if} \quad l \neq j$$

and

$$P(X_{n+1} = (j,k)|X_n = (i,j)) = \int_0^\infty e_j^T \exp(D_0 t) D_1 e_k \, dt$$

as well as

$$P(Y_n \in dt | X_n = (i,j)) = e_i^T \exp(D_0 t) D_1 e_j$$

Clearly, this representation is rather theoretical, as the resulting parametrization of the MAP could hardly be called a natural one.

Suppose a time series y is observed as a sample of the marginal chain Y of a hidden Markov model Z. The marginal chain X of Z is not observable. Then y is called an incomplete data sample of Z. A complete data sample would be a set $(x_1, y_1), \ldots, (x_N, y_N)$ with x_n and y_n denoting the state of the regime and the observed value of the marginal chain Y at time index n, respectively. Let $y = (y_1, \ldots, y_N)$ denote an incomplete data sample of size N to be fitted to a hidden Markov model. The statistical problem lies in finding the true parameters out of the parameter space $\mathcal{M}_1 \times \Theta^m$ for Z.

A likelihood oriented approach may be gained from an adaptation of the procedure derived in section 3. First, the data $y_1, \ldots, y_N$ are seen as a sample from a finite mixture distribution. They are used to gain information on the mixing distribution and the parameters of the component distributions in the mixture. Second, discriminant analysis yields an estimate of the current mixture (or the current state of the regime) at any time index $1, \ldots, N-1$. Third, using these estimates the transition matrix of the underlying Markov chain is estimated by an adaptation of the classical empirical estimator by Anderson, Goodman (1957). These steps are explicated in the rest of this section. All are similar in reasoning and methods to the respective steps in section 3. Therefore, their descriptions are kept short.

5.1. The sample as a finite mixture

Not knowing the regime X, the data sample must be seen as a sample of some finite mixture distribution with density function

$$f(x) = \sum_{i=1}^m \pi_i f(x; \theta_i)$$

for all x in the sample space $\mathcal{X}$. In this first step we assume that there is no correlation between $y_1, \ldots, y_N$. Then we can estimate the mixing distribution π as well as the parameters θ_i of the component distributions via an appropriate specification of the EM algorithm (see McLachlan, Peel (2000), ch.2). Denote the derived estimates by $\hat{\pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_m)$ and $\hat{\theta}_i$.

5.2. Estimating the regime sample

In order to discover the underlying correlation between $y_1, \dots, y_N$, we need to estimate the state of the regime at every time index $1, \dots, N$. This can be done by discriminant analysis (cf. Titterington et al. (1985), pp.168f), yielding

$$\hat{\pi}_n(i) = \frac{\hat{\pi}_i f(y_n; \hat{\theta}_i)}{\sum_{i=1}^m \hat{\pi}_i f(y_n; \hat{\theta}_i)}$$

for all $n = 1, \dots, N$ and $i = 1, \dots, m$. Here, $\hat{\pi}_n(i)$ denotes the conditional expectation that the regime was in state i at time index n, given the observation y_n and the estimates $\hat{\pi}$ and $\hat{\theta}_i$ from section 5.1.

5.3. Estimating the regime transition matrix

A standard estimator for Markov chains without any additional structure is given in Anderson, Goodman (1957). Analogous to step 3 in the procedure of section 3, the regime states cannot be observed directly but must be estimated by the vectors $\hat{\pi}_n$ computed in the preceding section 5.2. Plugging these into the classical formulas yields

$$\hat{N}_{ij} = \sum_{n=1}^{N-1} \hat{\pi}_n(i)\hat{\pi}_{n+1}(j) \qquad \text{and} \qquad \hat{N}_i = \sum_{n=1}^{N-1} \hat{\pi}_n(i) = \sum_{j=1}^{m} \hat{N}_{ij}$$

for $i, j = 1, \dots, m$. Now the empirical estimators under the condition of the estimates from section 5.2 are given as

$$\hat{p}_{ij} = \frac{\hat{N}_{ij}}{\hat{N}_i}$$

for all $i, j = 1, \dots, m$. This completes the estimation procedure.

Acknowledgements

The author wishes to thank the referees for their insightful comments which helped to considerably improve the first version of this paper.

References

Albert, A.: 1962, 'Estimating the infinitesimal generator of a continuous time, finite state Markov process'. *Ann. Math. Stat.* **33**, 727–753.

Anderson, T. and L. A. Goodman: 1957, 'Statistical inference about Markov chains.'. *Ann. Math. Stat.* **28**, 89–110.

Asmussen, S.: 1997, 'Phase-type distributions and related point processes: Fitting and recent advances.'. In Chakravarthy, Srinivas R. (ed.) et al., Matrix-analytic methods in stochastic models. Proceedings of the first international conference held in Detroit, MI, USA, August 24–25, 1995. New York, NY: Marcel Dekker. Lect. Notes Pure Appl. Math. 183, pp.137–149 .

Asmussen, S. and G. Koole: 1993, 'Marked point processes as limits of Markovian arrival streams.'. *J. Appl. Probab.* **30**(2), 365–372.

Asmussen, S., O. Nerman, and M. Olsson: 1996, 'Fitting phase-type distributions via the EM algorithm.'. *Scand. J. Stat.* **23**(4), 419–441.

Basawa, I. V. and B. Prakasa Rao: 1980, *Statistical inference for stochastic processes.* London etc.: Academic Press.

Baum, D.: 1996, 'Ein Faltungskalkül für Matrizenfolgen und verallgemeinerte Poisson–Gruppenprozesse'. Research Report No.96–36, Dept. of Mathematics and Computer Science, University of Trier, Germany.

Botta, R. F., C. M. Harris, and W. G. Marchal: 1987, 'Characterizations of generalized hyperexponential distribution functions.'. *Commun. Stat., Stochastic Models* **3**, 115–148.

Breuer, L.: 2000a, 'Parameter Estimation for a Class of BMAPs'. In: G. Latouche and P. Taylor (eds.): *Advances in Algorithmic Methods for Stochastic Models - Proceedings of the 3rd International Conference on Matrix Analytical Methods*. New Jersey, USA, pp. 87–97, Notable Publications.

Breuer, L.: 2000b, 'Spatial Queues'. Ph.D. thesis, University of Trier, Germany.

Breuer, L. and M. Gilbert: 2000, 'Parameter Estimation for BMAPs'. Research Report No.00–09, Department of Mathematics and Computer Science, University of Trier, Germany.

Cinlar, E.: 1969, 'On semi-Markov processes on arbitrary spaces'. *Proc. Camb. Philos. Soc.* **66**, 381–392.

Dempster, A., N. Laird, and D. Rubin: 1977, 'Maximum likelihood from incomplete data via the EM algorithm. Discussion.'. *J. R. Stat. Soc., Ser. B* **39**, 1–38.

Elliott, R. J., L. Aggoun, and J. B. Moore: 1995, *Hidden Markov models. Estimation and control.* Berlin: Springer-Verlag.

Gilbert, M.: 2000, 'Zur Parameterschätzung für eine Klasse von Markov–additiven Prozessen'. Diploma Thesis, University of Trier.

Jewell, N. P.: 1982, 'Mixtures of exponential distributions.'. *Ann. Stat.* **10**, 479–484.

Latouche, G. and V. Ramaswami: 1999, *Introduction to matrix analytic methods in stochastic modeling.* Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics.

Lucantoni, D. M.: 1991, 'New results on the single server queue with a batch Markovian arrival process.'. *Commun. Stat., Stochastic Models* **7**(1), 1–46.

Lucantoni, D. M.: 1993, 'The BMAP/G/1 Queue: A Tutorial'. In: L. Donatiello and R. Nelson (ed.): *Models and Techniques for Performance Evaluation of Computer and Communication Systems*. Springer, pp. 330–358.

McLachlan, G. and D. Peel: 2000, *Finite mixture models.* Chichester: Wiley.

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

Meng, X.-L. and D. van Dyk: 1997, 'The EM algorithm - an old folk-song sung to a fast new tune.'. *J. R. Stat. Soc., Ser. B* **59**(3), 511–567.

Neuts, M. F.: 1979, 'A versatile Markovian point process.'. *J. Appl. Probab.* **16**, 764–774.

Neuts, M. F.: 1986, 'The caudal characteristic curve of queues.'. *Adv. Appl. Probab.* **18**, 221–254.

Ryden, T.: 1993, *Parameter estimation for Markov modulated Poisson processes and overload control of SPC switches. Statistics and telecommunications.* Lund Inst. of Technology, Dept. of Math. Statistics.

Ryden, T.: 1994, 'Parameter estimation for Markov modulated Poisson processes.'. *Commun. Stat., Stochastic Models* **10**(4), 795–829.

Ryden, T.: 1996, 'An EM algorithm for estimation in Markov-modulated Poisson processes.'. *Comput. Stat. Data Anal.* **21**(4), 431–447.

Ryden, T.: 1997, 'Estimating the order of continuous phase-type distributions and Markov-modulated Poisson processes.'. *Commun. Stat., Stochastic Models* **13**(3), 417–433.

Titterington, D., A. Smith, and U. Makov: 1985, *Statistical analysis of finite mixture distributions.* Chichester etc.: John Wiley & Sons.