State Estimation for Partially Observed Markov Chains

MATS RUDEMO

Research Institute of National Defense, Stockholm, Sweden
Submitted by K. J. Åström

Recursive equations are derived for the conditional distribution of the state of a Markov chain, given observations of a function of the state. Mainly continuous time chains are considered. The equations for the conditional distribution are given in matrix form and in differential equation form. The conditional distribution itself forms a Markov process. Special cases considered are doubly stochastic Poisson processes with a Markovian intensity, Markov chains with a random time, and Markovian approximations of semi-Markov processes. Further the results are used to compute the Radon–Nikodym derivative for two probability measures for a Markov chain, when a function of the state is observed.

1. Introduction

The theory of conditional Markov processes, where recursive equations are derived for the conditional distribution of the state of an incompletely observed Markov process, was initiated by Stratonovich [1], see also [2]. Using a terminology common in control theory we shall refer to the computation of the conditional distribution of the state as state estimation. The main cases studied in [1] are discrete time Markov chains and continuous time processes of the diffusion type. State estimation for diffusion processes have also been studied by Kushner [3]. Discrete time chains including controlled chains were treated by Åström [4]. Continuous time Markov chains affecting an observed diffusion process were studied by Wonham [5] and Shiryaev [6]. This short list of works on state estimation is far from complete and only points out some of the earlier contributions. Furthermore papers treating Gaussian processes, that is Kalman–Bucy filtering, have not been mentioned.

The case with continuous time and discrete state space for the observations has not received much attention in the literature, at least until recently. In [7] Yashin discusses state estimation for a continuous time Markov chain with two components, of which the first forms a Markov chain itself. The second component is observed. Let \( \{\xi_t\} \) and \( \{\xi^1_t, \xi^2_t\} \) be Markov chains with finite state spaces \( S_1 \) and \( S_1 \times S_2 \), and with right-continuous sample functions.
The transition intensities (see Sections 3 and 7 below for definitions) of \{\xi_t\} are \( q_{ki}, k, i \in S_1 \), and given that \( \xi_t = k \) the transition intensities of \{\xi_t\} are \( \lambda_{ab}(k), a, b \in S_2 \). Define

\[
\hat{\lambda}_{k}(t) = \lambda_{k}(t) = \frac{q_{sk} - k}{s > 0, (1.1)}
\]

\[
K \in S_1. \quad \text{Then the equation in Note 2 in [7, p. 730] may be written}
\]

\[
\dot{p}_k(t) = p_k(0) + \int_0^t \sum_i \dot{p}_i(u) q_{ik} \, du
\]

\[
+ \int_0^t \dot{p}_k(u) [\dot{\lambda}_{k}(u) - \lambda_{ab}(k)] \, du
\]

\[
+ \sum_{0 < r < t} \dot{p}_k(r - 0) \left[ \frac{\lambda_{ab}(k)}{\dot{\lambda}_{ab}(r - 0)} - 1 \right],
\]

where the sum is to be extended over the jump epochs \( \tau \) of \{\xi_u; 0 < u \leq t\}, \( \alpha = \xi_{\tau - 0}, \beta = \xi_{\tau} \),

\[
\hat{\lambda}_{ab}(t) = \sum_k \hat{p}_k(t) \lambda_{ab}(k), (1.3)
\]

and

\[
\hat{\lambda}_a(t) = \sum_k \hat{p}_k(t) \lambda_{a}(k) = \sum_k \hat{p}_k(t) \sum_{b \in a} \lambda_{ab}(k). (1.4)
\]

The proof in [7] is given for the case when \( S_1 \) and \( S_2 \) consist of two states each, and is based on the theories of stochastic differential equations and derivatives of measures on function spaces. Representations with stochastic differential equations are also used by Galchuk [8] for more general Markov jump processes. Detailed results are given in a special case of a doubly stochastic Poisson process by Galchuk and Rozovskii [9].

State estimation for doubly stochastic Poisson processes, with an intensity which is a vector Markov process, is discussed by Snyder [10]. A basic result in [10] is a differential equation for the characteristic function of the conditional distribution of the state of the Markov process, given observations of the doubly stochastic Poisson process. In [11] a more general observation process is considered. A related approach to conditional distribution computation is also given in an example on prediction of point processes in Jowett and Vere-Jones [12, Section 5].

If \{\xi_t\} in the model above is a doubly stochastic Poisson process with intensity \( \lambda_k \) if \( \xi_t = k \), it follows from Eq. (1.2) that, in intervals between events, \( \hat{p}_k(t), k \in S_1 \), satisfy the system of ordinary differential equations

\[
\dot{\hat{p}}_k(t) = \sum_i \hat{p}_i(t) q_{ik} + [\hat{\lambda}(t) - \lambda_k] \hat{p}_k(t), (1.5)
\]
where

\[ \dot{\lambda}(t) = \sum_k \dot{\tau}_k(t) \lambda_k. \]  

(1.6)

At events

\[ \dot{\tau}_k(t) = \dot{\tau}_k(t - 0) \lambda_k \lambda(t - 0). \]  

(1.7)

A short derivation of (1.5)–(1.7) is given in [13]. To conform with the notation used later in this paper the Markov chains and \( \dot{\tau}_k(t) \) have been defined to make \( \dot{\tau}_k(t) \) right-continuous instead of left-continuous as in [13]. There it is also shown that the solution of (1.5) may be obtained from an associated linear system of differential equations. Specifically if no events of the point process occur in \((s, t]\) we have

\[ \dot{\tau}_k(t) = \frac{\tau_k^*(t)}{\sum_i \tau_i^*(t)}, \]  

(1.8)

where \( \tau_k^*(t), k \in S_1 \), satisfy the linear system of differential equations

\[ \frac{d\tau_k^*}{dt} = \sum_i \tau_i^*(t) q_{ik} - \lambda_k \tau_k^*(t), \]  

(1.9)

with the initial value

\[ \tau_k^*(s) = \dot{\tau}_k(s). \]

Loosely speaking the relation between the Eqs. (1.5) and (1.9) may be interpreted as follows. To give the correct conditional probabilities the solution of (1.9) must be normalized, see (1.8). The Eqs. (1.5) may be viewed as equations obtained from (1.9) by instant normalization.

In this paper the problem of computing the conditional probabilities of the states of a Markov chain is considered for the following model. The Markov chain \( \eta_t \) is allowed to have countably many states but no discontinuities other than jumps, and the observed process is a function \( \xi_t = g(\eta_t) \) of the Markov chain. We shall show that the row vector \( \dot{\tau}(t) \) with components

\[ \dot{\tau}_k(t) = P(\eta_t = k \mid \xi_s, s \leq t) \]

may be computed recursively by post-multiplication of matrix operators of two kinds. One kind corresponds to intervals between jumps and the other to jumps. The former operators satisfy Kolmogorov's forward and backward differential equations. The forward equations give the generalization of (1.9) and from these equations we derive the generalization of (1.5).

In Section 2 Markov chains with discrete time are considered. The results obtained resemble the results obtained later for continuous time. They are
however simpler both in proof and formulation. For instance only one kind
of matrix operators is involved in the recursive formula for $\hat{f}(t)$.

The continuous time model is described in Section 3. The transition
intensities are supposed to be constant in time. In order to include such
examples as birth and death processes with linear birth and death rates, they
are not required to be bounded. The chains are, however, supposed to have
sample functions which are step functions with a finite number of jumps in
every finite time interval. Relevant references on the theory of Markov
chains are [14] and [15], and for a more elementary treatment [16, Chap. 17].
Several of the proofs below would be simpler if the transition intensities were
supposed to be bounded. Further, most of the results can be modified to hold
for time-dependent transition intensities $\{q_{id}(t)\}$ if these are bounded and
satisfy suitable continuity conditions ensuring that Kolmogorov's forward
and backward equations remain valid, see for instance [17, p. 316].

The recursive matrix formula, with two kinds of matrix operators, is
derived in Section 4. This formula forms the starting point in Section 6
for the derivation of differential equations for the conditional state prob-
abilities. Further a result given in an appendix, on Kolmogorov's forward
equation for the absolute state probabilities, is used. In the appendix it is
shown that the integrated version of this equation always holds true under
the conditions of this paper. (For bounded transition intensities the equation
is valid in the usual sense.) Hence the general equations corresponding to
(1.5) and (1.9) are given in integrated form.

In Section 5 it is shown that the state estimates themselves, form a Markov
process $\{\hat{f}(t)\}$. If the chain $\{\eta\}$ has infinitely many states, the state space of
$\{\hat{f}(t)\}$ is infinite-dimensional.

The model with $\eta_t = (\xi_t, \xi_t)$, where the second component is observed, is
treated in Section 7. In particular a generalization of (1.2) is obtained for the
case with countably infinite state space. Further the transition intensities
of the first component may depend on the values of the second, and hence we
do not require that $\{\xi\}$ is a Markov chain. We note that if we allow $\xi_t$ and $\xi_t$
to jump simultaneously, then the models with $\xi_t = g(\eta_t)$ and $\eta_t = (\xi_t, \xi_t)$
are equivalent. A discussion of the equivalence of Markovian models, where
the observations are either deterministic functions of a system with large
state space, or stochastic functions of a system with smaller state space, is
given for discrete time in [18].

Two examples in Section 7 concern the doubly stochastic Poisson process
of [13] and Markov chains with a random time. The last model is a special
case of a model of semi-Markov processes with a random time, see [19].

In Section 8 the problem of approximating a semi-Markov process with a
partially observed Markov chain is discussed.

Statistical inference for continuous time Markov chains is discussed by
Albert [20] and for a more general model with Markov processes of jump type by Billingsley [21]. Let us consider a right-continuous Markov chain \( \{ \eta_t \} \) with transition intensities \( q_j(i) \) and \( q_j(i, k) \), \( i, k \in S, j = 0, 1 \), corresponding to two probability measures \( P_0 \) and \( P_1 \). Let \( N_t \) be the number of jumps in \( (0, t] \) and let \( \tau_1 < \tau_2 < \cdots \) be the successive jump epochs. Suppose that \( P_k \) is absolutely continuous with respect to \( P_0 \) when the measures are considered on the \( \sigma \)-algebra generated by \( \{ \eta_s; 0 \leq s \leq t \} \). Then the Radon–Nikodym derivative of \( P_k \) with respect to \( P_0 \) is given by

\[
\frac{dP_k}{dP_0} = \frac{p_k(\xi_0)}{p_0(\xi_0)} \prod_{k=1}^{N_t} \frac{q_k(\xi_{k-1}, \xi_k)}{q_0(\xi_{k-1}, \xi_k)} \cdot \exp \left\{- \int_0^t \left[ q_k(\eta_s) - q_0(\eta_s) \right] ds \right\},
\]

(1.10)

where \( p_j(i) = p_j(\eta_0 = i), j = 0, 1, \alpha_0 = \eta_0 \) and \( \alpha_k = \eta_k, k \geq 1 \). The formula may be obtained from Theorem 3.2 in [20, p. 732] or by specialization to a Markov chain in (7.7) in [21, p. 381]. In Section 9 of the present paper a formula similar to (1.10) is derived for a partially observed Markov chain.

2. Discrete Time Markov Chains

Let \( \eta_t, t = 0, 1, \ldots, \) be a discrete time Markov chain with discrete (finite or countably infinite) state space \( S \) and stationary transition probabilities

\[ P_{ti} = P(\eta_{t+1} = j \mid \eta_t = i). \]

Suppose we observe

\[ \xi_t = g(\eta_t), \]

(2.1)

which in general is a non-Markovian process. The state space of \( \xi_t \) is

\[ B = g(S), \]

(2.2)

and if \( \xi_t = a \) we know that \( \eta_t \) belongs to the set

\[ S_a = g^{-1}(\{a\}) = \{i \in S : g(i) = a \}. \]

(2.3)

Given the observations of \( \xi_0, \xi_1, \ldots, \xi_t \) the conditional probability of \( \eta_t = k \) is

\[ P_k(t) = P(\eta_t = k \mid \xi_s, s \leq t). \]

(2.4)

A recursive formula for \( P_k(t + 1) \) may be obtained from Bayes formula, see [1] or [4]. Suppose \( \xi_s = a_s, s \leq t + 1 \). Then

\[ P_k(t + 1) = \frac{P(\eta_{t+1} = k \mid \xi_s = a_s, s \leq t)}{P(\xi_{t+1} = a_{t+1} \mid \xi_s = a_s, s \leq t)}, \quad k \in S_{a_{t+1}}, \]

(2.5)
and $\dot{P_k}(t + 1) = 0$ else. Further
\[ P(\eta_{t+1} = k \mid \xi_s = a_s, s \leq t) = \sum_i \dot{p}_i(t) P_{ik}, \tag{2.6} \]
and the denominator in (2.5) may be obtained by summation of the right members of (2.6) for all $k$ such that $g(k) = a_{t+1}$.

To simplify the formula obtained for $\dot{p}_k(t + 1)$, let $P(a, b)$, where $a, b \in B$, be the substochastic matrix on $S \times S$ with elements
\[ P_{ik}(a, b) = P_{ik}, \quad i \in S_a, \quad k \in S_b, \tag{2.7} \]
and $P_{ik}(a, b) = 0$ if $(i, k) \not\in S_a \times S_b$. Further if $p$ is a row vector with elements $p_k, k \in S$, such that
\[ 0 < \sum_k p_k < \infty, \tag{2.8} \]
we let $pN$ be the row vector with elements
\[ p_k / \sum_{i \in S} p_i, \quad k \in S. \tag{2.9} \]
The nonlinear operator $N$ will be referred to as the normalizer. We then have the following lemma, the proof of which is given by a straightforward computation.

**Lemma 1.** Let $P$ be a matrix with nonnegative elements and $p$ a row vector such that $p$ and $pP$ satisfy (2.8). Then
\[ (pP)N = ((pN)P)N. \]

The lemma may be interpreted in the following way. If we normalize a vector matrix product, we may also normalize the vector before matrix multiplication if we wish. Using the convention that all operations should proceed from the left to the right, we may write the assertion of Lemma 1 as
\[ pPN = pNPN. \]
Using Lemma 1 we obtain from (2.5)–(2.7) the following theorem.

**Theorem 1.** The row vector $\dot{p}(t)$ of conditional state probabilities satisfies the recursive equation
\[ \dot{p}(t + 1) = \dot{p}(t) P(\xi_t, \xi_{t+1}) N. \]
If \( p(0) \) is the row vector of initial probabilities, \( p_k(0) = P(\eta_0 = k) \), then

\[
\hat{p}(t) = p(0) P(\xi_0, \xi_1) P(\xi_1, \xi_2) \cdots P(\xi_{t-1}, \xi_t) N.
\]

3. THE CONTINUOUS TIME MODEL

Let \( \{\eta_t; t \in [0, \infty)\} \) be a continuous time Markov chain with discrete state space \( S \) and stationary transition probabilities

\[
P_{ij}(t) = P(\eta_{t+} = j \mid \eta_s = i), \quad t \geq 0,
\]

satisfying

\[
\lim_{t \to 0} P_{ij}(t) = \delta_{ij}.
\]

Then the limits

\[
 q_i = -q_{ii} = \lim_{t \to 0} [1 - P_{ii}(t)]t
\]

and

\[
 q_{ij} = \lim_{t \to 0} P_{ij}(t)/t, \quad j \neq i,
\]

exist, see [14] or [15]. We shall assume that the \( q_i \)'s are finite and that the chain is conservative in the sense that

\[
\sum_{j \in S} q_{ij} = 0 \quad (3.1)
\]

for all \( i \in S \). We shall further assume that \( \{\eta_t\} \) has right-continuous paths, that it is the minimal chain corresponding to \( \{q_{ij}\} \), see [14, p. 251], and that if \( \chi_n \) denotes the state after the \( n \)-th jump, then

\[
P \left( \sum_{n=1}^{\infty} q_{\chi_n}^{-1} = \infty \right) = 1.
\]

This implies, see [14, p. 260], that the sample functions a.s. are step functions with a finite number of jumps in every finite time interval. Further the backward and forward Kolmogorov equations

\[
P'_{ik}(t) = \sum_j q_{ij} P_{jk}(t) \quad (3.2)
\]

and

\[
P'_{ik}(t) = \sum_j P_{ij}(t) q_{jk} \quad (3.3)
\]

are satisfied, see [14, Theorem 5 on p. 249].
Let us suppose that we observe a function

$$\xi_t = g(\eta_t)$$

for $$t \geq 0$$, and define $$B$$ and $$S_a$$, $$a \in B$$, by (2.2) and (2.3). Put

$$P_{ij}(a, t) = P(\eta_{s+t} = j, \eta_u \in S_a, s \leq u \leq s + t | \eta_s = i), \quad (3.4)$$

for $$t \geq 0$$ and $$i, j \in S$$. Obviously $$P_{ij}(a, t) = 0$$, if $$i$$ or $$j$$ does not belong to $$S_a$$. Let $$P(a, t)$$ be the matrix with elements $$P_{ij}(a, t)$$. Sometimes we shall use the same notation for the submatrix of $$P(a, t)$$ with element indices in $$S_a \times S_a$$. Like the matrix $$P(t)$$ with elements $$P_{ij}(t)$$, the matrix $$P(a, t)$$ satisfies the Chapman–Kolmogorov equation

$$P(a, t + s) = P(a, t)P(a, s) \quad (3.5)$$

and further, when regarded on $$S_a \times S_a$$

$$\lim_{t \to 0} P(a, t) = I,$$

where $$I$$ is the unit matrix and the convergence considered for matrices in this paper is element-wise. The matrix $$P(t)$$ is stochastic, that is the row sums are equal to one, but the matrix $$P(a, t)$$ is substochastic,

$$\sum_j P_{ij}(a, t) \leq 1,$$

and in general we have strict inequality for all $$t > 0$$. Further the backward and forward Kolmogorov equations

$$P'_{ik}(a, t) = \sum_{j \in S_a} q_{ij}P_{jk}(a, t), \quad i \in S_a, \quad (3.6)$$

and

$$P'_{ik}(a, t) = \sum_{j \in S_a} P_{ij}(a, t)q_{jk}, \quad k \in S_a, \quad (3.7)$$

are satisfied. Equations (3.6) and (3.7) follow from the fact that $$P_{ij}(a, t)$$ are taboo transition probabilities, see [14, p. 187], with the taboo set $$S \setminus S_a$$. Alternatively (3.6) and (3.7) may be derived by consideration of the Markov chain derived from $$\{\eta_t\}$$ by transformation of $$S \setminus S_a$$ into an absorbing state.

For $$a \neq b$$ let $$Q(a, b)$$ be the matrix with elements

$$Q_{ik}(a, b) = q_{ik}, \quad i \in S_a, \quad k \in S_b, \quad (3.8)$$

and $$Q_{ik}(a, b) = 0$$ if $$(i, k) \notin S_a \times S_b$$. 
4. The Conditional State Distribution

Let us consider the partially observed Markov chain \( \{ \eta_t; 0 \leq t < \infty \} \) described in the previous section. We recall that \( \{ \eta_t \} \) and hence also \( \{ \xi_t \} \) is supposed to have sample functions which are right-continuous step functions.

Let \( N_t \) be the number of jumps of \( \{ \xi_s; 0 < s \leq t \} \) and let \( \hat{\eta}(t) \) be the row vector with components

\[
\hat{\eta}_k(t) = P(\eta_t = k \mid \xi_s, 0 \leq s \leq t). \tag{4.1}
\]

We note that \( \hat{\eta}(t) \) is uniquely defined up to changes on sets of probability zero. Hence relations involving \( \hat{\eta}(t) \) may only be expected to hold almost surely. We shall usually omit that qualifier.

**Theorem 2.** Let \( \{ \xi_t \} \) have jumps at \( \tau_1 < \tau_2 < \cdots \). Put \( \alpha_0 = \xi_0 \) and \( \alpha_\nu = \xi_{\tau_\nu}, \nu \geq 1 \). Then

\[
\hat{\eta}(t) = \hat{\eta}(0) P(\alpha_0, \tau_1) Q(\alpha_0, \alpha_1) P(\alpha_1, \tau_2 - \tau_1) Q(\alpha_1, \alpha_2) \cdots Q(\alpha_{N_t-1}, \alpha_{N_t})
\times P(\alpha_{N_t}, t - \tau_{N_t}) N, \tag{4.2}
\]

where \( \hat{\eta}(0) \) is the distribution of \( \eta_0 \), \( P(\alpha, t) \) is given by (3.4), \( Q(\alpha, b) \) is given by (3.8), and \( N \) is the normalizer, see (2.9).

**Remark 1.** From (4.2) recursive formulas for \( \hat{\eta}(t) \) may be obtained. Suppose that \( \xi_t \) has no jumps for \( s < u \leq t \). Then it follows from (4.2), (3.5), and Lemma 1 that

\[
\hat{\eta}(t) = \hat{\eta}(s) P(\alpha, t - s) N, \tag{4.3}
\]

where \( \alpha = \xi_s \). Further we may then compute \( \hat{\eta}(t) \) from \( \hat{\eta}(s) \) by the solution of systems of differential equations, see Section 6 below. At a jump epoch \( \tau \) we have

\[
\hat{\eta}(\tau) = \hat{\eta}(\tau - 0) Q(\alpha, \beta) N \tag{4.4}
\]

where \( \alpha = \xi_{\tau-0} \) and \( \beta = \xi_\tau \). Some care has to be exercised in the interpretation of (4.4). The problem is similar to the problem of a rigorous interpretation of Palm probabilities in the theory of point process, see [23, p. 353]. Note however that if we want to compute \( \hat{\eta}(t) \), we can use (4.4) at the jump epochs and (4.3) in the intervals between. Compare also Remark 4 below.

The proof of Theorem 2 will be based on two Lemmas which also have some independent interest. We note that \( \tau_n, n \geq 1 \), are stopping times with respect to \( \{ \xi_t \} \), that is

\[
\{ \tau_n \leq t \} \in \mathcal{F}_t
\]
where $\mathcal{F}_t$ is the $\sigma$-algebra generated by $\{\xi_s: 0 \leq s \leq t\}$, compare [22, p. 268]. The following lemma specifies the distribution of the stopping time $\tau_1$.

**Lemma 2.** Suppose that $i, j \in S_a$, $k \not\in S_a$ and that $0 \leq u < v$. Then

$$P(\eta_0 = i, \eta_{\tau_1} = j, \eta_{\tau_1} = k, u < \tau_1 \leq v) = p_i(0) \int_u^v P(c_s(a, s) q_{jk} ds. \quad (4.5)$$

**Proof.** Let $A$ be the event in the left member of (4.5). Put

$$B(s, h) = \{\eta_0 = i, N_s = 0, \eta_s = j, \eta_{s+h} = k\}$$

and with $h_r = (v - u)/r$

$$B(r) = \bigcup_{m=0}^{r-1} B(u + mh_r, h_r).$$

We note that the sets in the union are disjoint. As $B(s, h)$ implies that $s < \tau_1 \leq s + h$ we find that $B(r)$ tends to $A$ as $r \to \infty$. Hence

$$P(A) = \lim_{r \to \infty} \sum_{m=0}^{r-1} P(B(u + mh_r, h_r))$$

$$= \lim_{r \to \infty} \sum_{m=0}^{r-1} p_i(0) P(e_s(a, u + mh_r) P(j)(h_r)$$

$$= p_i(0) \int_u^v P(\omega(a, s) q_{jk} ds. \quad \blacksquare$$

**Remark 2.** Summation of the members in (4.5) gives

$$P(u < \tau_1 \leq v) = \sum_{a \in B} \sum_{i \in S_a} \sum_{j \in S_a} \sum_{k \in S_a} p_i(0) \int_u^v P(e_s(a, s) q_{jk} ds.$$

Using Kolmogorov's forward Eq. (3.7) we may reduce this to

$$P(u < \tau_1 \leq v) = \sum_{a \in B} \sum_{k \in S_a} p_i(0) \left[ \sum_{k \in S_a} P(e_k(a, u) - \sum_{k \in S_a} P(e_k(a, v)) \right],$$

a formula which may be deduced in a more direct way.

Let us introduce the matrices

$$P(n, a_0 , \ldots , a_n , s_1 , \ldots , s_n , t)$$

$$= P(a_0 , s_1) Q(a_0 , a_1) P(a_1 , s_2 - s_1) \cdots Q(a_{n-1} , a_n) P(a_n , t - s_n), \quad (4.6)$$
\( n \geq 0, a_i \in B, i = 0, \ldots, n, 0 < s_1 < \cdots < s_n < t, \) with \( P(0, a, t) = P(a, t), \) and the vectors

\[
p(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t) = p(0) P(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t), \tag{4.7}
\]

\( n \geq 0. \) In the following lemma the joint distribution of \( N_t, \xi_0, \tau_v \) and \( \xi_{\tau_v}, \nu = 1, \ldots, N_t, \) and \( \eta_t \) is specified.

**Lemma 3.** Suppose that

\[
0 \leq u_1 < v_1 \leq u_2 < \cdots < u_n < v_n \leq t.
\tag{4.8}
\]

Then

\[
P(N_t = n, \xi_0 = a_0, \xi_{\tau_v} = a_v, u_v < \tau_v \leq v_v, \nu = 1, \ldots, n, \eta_t = k)
= \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} p_{ik}(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t) \, ds_1 \cdots ds_n. \tag{4.9}
\]

**Proof.** By induction we shall prove that

\[
P(N_t = n, \eta_0 = i, \xi_0 = a_0, \xi_{\tau_v} = a_v, u_v < \tau_v \leq v_v, \nu = 1, \ldots, n, \eta_t = k)
= \rho_i(0) \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} p_{ik}(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t) \, ds_1 \cdots ds_n. \tag{4.10}
\]

For \( n = 0 \) we have to show that for \( g(i) = a \)

\[
P(N_t = 0, \eta_0 = i, \eta_t = k) = \rho_i(0) P_{ik}(a, t),
\]

but this relation follows directly from the definition of \( P(a, t). \) Suppose now that (4.10) is valid if \( n \) is replaced by \( n - 1. \) Let \( A \) denote the event in the left member of (4.10). Using Lemma 2 and the strong Markov property, see [22, p. 323 and p. 328], we get

\[
P(A) = \sum_{j \in S_{\eta_0}} \sum_{l \in S_{\eta_{t-1}}} \int_{u_1}^{v_1} f(s) \rho_i(0) P_{ij}(a_0, s) q_{jl} \, ds,
\]

where

\[
f(s) = P(A \mid \tau_1 = s, \eta_0 = i, \eta_{t-0} = j, \eta_s = l)
= P(N_{t-s} = n - 1, \xi_0 = a_1, \xi_{\tau_v} = a_{v+1}, u_{v+1} - s < \tau_v \leq v_{v+1} - s, \nu = 1, \ldots, n - 1, \eta_{t-s} = k \mid \eta_0 = l).
\]
Now (4.10) follows from the induction hypothesis and the relation
\[
\sum_{j \in S_{a_0}} \sum_{k \in S_{a_1}} P_{jk}(a_0, s) q_{it} P_{ik}(n - 1, a_1, ..., a_n, s_1, ..., s_{n-1}, t)
= P_{ik}(n, a_0, ..., a_n, s, s_1 + s, ..., s_{n-1} + s, t + s),
\]
which in turn follows from definition (4.6).

**Remark 3.** If we apply (4.9) with \( g \) as the identity function, that is if \( g(\eta_t) = \eta_t \), we get the equation
\[
P(N_t = \eta_t, \eta_0 = i_0, \eta_{\tau_\nu} = i_\nu, u_\nu \leq \tau_\nu \leq v_\nu, \nu = 1, ..., n)
= \prod_{\nu=1}^{v_n} \prod_{u_n}^{v_n} P_{i_\nu} q_{i_\nu, i_{\nu-1}} q_{i_{\nu-1} i_{\nu}}
\times \exp\left(-q_{i_0} s_1 - q_{i_1} (s_2 - s_1) - \cdots - q_{i_n} (t - s_n)\right) dt \cdots ds,
\]
compare Theorem 3.1 in [20, p. 730].

**Proof of Theorem 2.** Let \( \mathcal{F}_t \) be the \( \sigma \)-algebra generated by \( \{\xi_s: 0 \leq s \leq t\} \) and let \( \mathcal{C}_t \) consist of the empty set and sets of the form
\[
\{N_t = n, \xi_0 = a_0, \xi_{\tau_\nu} = a_\nu, u_\nu < \tau_\nu \leq v_\nu, \nu = 1, ..., n\}, \tag{4.11}
\]
subject to (4.8), \( n \geq 0 \), and \( a_\nu \in B, \nu = 0, ..., n \). We note that \( \mathcal{C}_t \) generates \( \mathcal{F}_t \) and that \( \mathcal{C}_t \) is closed under finite intersections. Suppose we can show that the sure event \( \Omega \) is a countable union of disjoint events of the form (4.11). Then it follows that two finite measures on \( \mathcal{F}_t \), that coincide on \( \mathcal{C}_t \), are identical, see for instance [15, p. 335].

To express \( \Omega \) as a countable union of sets of the form (4.11) we first note that
\[
\Omega = \bigcup_{\nu=0}^{n} \{N_t = n, \xi_0 = a_0, \xi_{\tau_\nu} = a_\nu, \nu = 1, ..., n\},
\]
where the countable union is extended over \( n = 0, 1, ..., \), and \( a_\nu \in B, \nu = 0, ..., n \). Hence it is sufficient to show that the following subset of \( n \)-space,
\[
T_n(t) = \{x \in \mathbb{R}^n: 0 < x_1 < x_2 < \cdots < x_n \leq t\}
\]
can be expressed as a countable union of disjoint subsets of \( T_n(t) \) of the form
\[
\prod_{i=1}^{n} (u_i, v_i).
\]
A solution for \( n = 2 \) is easy to visualize and some reflection shows that this geometric problem has a solution for arbitrary \( n \). We shall however refrain from giving a formal construction.

For the rest of the proof, let \( \hat{p}(t) \) denote the right member of (4.2). It follows that if we can show that

\[
P(\eta_t = k, A) = \int_A \hat{p}_k(t) \, dP \tag{4.12}
\]

for \( A \in \mathcal{G}_t \), then the theorem follows.

Let \( A \) be the set in (4.11). Then the left hand side of (4.12) is given by (4.9). For \( N_t = n, \xi_0 = a_0, \xi_\tau_v = a_v, \tau_v = \tau_v, v = 1, \ldots, n \), we have

\[
\hat{p}_k(t) = \frac{p_k(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t)}{\sum_i p_i(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t)}, \tag{4.13}
\]

see (4.2), (4.6), and (4.7). Then it follows from Lemma 3 that

\[
\int_A \hat{p}_k(t) \, dP
\]

is given by the right member of (4.9). Hence (4.12) is satisfied.

**Remark 4.** Let us put \( \xi_n = \eta_0 \) and

\[
\zeta_n = n, \tau_n,
\]

that is we regard the Markov process \( \{\eta_t\} \) at the epochs of jump of \( \{\xi_t\} \). From the strong Markov property of \( \{\eta_t\} \) it follows that \( \{\zeta_t\} \) is a discrete time Markov chain. Let us derive its transition probabilities

\[
P(\zeta_n = k \mid \zeta_{n-1} = i) = P(\eta_{\tau_n} = k \mid \eta_0 = i).
\]

Let \( i \in S_a \) and \( k \in S_b \). From (4.5) we get

\[
P(\tau_1 \leq t, \eta_{\tau_1} = k \mid \eta_0 = i) = \int_0^t \sum_{j \in S_a} P_{ij}(a, s) \, q_{jk} \, ds.
\]

Letting \( t \to \infty \) we get

\[
P(\zeta_n = k \mid \zeta_{n-1} = i) = \int_0^\infty \sum_{j \in S_a} P_{ij}(a, s) \, q_{jk} \, ds, \tag{4.14}
\]

with \( a = g(i) \neq g(k) \).

Further \( \{(\zeta_n, \tau_n) ; n = 0, 1, \ldots\} \) is a discrete time Markov process. By using properties of partially observed Markov processes with discrete time it is possible to show, that the row vector \( p^\alpha(n) \) with components

\[
p_k^\alpha(n) = P(\eta_{\tau_n} = k \mid \xi_0, \tau_v, \xi_\tau_v, v = 1, \ldots, n)
\]

for \( A \in \mathcal{G}_t \), then the theorem follows.
is given by the equation
\[
p^d(n) = p(0) P(\alpha_0, \tau_1) Q(\alpha_0, \alpha_1) P(\alpha_1, \tau_2 - \tau_1) Q(\alpha_1, \alpha_2) \ldots \\
\times P(\alpha_{n-1}, \tau_n - \tau_{n-1}) Q(\alpha_{n-1}, \alpha_n) N,
\]
with \(\alpha_0 = \xi_0\) and \(\alpha_\nu = \xi_\nu\), \(\nu \geq 1\), compare (4.2).

5. The Markov Process Formed by the State Estimates

Let \(\hat{\theta}(t)\) be the row vector of state estimates, discussed in the previous section. As the following theorem shows \(\{\hat{\theta}(t)\}\) is a Markov process. A similar result for discrete time chains is given in [4, Lemma 1, p. 187]. Further in [2, p. 133–142] results of this kind are given for general partially observed Markov processes.

**Theorem 3.** The state estimates form a Markov process \(\{\hat{\theta}(t): t \in [0, \infty)\}\) with \([0, 1]^S\) as state space.

**Proof.** Let \(A \subset [0, 1]^S\) be a set in the \(\sigma\)-algebra generated by the cylinder sets corresponding to finite-dimensional Borel sets. We shall show that for \(t > s\)
\[
P(\hat{\theta}(t) \in A \mid \hat{\theta}(u), u \leq s) = P(\hat{\theta}(t) \in A \mid \hat{\theta}(s))
\]
with probability one. From Theorem 2 we see that \(\hat{\theta}(t)\) is uniquely determined by \(\hat{\theta}(s)\) and \(\{\eta_u: s \leq u \leq t\}\). Using also the Markovian property of \(\{\eta_t\}\), we find that
\[
P(\hat{\theta}(t) \in A \mid \hat{\theta}(u), u \leq s) = \sum_k P(\eta_s = k \mid \hat{\theta}(u), u \leq s) P(\hat{\theta}(t) \in A \mid \hat{\theta}(s), \eta_s = k).
\]
As \(\xi_u\) is determined uniquely by \(\hat{\theta}(u)\), in fact \(\xi_u = a\) if and only if \(\hat{\theta}_i(u) = 0\) for all \(i \notin S_u\), we have
\[
P(\eta_s = k \mid \hat{\theta}(u), u \leq s) = P(\eta_s = k \mid \xi_u, u \leq s) = \hat{f}_k(s).
\]
Hence the left member of (5.1) is
\[
\sum_k \hat{f}_k(s) P(\hat{\theta}(t) \in A \mid \hat{\theta}(s), \eta_s = k).
\]
A similar computation gives the same result for the right member. 

**Remark 5.** In a similar way it may be shown that the joint process \(\{((\eta, \hat{\theta}(t)): 0 \leq t < \infty\}\) is a Markov process.
As mentioned in Section 10 below, prediction and smoothing will be treated elsewhere. Let us here only give a corollary of Theorem 3, which shows that for prediction of \( \{\xi_t\} \), all information in \( \{\xi_u : u \leq s\} \) is summed up in \( \hat{p}(s) \).

**Corollary 1.** *For* \( t > s \)

\[
P(\xi_t = a | \xi_u, u \leq s) = P(\xi_t = a | \hat{p}(s)) = \sum_i \sum_{k \in S_a} \hat{p}_i(s) P_{ik}(t - s).
\]

**Proof.** As mentioned in the proof of Theorem 3, the variable \( \xi_t \) is determined by \( \hat{p}(t) \). Using Theorem 3 we get

\[
P(\xi_t = a | \xi_u, u \leq s) = P(\xi_t = a | \hat{p}(u), u \leq s) = P(\xi_t = a | \hat{p}(s))
\]

\[
= \sum_i P(\eta_s = i | \hat{p}(s)) P(\xi_t = a | \eta_s = i)
\]

\[
= \sum_i \hat{p}_i(s) \sum_{k \in S_a} P_{ik}(t - s).
\]

6. **Differential Equations for the State Estimates**

In the appendix it is shown that the absolute probabilities \( p_k(t), k \in S, \) of the Markov chain \( \{\eta_t\} \) satisfy the integrated version of Kolmogorov's forward equation

\[
dp_k(t) - \sum_i p_i(t) q_{ik} \, dt. \quad (6.1)
\]

An equation of the type

\[
dx(t) = f(x, t) \, dt, \quad (6.2)
\]

where \( x \) and \( f \) are at most countably infinite-dimensional deterministic or stochastic functions, is here interpreted in the following way for each component \( x_k \),

\[
x_k(t) - x_k(s) = \int_s^t f_k(x(u), u) \, du, \quad (6.3)
\]

where the integral is a Lebesgue integral. The notation (6.2) for Eq. (6.3) is similar to the one used for Itô's differential equation, see [17, p. 391]. In fact (6.2) may be viewed as a degenerated Itô differential equation with vanishing diffusion part. Equations of the type (6.2) have many properties in common with the usual type of ordinary differential equation systems. For instance if \( x \) in (6.2) and \( y \) in

\[
dy = g(y, t) \, dt
\]
are one-dimensional, and \( z(t) = x(t) y(t) \), then
\[
dz = x \, dy + y \, dx. \tag{6.4}
\]
A similar property which we will need in the future is the following. Suppose that \( \inf_t y(t) > 0 \). Then
\[
d(x/y) = (1/y) \, dx - (x/y^2) \, dy. \tag{6.5}
\]
Equation (6.4) is the formula for integration by parts for Lebesgue integrals, see for instance [24, p. 104], where it is proved by an application of Fubini's Theorem. Another way of proving relations like (6.4) and (6.5) is to use approximations in \( L_1 \)-norm of \( f(x(\cdot), \cdot) \) and \( g(y(\cdot), \cdot) \) with continuous functions.

Let us suppose that \( \xi_u = g(\eta_u) \) has no jumps for \( s < u \leq t \), and that \( \xi_s = a \). Then
\[
\hat{P}(t) = \hat{P}(s) \, P(a, t, s) \tag{6.6}
\]
according to (4.3). To use the result in the appendix, let us modify our Markov chain by lumping together all states outside \( S_a \) into one absorbing state \( \delta \). The transition intensities \( \lambda_{ik} \) remain unchanged for \( i, k \in S_a \), and the intensity for a transition from \( k \) to \( \delta \) becomes
\[
\lambda_k = \sum_{i \in S_a} q_{ki} . \tag{6.7}
\]
Let \( p_k^* \), \( k \in S_a \cup \{ \delta \} \), denote the absolute probabilities of this new chain. It follows from Eq. (A.5) in the appendix that
\[
dp_k^*(u) = \sum_{i \in S_a} p_i^*(u) \, q_{ik} \, du, \quad k \in S_a , \tag{6.8}
\]
while
\[
dp_\delta^*(u) = \sum_{i \in S_a} p_i^*(u) \, \lambda_i \, du. \tag{6.9}
\]
Further if we use the initial values
\[
\hat{p}_k^*(s) = \hat{P}_k(s)
\]
it follows that
\[
\hat{P}_k(u) = p_k^*(u) \left/ \sum_{i \in S_a} p_i^*(u) \right. = p_k^*(u) [1 - p_\delta^*(u)], \tag{6.10}
\]
for \( s < u \leq t \). As
\[
\inf_{s < u < t} [1 - p_\delta^*(u)] > 0
\]
it follows from (6.5) that
\[ d\hat{p}_k(u) = \sum_{i \in S_a} p_i^* q_{ik}(1 - p_i^*) \, du + p_k^* \sum_{i \in S_a} p_i^* \lambda_i(1 - p_i^*)^2 \, du \]
\[ = \sum_{i \in S_a} \hat{p}_i(u) q_{ik} \, du + \hat{p}_k(u) \hat{\lambda}(u) \, du, \]
where
\[ \hat{\lambda}(t) = \sum_{k \in S} \hat{p}_k(t) \lambda_k. \] (6.12)

We have now proved the following theorem.

**Theorem 4.** Suppose that the observed process \( \xi_u = g(\eta_u) \) has no jumps for \( s < u \leq t \). Then for \( k \in S_a \), where \( \alpha = \xi_s \),
\[ \hat{p}_k(t) = \frac{p_k^*(t)}{\sum_{i \in S_a} p_i^*(t)}, \] (6.13)
where \( p_k^*, k \in S_a \), satisfy the equations
\[ p_k^*(t) = \hat{p}_k(s) + \int_s^t \sum_{i \in S_a} p_i^*(u) q_{ik} \, du. \] (6.14)

Further
\[ \hat{p}_k(t) = \hat{p}_k(s) + \int_s^t \left[ \sum_{i \in S_a} \hat{p}_i(u) q_{ik} + \hat{p}_k(u) \hat{\lambda}(u) \right] \, du, \] (6.15)
where \( \hat{\lambda}(t) \) is given by (6.12).

**Remark 6.** Note that the differential equation system
\[ dp_k^* = \sum_{i \in S_a} p_i^* q_{ik} \, dt, \] (6.16)
\[ k \in S_a, \] corresponding to (6.14) is linear, while the system
\[ d\hat{p}_k = \sum_{i \in S_a} \hat{p}_i q_{ik} \, dt + \hat{p}_k \hat{\lambda} \, dt \] (6.17)
is nonlinear, as \( \hat{p}_i, i \in S_a \), enter in \( \hat{\lambda} \), see (6.12).

**Remark 7.** Let \( N_t \) be the number of jumps of \( \{\xi_s; 0 < s \leq t\} \) as before.
With an obvious interpretation of \(dN_t\) we can put (4.4) and (6.17) together. Set \(\alpha = \xi_{t-o}\) and \(\beta = \xi_t\). Note that \(\alpha = \beta\) except at jumps. Then

\[
d\hat{p}_k(t) = \sum_{i \in \mathcal{S}_\alpha} \hat{p}_i(t) q_{ik} \, dt + \hat{p}_k(t) \hat{\lambda}(t) \, dt
\]

\[
+ \left[ \left( \sum_{i \in \mathcal{S}_\alpha} \hat{p}_i(t-0) \sum_{j \in \mathcal{S}_\beta} q_{ij} \right) - \hat{p}_k(t-0) \right] dN_t
\]

for \(k \in \mathcal{S}_\beta\).

From (6.15) we see that \(\hat{\lambda}\) is integrable. This is stated in the next theorem which also gives an interpretation of \(\hat{\lambda}(t)\).

**Theorem 5.** With probability one \(\lambda\) is integrable over all finite time intervals. Further if no jumps occur in \((s, t), s < t\), then

\[
\exp\left[-\int_s^t \lambda(u) \, du\right] = P(N_t - N_s = 0 \mid u \leq s),
\]

(6.19)

and if \(\hat{\lambda}\) is continuous at \(t\),

\[
\hat{\lambda}(t) = \lim_{h \to 0^+} P(N_{t-h} - N_t > 0 \mid \xi_u, u < t)/h.
\]

(6.20)

**Proof.** Suppose \(\xi_s = a\), set for \(s < u < t\)

\[
p^*(u) = \hat{p}(s) P(a, u - s),
\]

and introduce the absorbing state \(\delta\) with

\[
p_\delta^*(u) = 1 - \sum_{k \in \mathcal{S}_a} p_k^*(u)
\]

as in the discussion preceding Theorem 4, see (6.9). From (6.9), (6.10), and (6.12) we see that

\[
dp_\delta^*(u) = [1 - p_\delta^*(u)] \hat{\lambda}(u) \, du,
\]

which gives

\[
1 - p_\delta^*(t) = \exp\left[-\int_s^t \lambda(u) \, du\right].
\]

(6.21)

As

\[
P(N_t - N_s = 0 \mid \xi_u, u \leq s) = \sum_{k \in \mathcal{S}_a} p_k^*(t)
\]
we get (6.19). The right continuity of $\xi_t$ implies that there is an interval $(t, t+h)$ without jumps. For such an interval

$$P(N_{t+h} - N_t > 0 \mid \xi_u, u \leq t) = 1 - \exp \left[ - \int_t^{t+h} \lambda(u) \, du \right].$$

Dividing by $h$ and letting $h$ decrease to zero we get (6.20).

Remark 8. From (6.19) we see that given $\xi_u, u \leq s$, the conditional distribution function of the time until the next jump of $\{\xi_t\}$ is given by

$$F^\eta(t) = 1 - \exp \left[ - \int_s^t \lambda(u) \, du \right]$$

with $\lambda(u), u \geq s$, computed as if no jumps occur in $(s, t)$, compare [13, Eq. (4.1)].

7. Markov Chains with an Unobserved Component

In this section we shall suppose that the Markov chain $\{\eta_t\}$ has the form

$$\eta_t = (\xi_t, \xi_t),$$

where $\{\xi_t\}$ is an unobserved component with state space $S_1$ and $\{\xi_t\}$ is the observed component with state space $S_2$. Slightly changing the notation used earlier, we shall for $i, k \in S_1, i \neq k$, and $a, b \in S_2, a \neq b$, let $q_{ki}(a)$ and $\lambda_{ab}(k)$ denote the transition intensities from the state $\eta_t = (k, a)$ to the states $(i, a)$ and $(k, b)$ respectively. The probability of a simultaneous change of $\{\xi_t\}$ and $\{\xi_t\}$ is supposed to be zero. We set

$$q_{ki}(a) = -q_{kk}(a) = \sum_{i \neq k} q_{ki}(a)$$

and

$$\lambda_{ab}(k) = -\lambda_{aa}(k) = \sum_{b \neq a} \lambda_{ab}(k).$$

If $q_{ki}(a), k, i \in S_1$, and hence also $q_{k}(a), k \in S_1$, are independent of $a$, we shall use the notations $q_{ki}$ and $q_{k}$. In this case the unobserved component forms a Markov chain itself.

Further we now let $P(a, t)$ denote the matrix on $S_1 \times S_1$ with elements

$$P_{ki}(a, t) = P(\xi_{s+t} = i, \xi_s = a, 0 \leq u \leq t \mid \xi_u = k, \xi_s = a)$$

and $Q(a, b)$ the diagonal matrix on $S_1 \times S_1$ with diagonal elements

$$Q_{ kk}(a, b) = \lambda_{ab}(k).$$
Then the formula (4.2) remains valid with \( p(0) \) as the initial distribution over \( S_1 \) of \( \xi_s \), and \( \dot{p}(t) \) as the row vector over \( S_1 \) with elements

\[
\dot{p}_k(t) = P(\xi_t = k \mid \xi_s, 0 \leq s \leq t).
\] (7.3)

The change in \( \dot{p}_k \) at a jump, see (4.4), is now determined by

\[
\dot{p}_k(\tau) = \dot{p}_k(\tau - 0) \lambda_{ab}(k) \left/ \sum_{i \in S_1} \dot{p}_i(\tau - 0) \lambda_{ab}(i) \right.,
\] (7.4)

where \( \alpha = \xi_{\tau-0} \) and \( \beta = \xi_\tau \). In an interval \((s, t]\) without observed jumps, we have, see (6.13) and (6.16),

\[
\dot{p}_k(t) = p_k^*(t) \left/ \sum_{i \in S_1} p_i^*(t),
\]

where \( p_k^* \) satisfies the equation

\[
dp_k^* = \sum_{i \in S_1} p_i^* q_{ik}(\alpha) \, dt - p_k^* \lambda_a(k) \, dt,
\] (7.5)

with \( \alpha = \xi_s \) and \( p_k^*(s) = \dot{p}_k(s) \). Furthermore, see (6.17)

\[
d\dot{p}_k = \sum_{i \in S_1} \dot{p}_i q_{ik}(\alpha) \, dt + \dot{p}_k [\dot{\lambda}_a(t) - \lambda_a(k)] \, dt,
\] (7.6)

where

\[
\dot{\lambda}_a(t) = \sum_{k \in S_1} \dot{p}_k(t) \lambda_a(k).
\] (7.7)

The general Eq. (6.18), valid over arbitrary time intervals becomes

\[
d\dot{p}_k(t) = \sum_{i \in S_1} \dot{p}_i(t) q_{ik}(\beta) \, dt + \dot{p}_k(t) [\dot{\lambda}_a(t) - \lambda_a(k)] \, dt
\]

\[
+ \dot{p}_k(t - 0) [\lambda_{ab}(k)/\dot{\lambda}_{ab}(t - 0) - 1] \, dN_t,
\] (7.8)

where \( \alpha = \xi_{t-0}, \beta = \xi_t \) and

\[
\dot{\lambda}_{ab}(t) = \sum_{k} \dot{p}_k(t) \lambda_{ab}(k).
\] (7.9)

Specializing to the case where \( q_{ik}(a) \), \( i, k \in S_1 \), are independent of \( a \) we get a formula equivalent to (1.2).

From Theorem 3 it follows that for the model considered in this section with \( \dot{p}(t) \) defined by (7.3), the process \( \{\dot{p}(t), \xi_t; t \in [0, \infty)\} \) is a Markov process. In general \( \{\dot{p}(t)\} \) is not a Markov process. For the doubly stochastic
Poisson process discussed in the introduction and in the following example the process \( \hat{\eta}(t) \) is however Markovian.

**Example 1** (A doubly stochastic Poisson process). Let \( \{\xi_t\} \) be a doubly stochastic Poisson process of the type mentioned in the introduction in connection with (1.5)–(1.9). The Markov chain \( \{\xi_t\} \), controlling the intensity of \( \{\xi_t\} \) is now allowed to have infinitely many states. Then Eqs. (1.5), (1.7) and (1.9) are easily seen to be special cases of (7.6), (7.4) and (7.5), respectively. Further \( \{\hat{\eta}(t)\} \) is a Markov process, that is for measurable \( A \) and \( t > s \)

\[
P(\hat{\eta}(t) \in A \mid \hat{\eta}(u), u \leq s) = P(\hat{\eta}(t) \in A \mid \hat{\eta}(s)).
\]

This follows, compare the proof of Theorem 3, from the fact that \( \hat{\eta}(t) \) is determined by \( \hat{\eta}(s) \) and the increments \( \xi_u - \xi_s, s \leq u \leq t \). Further given \( \eta_u \) these increments are conditionally independent of the increments \( \xi_s - \xi_u \) for \( u \leq s \).

**Example 2** (A Markov chain with random time). Let \( \{(\xi_t, \eta_t)\} \) be a Markov chain of the type described in the beginning of this section, with

\[
q_{ik}(a) = q_{ik}
\]

independent of \( a \), and

\[
\lambda_{ab}(k) = r_k \lambda_{ab},
\]

where \( \lambda_{ab} \) is independent of \( k \). If all \( r_k \) were replaced by 1, then \( \{\xi_t\} \) would be a Markov chain with transition intensities \( \{\lambda_{ab}\} \). In the present case we may interpret the unobserved component \( \{\xi_t\} \) as a random time with rate \( r_k \) at time \( t \), if \( \xi_t = k \). From Eqs. (7.6) and (7.7) we get for intervals between jumps

\[
d\hat{\eta}_k(t) = \sum_{i \in S_k} \hat{\eta}_i(t) q_{ik} dt + \hat{\eta}_k(t) [\hat{r}(t) - r_k] \lambda_k dt,
\]

with \( \alpha = \xi_t \), and where

\[
\hat{r}(t) = \sum_{k \in S_k} \hat{\eta}_k(t) r_k
\]

is the average conditional rate. At jump epochs, see (7.4), we have

\[
\hat{\eta}_k(\tau) = \hat{\eta}_k(\tau - 0) r_k/\hat{r}(\tau - 0).
\]

Note that the model in this example includes the doubly stochastic Poisson process in Example 1 as a special case.
A restriction of the applicability of Markov chain models is that the time spent in a state is required to be exponentially distributed. Dropping this requirement we get a semi-Markov process. However, then some powerful analytical tools, as for instance Kolmogorov's differential equations, are no longer available. The object of this section is to sketch a method of approximating a semi-Markov process with a function of a Markov chain. If then the Markov chain model is used, the state of the chain will not be directly observable. Hence the methods of the present paper should be useful. The method of approximating a semi-Markov process with a function of a Markov chain may be conceived as a generalization of the methods in queueing theory of service in phases and arrivals in stages, see [25, pp. 24–32].

Let \( \{\zeta_t\} \) be the semi-Markov process corresponding to a Markov renewal process \( \{(\theta_n, \tau_n)\} \), that is

\[
\zeta_t = \theta_n, \quad \tau_n \leq t < \tau_{n+1},
\]

see [26]. Note that \( \{\theta_n\} \) is a discrete time Markov chain. The sojourn time \( \tau_{n+1} - \tau_n \) corresponding to the pair \( (\theta_n, \theta_{n+1}) \) has the distribution function

\[
F_{i,j}(t) = P(\tau_{n+1} - \tau_n \leq t \mid \theta_n = i, \theta_{n+1} = j),
\]

\( t \geq 0 \). The idea on which the approximation is based is to approximate the distribution function \( F_{i,j} \) with the distribution function of the time to absorption of a finite state continuous time Markov chain. We shall need the following lemma. Approximation is taken in Levy metric, see [27, p. 253], but of course other measures of fit, ensuring for instance also good approximations of moments, can be used.

**Lemma 4.** Let \( F \) be the distribution of a nonnegative random variable. Then \( F \) can be approximated arbitrarily closely in Levy metric by the distribution of the time to absorption of a finite state, continuous time Markov chain with one absorbing state.

**Proof.** We first note that \( F \) can be approximated arbitrarily closely in Levy metric by a finite linear combination

\[
\sum_{i=1}^{n} p_i H(t - t_i),
\]

\( p_i \geq 0, \sum p_i = 1 \), of one-point distributions at \( t_i > 0, i = 1, \ldots, n \), that is \( H(t) = 0 \) for \( t < 0 \) and \( H(t) = 1 \) for \( t \geq 0 \). To see this, think for instance
of the graph of the empirical distribution of a sample of independent random variables with distribution $F$. Further a one-point distribution at $t_i > 0$ can be approximated arbitrarily closely in Levy metric by the distribution of a sum of $n_i$ exponentially distributed variables with expectation $t_i/n_i$. Just choose $n_i$ large enough.

The approximating Markov chain is constructed in the following way. The state space consists of 0 and the pairs $(i, j), i = 1, \ldots, n, j = 1, \ldots, n_i$. The chain starts in state $(i, 1)$ with probability $p_i$. The only possible transitions are $(i, j) \rightarrow (i, j + 1)$ if $j < n_i$ and $(i, j) \rightarrow 0$ if $j = n_i$. The corresponding transition intensities depend only on $i$ and equal $n_i/t_i, i = 1, \ldots, n$. The verification that the Markov chain has the properties claimed in the lemma is straightforward.

The approximation method in the proof above is constructive, but definitely not the most effective one. Think of the case when the original distribution is exponential!

The lemma will now be used to prove the following theorem.

**Theorem 6.** Let $\{\xi_t\}$ be a semi-Markov process and let $\epsilon > 0$ be given. Then there exists a Markov chain $\{\eta_t\}$ and a function $g$, such that the process $\{\xi_t\}$ with $\xi_t = g(\eta_t)$ has the following properties.

(i) The sequences of states visited by $\{\xi_t\}$ and $\{\eta_t\}$ are discrete time Markov chains with identical distributions.

(ii) Given two successive states, the distributions of the sojourn times of $\{\xi_t\}$ and $\{\eta_t\}$ differ at most by $\epsilon$ in Levy distance.

**Proof.** Let $S$ be the state space of $\{\xi_t\}$. Let $S_1$ contain the state spaces of all the approximating Markov chains of Lemma 4 corresponding to the distribution functions $F_{ij}, i, j \in S$, and the given $\epsilon$. The absorbing state of such a chain we denote by 0 as in proof of the lemma. Hence $0 \in S_1$. Further for fixed $(i, j)$, we let $p_k(i, j), k \in S_1$, denote the initial distribution, and $q_{kl}(i, j), k, l \in S_1$, the transition intensities of the chain in the lemma.

The Markov chain $\{\eta_t\}$ which we shall define has state space $S \times S \times S_1$. Let $P_{ij}$ denote the transition probabilities of the discrete time Markov chain $\{\theta_n\}$ corresponding to $\{\xi_t\}$, see (8.1). Then the transition intensity from $(i, j, k) \in S \times S \times S_1$ with $k \neq 0$, to a state of the type $(i, j, l)$ is $q_{kl}(i, j)$, and no other transitions are possible. From a state of the type $(i, j, 0)$ transition is immediate, and it leads with probability

$$P_{jj'}p_k(j, j')$$

to the state $(j, j', k)$. Further we set

$$g((i, j, k)) = i.$$
It follows from Lemma 4 that \( \{\xi_t\} \) and \( \{\eta_t\} \) have the properties described in the theorem.

We note that the constructed chain \( \{\eta_t\} \) does not satisfy the conditions of Section 3 as the states of the type \((i, j, 0)\) are instantaneous. We can however eliminate these states by replacing a transition sequence

\[
(i, j, k) \rightarrow (i, j, 0) \rightarrow (j, j', l)
\]

by the direct transition

\[
(i, j, k) \rightarrow (j, j', l).
\]

Above we have only regarded approximation of the distributions of the times between jumps. A more thorough study of the approximation problem might include approximation of the distribution of the whole process with an appropriate topology. This topology could be chosen similar to the topology on the space \( D \) in [28, Chap. 3]. However, as we have processes with discrete state spaces some modification is suitable.

Let us conclude this section by remarking that if a function \( h(\xi_t) \) of the semi-Markov process is observed, the approximation method above leads us to consider the Markov chain \( \{\eta_t\} \) and the observed function \( h(g(\eta_t)) \). This problem is still of the type studied in this paper and computation of the conditional state probabilities leads to a system of ordinary differential equations.

On the other hand, a direct approach to computation of the conditional distribution of the state of the semi-Markov process leads to a system of partial differential equations. To get a conditional distribution which forms a Markov process, we namely have to consider not only the unknown state variable, but also the unknown time \( \tau \) which has elapsed since the last transition. Hence we have to derive equations for

\[
P(\tau \leq t, \xi_t = k | g(\eta_s), s \leq t),
\]

\( k \in S, t \geq 0 \), which in general leads to a system of partial differential equations.

9. Computation of Likelihood Ratios

Let \( \{\eta_t\} \) be a right-continuous Markov chain of the type described in Section 3 and 4 with \( \xi_t = g(\eta_t) \). We have, see (6.12),

\[
\lambda(t) = \sum_{k \in S} \hat{p}_k(t) \lambda_k.
\]

(9.1)
Further we set

$$\hat{\lambda}(t, a, b) = \sum_{k \in S_a} p_k(t) \sum_{i \in S_0} q_{ki} \quad (9.2)$$

for $a \neq b$. For a jump epoch $\tau$ we let $\hat{\lambda}(\tau - 0, a, b)$ denote the right member of (9.2) with $\hat{p}_k(t)$ replaced by $\hat{p}_k(\tau - 0)$.

**Theorem 7.** Let $P_0$ and $P_1$ be two probability measures such that $P_1$ is absolutely continuous with respect to $P_0$ on the $\sigma$-algebra $\mathcal{F}_t$ generated by $\xi_s = g(\eta_s)$ for $0 \leq s \leq t$, and let $\hat{\lambda}_j(t)$ and $\hat{\lambda}_j(t, a, b)$ be defined by (9.1) and (9.2) with respect to $P_j$, $j = 0, 1$. Let $N_t$ be the number of jumps of $\{\xi_s: 0 < s \leq t\}$, and let $\tau_1 < \tau_2 < \cdots$ be the epochs of jump. Then the Radon–Nikodym derivative of $P_1$ with respect to $P_0$ is given by

$$\frac{dP_1}{dP_0} = \frac{p_1(a_0)}{p_0(a_0)} \prod_{k=0}^{n} \frac{\hat{\lambda}_1(\tau_k - 0, a_{k-1}, a_k)}{\hat{\lambda}_0(\tau_k - 0, a_{k-1}, a_k)} \exp \left\{ - \int_0^t [\hat{\lambda}_1(s) - \hat{\lambda}_0(s)] \, ds \right\}, \quad (9.3)$$

where $p_j(a) = P_j(\xi_0 = a)$, $j = 0, 1$, $a_0 = \xi_0$ and $a_k = \xi_{\tau_k}$, $k \geq 1$.

**Proof.** We shall use a method of proof similar to the one used in the proof of Theorem 3.2 in [20]. We note that there is a natural mapping from the set

$$A(n, a_0, \ldots, a_n) = \{N_t = n, \xi_0 = a_0, \xi_{\tau_k} = a_k, k = 1, \ldots, n\} \quad (9.4)$$

to the subset of Euclidean $n$-space

$$T_n = \{(s_1, \ldots, s_n): 0 < s_1 < \cdots < s_n \leq t\}, \quad (9.5)$$

such that $\mathcal{F}_t \cap A(n, a_0, \ldots, a_n)$ corresponds to the $\sigma$-algebra $\mathcal{B}_n$ of Borel subsets of $T_n$. From (4.9) we see that the corresponding measure on $\mathcal{B}_n$ is absolutely continuous with respect to Lebesgue measure with density

$$\sum_k p_k(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t).$$

As the sure event is a countable union of disjoint events of the type (9.4), the theorem follows if we can show that for $N_t = n$, $a_k = a_k$, $k = 0, 1, \ldots, n$, and $\tau_k = s_k$, $k = 1, \ldots, n$, we have

$$\sum_k p_k(n, a_0, \ldots, a_n, s_1, \ldots, s_n, t)$$

$$= P(\xi_0 = a) \prod_{k=1}^n \hat{\lambda}(s_k - 0, a_{k-1}, a_k) \exp \left\{ - \int_0^t \hat{\lambda}(s) \, ds \right\}. \quad (9.6)$$
We shall give the details of the proof of (9.6) for \( n = 0 \) and \( n = 1 \). The arguments for \( n = 1 \) can in a straightforward way be transformed into a formal induction proof of (9.6).

Let us start with \( n = 0 \). We shall show that for \( \xi_0 = a \) and \( N_t = 0 \) we have

\[
\hat{p}(0) P(a, t) = P(\xi_0 = a) \exp \left[ - \int_0^t \hat{\lambda}(u) \, du \right] \hat{p}(t),
\]

which gives (9.6) for \( n = 0 \). For \( 0 \leq s \leq u \leq t \), put

\[
\hat{p}^*(u) = \hat{p}(s) P(a, u - s).
\]

From (6.10) and (6.21), it follows that

\[
\hat{p}(s) P(a, t - s) = \exp \left[ - \int_s^t \hat{\lambda}(u) \, du \right] \hat{p}(t).
\]

If we set \( s = 0 \) in (9.9) and observe that

\[
P_{k}(0) = \delta_{k,0} = \delta_{k_0}\hat{p}(0)
\]

for \( k \in S_a \), we get (9.7).

Let us now consider \( n = 1 \). We shall show that

\[
\hat{p}(0) P(a, s) Q(a, b) P(b, t - s) = P(\xi_0 = a) \hat{\lambda}(s - 0, a, b) \exp \left[ - \int_0^t \hat{\lambda}(u) \, du \right] \hat{p}(t),
\]

for \( \xi_0 = a, \tau_1 = s, \xi_1 = b \) and \( N_t = 1 \). From (9.2) it follows that

\[
\hat{p}(s - 0) Q(a, b) = \hat{\lambda}(s - 0, a, b) \hat{p}(s).
\]

The members of (9.7) are continuous element-wise in \( t \). Hence

\[
\hat{p}(0) P(a, s) = P(\xi_0 = a) \exp \left[ - \int_0^s \hat{\lambda}(u) \, du \right] \hat{p}(s - 0).
\]

Combining the last two equations with (9.9) we get (9.10).

Remark 9. It follows from the proof above that the vector-matrix product in (4.2) before normalization equals

\[
P(\xi_0 = a_0) \prod_{k=1}^{N_t} \hat{\lambda}(\tau_k - 0, \alpha_{k-1}, \alpha_k) \exp \left[ - \int_0^t \hat{\lambda}(s) \, ds \right] \hat{p}(t).
\]
Remark 10. As mentioned in the proof above, the sure event $\Omega$, may for fixed $k$ be written as a countable disjoint union

$$\Omega = \bigcup A(a_0, a_1, \ldots, a_n)$$

of sets of the type (9.4). Further, for each of these sets there is a finite measure corresponding to Lebesgue measure on $(T, \mathcal{B})$ in (9.5). Let $\sigma$ denote the $\sigma$-finite measure thus obtained on $(\Omega, \mathcal{F})$. Then the Radon-Nikodym derivative of a probability measure $P$ with respect to $\sigma$ is given by

$$\frac{dP}{d\sigma} = P(\xi_0 = \alpha_0) \prod_{k=1}^{N} \lambda(\tau_k - 0, \alpha_{k-1}, \alpha_k) \exp \left[ - \int_0^t \lambda(s) \, ds \right]. \quad (9.12)$$

This result may be viewed as a generalization of Theorem 3.2 in [20].

Remark 11. There is a striking similarity between (1.10) and (9.3). Results showing, that we for some purpose, for instance detection or control, may treat a partially observed process as a completely observed process, if we replace the process values with their estimates, are sometimes called separation theorems. See for instance [29] for detection in Wiener noise, or [10] for detection of doubly stochastic Poisson processes.

10. Extensions and Related Problems

In Section 5 we have briefly touched upon prediction of $\xi_t$ given observations of $\xi_u, \ u \leq s$, where $s < t$. A more systematic study of prediction and smoothing, based on the methods and results of the present paper will be given elsewhere.

In Section 5, Theorem 3 and Remark 6, it is ascertained that $\{\hat{p}(t)\}$ and $\{(\eta_t, \hat{p}(t))\}$ are Markov processes. In [13, Section 7] integrodifferential equations are derived for the state distributions of these processes. Those results could be generalized at least to partially observed chains with finite state space. Alternatively the Markov processes $\{\hat{p}(t)\}$ and $\{(\eta_t, \hat{p}(t))\}$ could be characterized by their characteristic operators, see [9, Corollary of Theorem 3] for a special case. One reason for studying the properties of $\{\hat{p}(t)\}$ is that $\hat{p}(s), s \leq t$, and $\xi_s - g(\eta_s), s \leq t$, generate the same $\sigma$-algebras. Hence by introduction of $\{\hat{p}(t)\}$ an optimal stopping problem for $\{\xi_t\}$ may be transformed into an optimal stopping problem for a Markov process, for which general methods exist, see for instance [30].

The results in Section 9 could be made a starting point for a study of statistical methods, in particular likelihood methods, for partially observed Markov chains, compare the results in [20] for completely observed Markov
chains. A problem related to statistical problems is the approximation problem shortly discussed in Section 8.

The doubly stochastic Poisson process in [13] is a point process. The methods of the present paper can be used to study point processes generated by transitions of Markov chains in a systematic way, see [31].

A problem which is at least of some theoretical interest is to extend the considered class of Markov chains \( \{\eta_t\} \) to include also chains with sample functions that are more complicated than step functions.

Finally the process \( \{g(\eta_t)\} \) may be regarded as an approximation of \( \{\eta_t\} \) and it might be of interest to regard sequences \( \{g_n\} \) of approximations that are more and more informative in the sense that, for all \( k \in S \), \( g_n^{-1}(\{g_n(k)\}) \), \( n = 1, 2, \ldots \), decreases to the one point set \( \{k\} \) as \( n \) tends to infinity. This type of approximation may be compared with the Markovian approximation in [32], where also the conditional distribution of the original process, given the trajectory of the approximation is considered, see [32, p. 51].

---

**APPENDIX: ON KOLMOGOROV'S FORWARD EQUATION FOR THE ABSOLUTE PROBABILITIES OF A MARKOV CHAIN**

Let \( \{\eta_t: t \in [0, \infty)\} \) be a continuous time Markov chain satisfying the conditions of Section 3. In particular the forward Kolmogorov equation

\[
P'_{ik}(t) = \sum_j P_{ij}(t) q_{jk}
\]

(A.1)

is satisfied. We set

\[
p_k(t) = P(\eta_t = k)
\]

(A.2)

and observe that

\[
p_k(t) = \sum_i p_i(0) P_{ik}(t).
\]

Multiplying (A.1) with \( p_i(0) \) and summing we formally get

\[
p_k(t) = \sum_j p_j(t) q_{jk}.
\]

(A.3)

Under what conditions is (A.3) true? The object of this appendix is to give a partial answer to this question by showing that the integrated version of (A.3) always holds true and to exhibit an example showing that the right member of (A.3) may be infinite. We note that a well-known sufficient condition for (A.3) to hold is that \( q_{jk} \) for fixed \( k \) is bounded with respect to \( j \). (For instance may the proof of (A.1) in [16, p. 472] be used.)
If we integrate (A.1) over the interval \((s, t)\) we get

\[
P_{ik}(t) - P_{ik}(s) = \sum_j q_{jk} \int_s^t P_{ij}(u) \, du.
\]

Multiply by \(p_i(0)\) and sum with respect to \(i\). By considering the positive and negative terms separately we see that it is permitted to exchange summation and integration. We get

\[
p_k(t) - p_k(s) = \sum_j q_{jk} \int_s^t p_i(u) \, du.
\] (A.4)

The relation (A.4) will be written symbolically

\[
dp_k(t) = \sum_j q_{jk} p_j(t) \, dt,
\] (A.5)

compare Section 6. In the sum all terms except one are nonnegative. Hence we may exchange summation and integration in (A.4).

Let us now give an example showing that the right member of (A.3) may be infinite. We note that this is easy for \(t = 0\). Just choose a chain with \(\sup j q_{jk} = \infty\) and choose \(p_j(0)\) such that

\[
\sum_j p_j(0) q_{jk} = \infty.
\]

In the following example the right member of (A.3) is infinite for \(t = 1\). The construction is based on the fact that the distribution of the sum of \(n\) independent exponentially distributed variables with expectation \(1/n\) tends to a one-point distribution at 1 as \(n\) tends to infinity.

**Example.** Consider a Markov chain \(\{\eta_t\}\) with state space

\[
S = \{0\} \cup \{(n, m) : m = 0, 1, \ldots, n, n = 1, 2, \ldots\}.
\]

The state 0 is absorbing and the transition intensity from the state \((n, m)\) is \(n\). Further from \((n, m)\) the only possible transition is to \((n, m + 1)\) if \(m < n\) and to 0 if \(m = n\). The chain starts in the state \((n, 0)\) with probability \(r_n\), \(\sum_n r_n = 1\). For instance from the properties of the Poisson process it follows that the probability of a transition from \((n, 0)\) at time zero to \((n, n)\) at time \(t\) is

\[
F_n(t) = \left(\frac{(nt)^n}{n!}\right) e^{-nt}.
\]

Hence

\[
\sum_{k \in S} p_k(t) q_{k0} = \sum_{k,j \in S} p_k(0) P_{kj}(t) q_{j0} = \sum_{n=1}^{\infty} r_n F_n(t) n.
\]
Using Stirlings formula

\[ n! \sim (2\pi n)^{1/2} (n/e)^n \]

we find that

\[ F_n(1) n \sim (n/2\pi)^{1/2}. \]

We see that

\[ \sum_{k=1}^{\infty} p_k(1) q_k = \infty \]

if the sequence \( \{r_n\} \) is chosen such that

\[ \sum_{n=1}^{\infty} r_n n^{1/2} = \infty. \]

ACKNOWLEDGMENTS

For valuable comments I wish to thank Karl Johan Åström and Jan Grandell.

REFERENCES

STATE ESTIMATION FOR MARKOV CHAINS