The Annals of Statistics
1995, Vol. 23, No. 4, 1073–1101

EXPLICIT LIMIT RESULTS FOR MINIMAL SUFFICIENT STATISTICS AND MAXIMUM LIKELIHOOD ESTIMATORS IN SOME MARKOV PROCESSES: EXPONENTIAL FAMILIES APPROACH

BY VALERI T. STEFANOV

Bulgarian Academy of Sciences and University of Western Australia

Finite-state Markov chains with either a discrete or continuous time parameter, Markov renewal processes and Markov-additive processes are considered. We prove that their likelihood functions, in the nonsequential as well as in various sequential cases, belong to special $(n + k, n)$-curved exponential families in general, for which limit results are easily established. Subsequently, asymptotic normality of the corresponding nonsequential and sequential maximum likelihood estimators is established. Also in the case of Markov renewal and Markov-additive processes, stopping times are determined which reduce the corresponding curved exponential families in general to noncurved ones. The latter, together with results of Stefanov, are combined with results of Serfozo to imply explicit solutions in functional limit theorems for the considered processes. In particular, we derive explicit solutions for the important variance parameter in the functional central limit theorems and functional laws of iterated logarithm for those processes. Indeed, our explicit solutions cover more general cases than the known ones, even in the case of finite-state Markov chains. Moreover, we supply explicit solutions, not previously available, in functional limit theorems for Markov renewal processes and Markov-additive processes.

0. Introduction. In the first five sections of the present paper we obtain asymptotic normality of the minimal sufficient statistics and subsequently of sequential and nonsequential maximum likelihood estimators in finite-state Markov chains with either a discrete or continuous time parameter, in Markov renewal processes and in Markov-additive processes. This is done in a unified manner using the special exponential structure of the corresponding likelihood functions. Actually, we prove that these likelihood functions belong to special $(n + k, n)$-curved exponential families in general. For the latter families, limit results are easily established. For the definitions of curved and noncurved exponential families, see Barndorff-Nielsen (1980). Exponential families are a valuable tool in asymptotic statistical theory, as pointed out by Brown (1986). An earlier reference is Berk (1972), who considered an i.i.d.

Received January 1994; revised December 1994.
AMS 1991 *subject classifications*. Primary 62L99, 62M05; secondary 60F17.
Key words and phrases. Curved exponential family, stopping time, sequential estimation, maximum likelihood estimator, Markov chain, Markov renewal process, Markov-additive process, functional limit theorems.

case of noncurved exponential families [cf. also Andersen (1969) and Johansen (1979)]. Sorensen (1986) explored the tools of exponential families for obtaining asymptotic normality of sequential maximum likelihood estimators in some $(2, 1)$-curved exponential families and special $(n + 1, n)$-curved exponential families of stochastic processes. However, these tools are still not fully explored. The results presented in the first five sections can be considered as extensions of Sorensen's (1986) results to the multivariate case of $(n + k, n)$-curved exponential families.

Relevant references to the topic treated in the first five sections of the paper follow. The asymptotic normality of nonsequential maximum likelihood estimators in finite-state Markov chains with either discrete or continuous time parameter were treated by Anderson and Goodman (1957), Billingsley (1961a, b) and Albert (1962) [cf. also Basawa and Prakasa Rao (1980)].

Billingsley (1961a) and Silvey (1961) seem to have been the first to recognize the role of martingale limit results in large-sample inference for stochastic processes. The martingale limit results became and are now the principal tool in asymptotic statistical theory for stochastic processes [see Hall and Heyde (1980), Jacod and Shiryaev (1987) and Karr (1991)]. Both the martingale and exponential families techniques are combined in the case of some exponential families of semimartingales by Küchler and Sorensen (1989) and Sorensen (1991) [cf. also Küchler and Sorensen (1994)]. Earlier references in that area are Heyde and Feigin (1975) and Feigin (1981).

Moore and Pyke (1968) obtained the asymptotic normality of a nonparametric estimator of the semi-Markov kernel in Markov renewal processes, by applying the limit results for the latter processes obtained by Pyke and Schaufele (1964). Except for nonparametric and Bayesian estimation [cf. Gill (1980), Phelan (1990a, b)] other statistical issues for Markov renewal processes have not received much attention [cf. Fygenson (1991)].

To the author's knowledge, large-sample properties of the maximum likelihood estimators in Markov-additive processes are not treated in the literature. In view of the large applicability of these processes in queueing theory [see Çinlar (1972a, b) and Prabhu (1991)], statistical issues related to them are of importance.

Let us remark that, once asymptotic normality is established in the nonsequential case, one gets asymptotic normality in the sequential case for any sequence of stopping times (say, τ_s) such that τ_s/s converges to a positive constant in probability as $s \uparrow +\infty$. This follows by applying well-known results (Anscombe's type of theorem) concerning random time changes in the central limit theorem [see Billingsley (1968)]. However, the above property has not been established in the literature for all stopping times considered in the present paper, although it does in fact hold true for all of them, as follows from our results (cf. Section 6). For example, in the case of a continuous-time Markov chain only, Höpfner (1987) established the abovementioned property for some of the stopping times considered here.

We pursue also the interesting problem of the existence of stopping times such that the corresponding sequential likelihood functions belong to noncurved exponential families in the case of Markov renewal processes and Markov-additive processes [cf. also Stefanov (1988a, b) and Stefanov (1991)]. Such stopping times do exist and are described in Propositions 4.1 and 5.1 below. In the case of Markov renewal processes, they are the same as those which appear in the case of finite-state Markov chains with discrete time parameter. In the case of Markov-additive processes, some other stopping times appear also to have that property. Generally those stopping times possess the optimal properties in noncurved exponential families, which are well known in statistical inference. Moreover, as shown below, they turn out to be very useful in deriving explicit solutions in functional limit laws too.

For the derivation of the limit results in the first five sections, only the special curved exponential representation of the considered processes is essentially used. However, if the above results are combined with a result of Stefanov [(1986), Theorem 1] and with the powerful results of Serfozo (1975), then one gets *explicit* limit results in functional limit laws for those processes too. By *explicit* we mean that an explicit expression of the important variance parameter in the functional central limit theorem and functional laws of iterated logarithm is obtained, which makes these results applicable in practice. Calculable expressions for the variance parameter in the functional central limit theorem for some Markov processes are found in Bhattacharya and Waymire (1990), pages 513–515] and Bhattacharya (1982); see also the references in the latter. An earlier reference in this direction is Gordin and Lifšic (1978). However, our method is much simpler, and, moreover, our explicit solutions cover more general limit results even in the case of finite-state Markov chains (discrete or continuous time parameter) than the known ones which have been derived using other techniques [cf. Chung (1967), Sections 14–16] and Iosifescu [(1980), page 138]. For example, for finite-state Markov chains we get explicit results in functional limit laws for any linear function of the components of the minimal sufficient statistic. They contain as special cases the known limit results concerning any function defined on the finite state space [cf. Chung (1967) and Iosifescu (1980), page 138].

The paper is organized as follows. In Section 1, a general $(n + k, n)$-curved exponential family for which limit results are easily established is introduced. In the subsequent Sections 2–5, it is shown that the nonsequential and various sequential versions of the likelihood functions of finite-state Markov chains with either discrete or continuous time parameter, Markov renewal processes and Markov-additive processes, belong to the curved-exponential family introduced in Section 1. Subsequently asymptotic normality of the corresponding canonical statistics and maximum likelihood estimators in those processes is established. The concepts, which are similar for all the processes, are fully considered in Section 2 only. Section 6 contains hints as to how to obtain more information about the limiting normal distributions; for example, asymptotic independence of components of the maximum likelihood estimators. In Section 7 we combine some of the above results with the powerful results of Serfozo (1975) in order to get explicit solutions in functional limit theorems for the considered processes. The key insights from

Serfozo (1975) and Stefanov (1986) are discussed briefly and explicit limit results in the functional central limit theorem as well as in Strassen's invariance principle are given. Two simple examples are provided too.

Finally, we would like to remark that the general limit results, presented in Section 1, are not only applicable to the processes considered in the present paper. It seems that the curved exponential structure considered here is hidden in many other statistical models, for example, such as models for branching processes and time series.

1. Preliminary results. Let $(X(t))_{t\geq 0}$ be a stochastic process defined on the probability space $(\Omega, \mathscr{A}, \mathscr{P}_\theta)$ with values in $(R^n, \mathscr{B}_{R^n})$, where θ is a parameter, $\theta \in \Theta \subset R^n$, and where Θ is an open set. The time parameter t may be either discrete or continuous. Suppose also that $(X(t))_{t\geq 0}$ is adapted to the filtration $(\mathscr{F}_t)_{t\geq 0}$. Let $P_{\theta,t}$ be the restriction of P_θ to the σ-algebra $\mathscr{F}_t$. We shall assume that for each t there is a σ-finite measure Q_t such that $P_{\theta,t} \ll Q_t$ for each $\theta \in \Theta$ and that the likelihood function $dP_{\theta,t}/dQ_t$ is given by

$$\frac{dP_{\theta,t}}{dQ_t} = \exp\left(\sum_{i=1}^{n} \theta_i X_i(t) - \varphi(\theta)t + \sum_{i=1}^{k} \psi_i(\theta) D_i(t)\right), \tag{1}$$

where $X(t) = (X_1(t), \dots, X_n(t))$, φ: $R^n \to R$ is a twice continuously differentiable function, ψ_i: $R^n \to R$, $i = 1, 2, \dots, k$, are continuous functions and $D_i(t)$, $i = 1, 2, \dots, k$, are random variables such that, P_θ-a.s., $\theta \in \Theta$,

$$|D_i(t)| \leq C(\theta) < +\infty, \qquad i = 1, 2, \dots, k, \tag{2}$$

where $C(\theta)$ is a continuous function which does not depend on t. It is easy to see from the proof of the next propositions that we can relax condition (2) to the following one:

$$|D_i(t)| \text{ stoch} \leq C_i(\theta) \tag{3}$$

where "stoch $\leq$ " means *stochastically not greater than* [for the definition of the latter, see Lehmann (1986), page 84], and $C_i(\theta)$ is a nonnegative random variable with the property that for each $\theta_1 \in \Theta$ there exists an $\varepsilon > 0$ and a neighbourhood $\mathscr{N}$ of θ_1 such that

$$E\big(\sup(\exp(\varepsilon C_i(\theta)) \colon \theta \in \mathscr{N})\big) < +\infty.$$

Note that we can always assume that $C_i(\theta)$, $i = 1, \dots, k$, are independent. Also, for the processes considered in the remaining sections we have $P_{\theta,t} \ll P_{\theta_0,t}$, for some θ_0. Thus, the measure Q_t may be viewed as the one produced from $P_{\theta_0,t}$ after absorbing that part of $dP_{\theta,t}/dP_{\theta_0,t}$ which does not depend on θ.

Conditions (2) and (3) are kinds of weak dependence assumptions which guarantee that $X(t)$ has asymptotic properties similar to those of a process with independent increments (cf. Proposition 1.2).

The propositions given below are easy but very useful, as will become clear in the next sections. They are multivariate extensions of results given by

Sorensen (1986) in the univariate case (see his Section 3) and noted in a special multivariate case (see his Section 7).

PROPOSITION 1.1. *For each i, $i = 1, 2, \dots, n$, and each $\theta \in \Theta$, the following result concerning convergence in probability* (P_θ) *holds*:

$$\frac{X_i(t)}{t} \to \varphi_i'(\theta) \quad P_\theta\text{-in probability as } t \uparrow +\infty,$$

where $\varphi_i'(\theta) = \partial\varphi(\theta)/\partial\theta_i$.

PROOF. Consider, for example, the Laplace transform of $X_1(t)/t$,

$$E_\theta \exp\left(\frac{s_1 X_1(t)}{t}\right)$$
$$= \int \exp\left(\sum_{i=2}^{n} \theta_i X_i(t) + \left(\theta_1 + \frac{s_1}{t}\right) X_1(t) - \varphi(\theta)t + \sum_{i=1}^{k} \psi_i(\theta) D_i(t)\right) dQ_t.$$

The Laplace transform exists in a neighbourhood of 0, as is easily seen from the assumptions made about the model. Let e_i be the unit vector in R^n whose i-th coordinate is 1. From the above equality we derive that

$$E_\theta \exp\left(\frac{s_1 X_1(t)}{t}\right) = \exp\left(t\left(\varphi\left(\theta + \frac{s_1 e_1}{t}\right) - \varphi(\theta)\right)\right)$$
$$\times E_{\theta + s_1 e_1/t} \exp\left(\sum_{i=1}^{k} \left(\psi_i(\theta) - \psi_i\left(\theta + \frac{s_1 e_1}{t}\right)\right) D_i(t)\right).$$

In view of the assumptions made above about $\varphi(\theta)$, $\psi_i(\theta)$, $D_i(t)$, $i = 1, 2, \dots, k$, and $C(\theta)$, it is easy to see that the Laplace transform of $X_1(t)/t$ tends to $\exp(\varphi_1'(\theta)s_1)$ as $t \uparrow +\infty$. This completes the proof. □

PROPOSITION 1.2. *Let*

$$Y(t) = \frac{X(t) - \nabla\varphi(\theta)t}{\sqrt{t}},$$

where $\nabla\varphi(\theta) = (\varphi_1'(\theta), \dots, \varphi_n'(\theta))$. *Then as* $t \uparrow +\infty$, $Y(t)$ *is asymptotically* $N(0, \Sigma)$, *where*

$$\Sigma = \left(\frac{\partial^2 \varphi}{\partial\theta_i\, \partial\theta_j}(\theta)\right)_{i,j=1}^{n}.$$

PROOF. Consider the Laplace transform of $Y(t)$. In view of (1) we have

$$\begin{aligned} E_\theta \exp\left(\sum_{i=1}^n s_i Y_i(t)\right) &= \exp\left(-\sum_{i=1}^n s_i\sqrt{t}\,\varphi_i'(\theta)\right) E_\theta \exp\left(\sum_{i=1}^n \frac{s_i}{\sqrt{t}} X_i(t)\right) \\ &= \exp\left(-\sum_{i=1}^n s_i\sqrt{t}\,\varphi_i'(\theta) + t\left(\varphi\left(\theta + \frac{s}{\sqrt{t}}\right) - \varphi(\theta)\right)\right) \\ &\quad\times E_{\theta+s/\sqrt{t}} \exp\left(\sum_{i=1}^n \left(\psi_i(\theta) - \psi_i\left(\theta + \frac{s}{\sqrt{t}}\right)\right) D_i(t)\right), \end{aligned}$$

where $s = (s_1, \dots, s_n)$ and $\theta + s/\sqrt{t} = (\theta_1 + s_1/\sqrt{t}, \dots, \theta_n + s_n/\sqrt{t})$. Of course in view of the assumptions on $\psi_i(\theta)$, $D_i(t)$, $i = 1, 2, \dots, k$ and $C(\theta)$, we have

$$E_{\theta+s/\sqrt{t}} \exp\left(\sum_{i=1}^n \left(\psi_i(\theta) - \psi_i\left(\theta + \frac{s}{\sqrt{t}}\right)\right) D_i(t)\right) \to 1 \quad \text{as } t \uparrow +\infty.$$

Applying the Taylor expansion for $\varphi(\theta + s/\sqrt{t})$, we get finally that

$$E_\theta \exp\left(\sum_{i=1}^n s_i Y_i(t)\right) \to \exp\left(\frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2\varphi}{\partial\theta_i\,\partial\theta_j}(\theta) s_i s_j\right) \quad \text{as } t \uparrow +\infty.$$

This completes the proof. □

REMARK 1.1. It is straightforward to see that the convergence results in both propositions above hold uniformly on compact subsets of Θ.

We complete this section by stating a well-known convergence result [see Rao (1973), page 338]), which will be referred to in the following sections.

LEMMA 1.1. *Let $S(t) = (S_1(t), \dots, S_k(t))$ be a k-dimensional statistic such that $\sqrt{t}(S(t) - \vartheta)$ is asymptotically k-variate normal $N(0, \Sigma_1)$ as $t \uparrow +\infty$, where ϑ is a k-dimensional parameter. Let $g_1, \dots, g_r$ be r functions of k variables with each g_i a twice continuously differentiable function. Then $\sqrt{t}(g(S(t)) - g(\vartheta))$ is asymptotically $N(0, G\Sigma_1 G^T)$, where*

$$G = \left(\frac{\partial g_i}{\partial\vartheta_j}(\vartheta) : i = 1, \dots, r; j = 1, \dots, k\right).$$

2. Finite-state Markov chains with discrete time. Let $(Z(t))_{t\geq 0}$ be a homogeneous and ergodic m-state Markov chain in discrete time. For simplicity we shall assume that all transition probabilities are positive. The case when some of them equal 0 is treated analogously. Also without loss of generality we assume that the chain starts from state 1 with probability 1. The likelihood function in this case is given by [see, e.g., Basawa and Prakasa Rao (1980) or Stefanov (1991)]

$$L(\pi, t) = \exp\left(\sum_{i,j=1}^m N_{i,j}(t) \ln p_{i,j}\right),$$

where $\pi = (p_{i,j})_{i,j=1}^m$ is the matrix of the one-step transition probabilities; $N(t) = (N_{i,j}(t);\, i, j = 1, \dots, m)$; and $N_{i,j}(t)$ means the number of the one-step transitions from state i to state j in the time interval $[0, t]$.

We shall consider the following type of stopping times:

$$\tau_s = \inf\left(t:\ \sum_{i,j=1}^m a_{i,j} N_{i,j}(t) > s\right), \qquad s > 0, \tag{4}$$

where $a_{i,j}$, $i, j = 1, \dots, m$, are prescribed arbitrary real numbers not all equal to 0. Of course in the particular case when $a_{i,j} = 1$ for each $i, j = 1, \dots, m$, we get the case of fixed sample size. The stopping time given by (4) is finite, that is,

$$P_\pi(\tau_s < +\infty) = 1 \quad \text{if } E_\pi\left(\sum_{i,j=1}^m a_{i,j} N_{i,j}(\eta_1)\right) > 0,$$

where η_1 means the first entry time to state 1 after the chain has left it for the first time. Actually, one gets this easily using the regeneration property of the successive entries to a fixed state, in particular state 1 in this case, and well-known properties of independent and identically distributed summands. Namely, if $\eta_1, \eta_2, \dots$ is the sequence of the successive entries to state 1, then in view of their regeneration property the process $(\sum_{i,j=1}^m a_{i,j} N_{i,j}(\eta_n))_{n \geq 0}$ has stationary independent increments. Of course if the embedded process crosses a level a.s., then the original process crosses that level a.s. too. Then the above claim follows directly from Gut [(1988), Theorem 1.1 on page 76]. Let

$$\Pi_1 := \left(\pi:\ E_\pi \sum_{i,j=1}^m a_{i,j} N_{i,j}(\eta_1) > 0\right).$$

REMARK 2.1. Observe that η_1 is a stopping time which reduces the corresponding curved-exponential family to a noncurved one [see Stefanov (1991), Proposition 1]. Thus, in any particular case, the expectations involved in Π_1 can be found explicitly using well-known analytic properties of noncurved exponential families [Barndorff-Nielsen (1978), page 114 or Brown (1986)]. Therefore, the criterion Π_1 is practicable.

In view of Jacod and Shiryaev [(1987), Theorem 3.4 on page 153] for the sequential version of the likelihood function we have

$$L(\pi, \tau_s) = \exp\left(\sum_{i,j=1}^m N_{i,j}(\tau_s) \ln p_{i,j}\right), \qquad \pi \in \Pi_1. \tag{5}$$

In the sequel we shall show that (5) has a representation of the type given by (1). Consider the following linear system:

$$\begin{aligned} x_{\cdot i} - x_{i\cdot} &= 0, \qquad i = 1, 2, \dots, m-1, \\ \sum_{i,j=1}^m a_{i,j} x_{i,j} - s &= 0, \end{aligned} \tag{6}$$

where $(x_{i,j})_{i,j=1}^m \in R^{m^2}$, $x_{\cdot i} = \sum_{j=1}^m x_{j,i}$, $x_{i\cdot} = \sum_{j=1}^m x_{i,j}$ and $a_{i,j}$, $i,j = 1,2,\dots,m$, are the constants given in (4). It is easy to see that the rank of the system given by (6) is m. Define the random variables $R_1(s),\dots,R_m(s)$ as follows:

$$R_i(s) = \mathbf{1}_{(i)}(Z(\tau_s)) - \mathbf{1}_{(i)}(Z(0)), \qquad i = 1,2,\dots,m-1,$$

$$R_m(s) = s - \sum_{i,j=1}^m a_{i,j} N_{i,j}(\tau_s),$$

where $\mathbf{1}_{(\cdot)}(\cdot)$ is the indicator function and s and $a_{i,j}$, $i,j = 1,2,\dots,m$, are as given in (4). It is easy to see that [cf. also Stefanov (1991)] P_π-a.s., $\pi \in \Pi_1$,

$$(7)\qquad \begin{aligned} N_{\cdot i}(\tau_s) - N_{i\cdot}(\tau_s) &= R_i(s), \qquad i = 1,2,\dots,m-1,\\ \sum_{i,j=1}^m a_{i,j} N_{i,j}(\tau_s) - s &= -R_m(s), \end{aligned}$$

where $N_{\cdot i}(\tau_s) = \sum_{j=1}^m N_{j,i}(\tau_s)$ and $N_{i\cdot}(\tau_s) = \sum_{j=1}^m N_{i,j}(\tau_s)$. In view of (7) the likelihood function given by (5) can be represented as follows:

$$(8)\qquad \exp\left(\sum_{i=1}^{m^2-m} \nu_i(\pi) T_i(N(\tau_s)) - \sum_{i=1}^m \mu_i(\pi) R_i(s) - \mu_{m+1}(\pi) s\right),$$

where $T(\cdot) = (T_1(\cdot),\dots,T_{m^2-m}(\cdot))$: $R^{m^2} \to R^{m^2-m}$ is a certain linear transformation and $\nu_i(\pi)$, $i = 1,\dots,m^2-m$, and $\mu_i(\pi)$, $i = 1,\dots,m$, are certain linear functions of $\ln p_{i,j}$, $i,j = 1,\dots,m$. Of course $T_i(N(\tau_s))$, $i = 1,\dots,m^2 - m$, may be considered as those (m^2-m) components of $N(\tau_s)$ through which the remaining m components are expressed linearly using the linear system given by (6). Any linear transformation produced using the linear system (6) may serve for the representation (8) at the cost only of changing the functions $\nu_i(\pi)$, $i = 1,\dots,m^2-m$, $\mu_i(\pi)$, $i = 1,\dots,m+1$, suitably. Manifestly, all $\mu_i(\cdot)$ and $\nu_i(\cdot)$ are differentiable arbitrarily many times, and

$$|R_i(s)| < \text{const} < +\infty, \qquad i = 1,\dots,m,\ P_\pi\text{-a.s.}$$

Now to prove that the representation (8) is of the type given by (1) it is enough to show that the Jacobian

$$\left|\left(\frac{\partial \nu_i}{\partial p_{u,v}}(\pi)\colon u = 1,\dots,m,\ v = 1,\dots,m-1,\ i = 1,\dots,m^2-m\right)\right|$$

is not equal to 0 for $\pi \in \Pi_1$. In the case of a two- or even three-state Markov chain, one may check by direct lengthy calculations that this is so. However, the general case of m-states (m arbitrary) is not straightforward. Observe first that from the form of (6) it follows that one may choose m components, one from each row of the matrix $N(\tau_s)$, to be linear functions of the remaining (m^2-m) components of $N(\tau_s)$. Denote these components by $N_{i,j_i}(\tau_s)$, $i = 1,\dots,m$. Then of course for each k, $k = 1,\dots,m^2-m$, the function $\nu_k(\pi)$ has the following representation:

$$(9)\qquad \nu_k(\pi) = \ln p_{k,l} - \sum_{i=1}^m c_i \ln p_{i,j_i},$$

for some pair (k,l) and some constants c_i, $i = 1,\dots,m$, where $(k,l) \neq (i, j_i)$ for each i, $i = 1,\dots,m$. Consider now the following matrix of $(m^2 - m)$-dimensional vectors:

$$\nabla \ln(p_{k,l})_{k,l=1}^{m} := \left(\frac{\partial \ln p_{k,l}}{\partial p_{u,v}} : u = 1,\dots,m, v = 1,\dots,m-1 \right)_{k,l=1}^{m}.$$

It is easy to see that if we arbitrarily select $(m-1)$ vectors from each row of the above matrix, we will get $(m^2 - m)$ linearly independent vectors. Bearing in mind this and the representation (9), we conclude that the following vectors are linearly independent:

$$\nabla \nu_i(\pi) = \left(\frac{\partial \nu_i}{\partial p_{u,v}}(\pi) : u = 1,\dots,m, v = 1,\dots,m-1 \right), \quad i = 1,\dots,m^2 - m.$$

Thus the Jacobian

$$\left\| \left(\frac{\partial \nu_i}{\partial p_{u,v}}(\pi) : u = 1,\dots,m, v = 1,\dots,m-1, i = 1,\dots,m^2 - m \right) \right\|$$

is not equal to 0. Furthermore, the likelihood function given by (8) has a representation like the one given by (1), where $n = m^2 - m$ and $k = m$. For any fixed linear transformation T: $R^{m^2} \to R^{m^2 - m}$, one can explicitly find the function $\varphi(\theta)$ in the representation given by (1), where

$$\theta = (\theta_1,\dots,\theta_{m^2-m}) = (\nu_1(\pi),\dots,\nu_{m^2-m}(\pi)).$$

From Proposition 1.1 we get that, as $s \uparrow +\infty$,

$$T(N(\tau_s))/s \to \nabla\varphi(\theta), \quad P_\pi\text{-in probability},$$

for each $\pi \in \Pi_1$, where $\nabla\varphi(\theta) = (\varphi_1'(\theta),\dots,\varphi_{m^2-m}'(\theta))$. Also from Proposition 1.2 we have the following.

PROPOSITION 2.1. *As* $s \uparrow +\infty$,

$$\frac{T(N(\tau_s)) - s\,\nabla\varphi(\theta)}{\sqrt{s}} \to N(0,\Sigma), \quad P_\pi\text{-weakly}, \pi \in \Pi_1, \tag{10}$$

where $\Sigma = (\partial^2\varphi(\theta)/\partial\theta_i\,\partial\theta_j)_{i,j=1}^{m^2-m}$.

Of course there exists a linear transformation G: $R^{m^2-m} \to R^{m^2}$, which is a generalized inverse of T whose rank is $m^2 - m$. For a given T, the matrix corresponding to G can be explicitly given. In view of Lemma 1.1 we have the following.

PROPOSITION 2.2. *As* $s \uparrow +\infty$,

$$\frac{N(\tau_s) - sG(\nabla\varphi(\theta))}{\sqrt{s}} \to N(0, G'\Sigma G'^T) \quad P_\pi\text{-weakly}, \pi \in \Pi_1, \tag{11}$$

where

$$G' = \left(\frac{\partial G_i}{\partial x_j} (\nabla \varphi(\theta)) : i = 1, \dots, m^2, \qquad j = 1, \dots, m^2 - m \right).$$

Of course it is well known that $(N_{i,j}(\tau_s)/N_{i\cdot}(\tau_s)$: $i, j = 1, \dots, m)$ is the sequential nonparametric maximum likelihood estimator of the transition matrix of the chain. Let H: $R^{m^2} \to R^{m^2}$ be defined as follows:

$$H_{i,j}\big((x_{k,l})_{k,l=1}^m\big) = \frac{x_{i,j}}{\sum_{k=1}^m x_{i,k}}, \qquad i, j = 1, \dots, m.$$

The above sequential maximum likelihood estimator equals $H(N(\tau_s))$. Once again using Lemma 1.1 we get that, as $s \uparrow +\infty$,

$$(12) \qquad \frac{H(N(\tau_s)) - s\pi}{\sqrt{s}} \to N(0, H'G'\Sigma G'^T H'^T), \quad P_\pi\text{-weakly}, \ \pi \in \Pi_1,$$

where $H' = ((\partial H_i/\partial x_j)(G(\nabla\varphi(\theta)))$: $i, j = 1, \dots, m^2)$, bearing in mind that $H(G(\nabla\varphi(\theta))) = \pi$. The latter follows, for example, from the fact that, as $s \uparrow +\infty$,

$$H(N(\tau_s)) \to H(G(\nabla\varphi(\theta))) \quad \text{and} \quad H(N(\tau_s)) \to \pi,$$

P_π-in probability, $\pi \in \Pi_1$.

In the parametric case, that is, when the transition matrix π depends on a k-dimensional unknown parameter ($k \le m^2 - m$), say, $\vartheta = (\vartheta_1, \dots, \vartheta_k)$, the rank of the matrix $(\partial p_{i,j}(\vartheta)/\partial\vartheta_l$: $i, j = 1, \dots, m$; $l = 1, \dots, k)$ equals k, and moreover the functions $p_{i,j}(\vartheta)$ are twice continuously differentiable, the asymptotic normality of the sequential maximum likelihood estimator of ϑ is obtained from the asymptotic normality of the sequential nonparametric maximum likelihood estimator in a similar way as was done above for the latter estimator using the asymptotic normality of the canonical statistic $N(\tau_s)$.

3. Finite-state Markov chains with continuous time. It should not be confusing for the reader if we use the same notation in the continuous-time case, as in the preceding section, for the corresponding quantities.

Let $(Z(t))_{t \ge 0}$ be a homogeneous and ergodic m-state Markov chain in continuous time. Once again without loss of generality we assume that the chain starts from state 1 with probability 1. The likelihood function is given by

$$(13) \qquad L(\pi, t) = \exp\left(\sum_{i,j=1,\, i \ne j}^{m} N_{i,j}(t) \ln \lambda_{i,j} - \sum_{i=1}^{m} S_i(t) \lambda_{i,i} \right),$$

where π is the transition intensity matrix $(\lambda_{i,j})_{i,j=1}^m$, $N_{i,j}(t)$ is the number of transitions from state i to state j in the time interval $[0, t]$ and $S_i(t)$ is the sojourn time at state i in the time interval $[0, t]$. The following type of stopping time is considered:

$$\tau_s := \inf\left(t: \sum_{i,j=1,\, i\neq j}^{m} a_{i,j} N_{i,j}(t) + \sum_{i=1}^{m} a_{i,i} S_i(t) > s\right), \qquad s > 0,$$

where $a_{i,j}$, $i, j = 1, \dots, m$, are as in the preceding section. Of course in the particular case when $a_{i,j} = 0$ if $i \neq j$ and $a_{i,i} = 1$, $i = 1, \dots, m$, we get the fixed-time case. Let

$$\Pi_1 := \left(\pi : E_\pi\left(\sum_{i,j=1,\, i\neq j}^{m} a_{i,j} N_{i,j}(\eta_1) + \sum_{i=1}^{m} a_{i,i} S_i(\eta_1)\right) > 0\right),$$

where η_1 is the first entry time to state 1 after the chain has left it for the first time. Criterion Π_1 is practicable (cf. Remark 2.1). In view of Jacod and Shiryaev [(1987), Theorem 3.4, page 153] the sequential version of (13) takes the following form:

$$(14) \quad L(\pi, \tau_s) = \exp\left(\sum_{i,j=1,\, i\neq j}^{m} N_{i,j}(\tau_s) \ln \lambda_{i,j} - \sum_{i=1}^{m} S_i(\tau_s)\lambda_{i,i}\right), \qquad \pi \in \Pi_1.$$

Deriving the asymptotic normality of all $(N(\tau_s), S(\tau_s))$ and the corresponding sequential maximum likelihood estimators works in exactly the same way as was done for the discrete-time case and is therefore omitted. We only remark that the equalities given in (7) now take the form

$$N_{\cdot i}(\tau_s) - N_{i\cdot}(\tau_s) = R_i(s), \qquad i = 1, \dots, m-1,$$

$$\sum_{i,j=1,\, i\neq j}^{m} a_{i,j} N_{i,j}(\tau_s) + \sum_{i=1}^{m} a_{i,i} S_i(\tau_s) - s = -R_m(s),$$

where $R_m(s)$ is suitably changed and

$$N_{i\cdot}(\tau_s) = \sum_{j=1,\, i\neq j}^{m} N_{i,j}(\tau_s), \qquad N_{\cdot i}(\tau_s) = \sum_{j=1,\, i\neq j}^{m} N_{j,i}(\tau_s).$$

Of course condition (2) is satisfied for $R_i(s)$, $i = 1, \dots, m-1$, and it is easy to see that condition (3) is satisfied for $R_m(s)$ if we take

$$C_m(\cdot) = \sum_{i,j=1,\, i\neq j}^{m} |a_{i,j}| + \sum_{i=1}^{m} |a_{i,i} Y_i|,$$

where $Y_1, \dots, Y_m$ are independent random variables such that for each i the density of Y_i is the classical exponential one with parameter $\lambda_{i,i}$.

Also, when applying the linear transformation T: $R^{m^2} \to R^{m^2 - m}$ on $(N(\tau_s), S(\tau_s))$, one should proceed as if $N_{i,i}(\tau_s)$ from the discrete-time case were replaced by $S_i(\tau_s)$. The function H: $R^{m^2} \to R^{m^2}$ from the preceding section then takes the form

$$H_{i,j}(x_{i,j})_{i,j=1}^{m} = \frac{x_{i,j}}{x_{i,i}}, \qquad i, j = 1, \dots, m.$$

The functions $\nu_i(\pi)$, $i = 1, \dots, m^2 - m$, $\mu_i(\pi)$, $i = 1, \dots, m$, in representation (8) are the same linear functions, but now of $\ln \lambda_{i,j}$, $i, j = 1, \dots, m$ and $i \neq j$, and $\lambda_{i,i}$, $i = 1, \dots, m$.

4. Markov renewal processes. Let $(Z(t), A(t))_{t \geq 0}$ be a Markov renewal process, where t is a discrete time parameter, $Z(t)$ is the embedded m-state Markov chain and $A(t)$ is the so-called additive part of the Markov renewal process. For $A(t)$ the following condition is assumed:

$$0 = A(0) < A(1) < \cdots .$$

For the definition of a Markov renewal process see, for example, Karr (1991) or Prabhu (1991). Furthermore, we assume that for each i, $i = 1, \dots, m$, the conditional density with respect to Lebesgue measure of $A(t+1) - A(t)$ given $Z(t) = i$ is given by

$$h(x)\exp(\lambda_i x - f_i(\lambda_i)), \tag{15}$$

where λ_i is a real parameter, $\lambda_i \in \Lambda_i \subset R$, and $h(x)$ is a Borel function. In other words the sojourn times have distributions belonging to one-dimensional exponential families. We assume that Λ_i is the interior of the natural parameter space of the respective exponential family. Without loss of generality we shall further assume that $Z(0) = 1$ with probability 1. Also we remark that we shall use the same notation as those used in the preceding sections for the same or similar quantities; it should not be confusing for the reader, and, moreover, the explanations become clearer when we refer to the considerations made in Section 2.

It is easy to find the likelihood function corresponding to the observation of a single realization of the process up to time t. It has the following form:

$$L(\pi, \lambda, t) = \exp\left(\sum_{i,j=1}^{m} N_{i,j}(t)\big(\ln p_{i,j} - f_i(\lambda_i)\big) + \sum_{i=1}^{m} S_i(t)\lambda_i \right), \tag{16}$$

where $\pi = (p_{i,j})_{i,j=1}^{m}$ is the transition probability matrix of the embedded Markov chain $Z(t)$, $\lambda = (\lambda_1, \dots, \lambda_m)$, $N_{i,j}(t)$ is the number of the one-step transitions from state i to state j in the time interval $[0, t]$ and $S_i(t)$ is the sojourn time at state i in the time interval $[0, t]$.

We consider the following type of stopping time:

$$\tau_s := \inf\left(t \colon \sum_{i,j=1}^{m} a_{i,j} N_{i,j}(t) + \sum_{i=1}^{m} b_i S_i(t) > s \right), \qquad s > 0, \tag{17}$$

where b_i and $a_{i,j}$, $i, j = 1, \dots, m$, are real numbers not all equal to 0.

The particular case when $a_{i,j} = 1$ for each i, j and $b_i = 0$ for each i corresponds to the fixed sample size case.

Let

$$\Pi_1 \times \Lambda := \left((\pi, \lambda) \colon E_{\pi,\lambda}\left(\sum_{i,j=1}^{m} a_{i,j} N_{i,j}(\eta_1) + \sum_{i=1}^{m} b_i S_i(\eta_1) \right) > 0 \right),$$

where η_1 is the first entry time to state 1 after the embedded Markov chain $Z(t)$ has left it for the first time. Criterion $\Pi_1 \times \Lambda$ is practicable (cf. Remark 2.1 and bear in mind that η_1 is one of the stopping times described in Proposition 4.1). Likewise (cf. Section 2), it is easy to show that

$$P_{\pi, \lambda}(\tau_s < +\infty) = 1$$

for each $(\pi, \lambda) \in \Pi_1 \times \Lambda$. Then, in view of Jacod and Shiryaev [(1987), Theorem 3.4, page 153], the sequential version of (16) takes the following form:

$$L(\pi, \lambda, \tau_s) = \exp\left(\sum_{i,j=1}^{m} N_{i,j}(\tau_s)\big(\ln p_{i,j} - f_i(\lambda_i)\big) + \sum_{i=1}^{m} S_i(\tau_s)\lambda_i \right), \quad (\pi, \lambda) \in \Pi_1 \times \Lambda. \tag{18}$$

First we shall show that there are stopping times such that the corresponding sequential likelihood functions belong to noncurved exponential families. Actually, these are the same stopping times which have the abovementioned property in the case of finite-state Markov chains with discrete time [cf. Stefanov (1991)].

Proposition 4.1. *Let I be the state space of the embedded Markov chain $Z(t)$. Then, for each $i \in I$ and each $J \subseteq I$ such that $\sum_{j \in J} p_{j,i} > 0$, the stopping times $\tau_J^i(s)$ given by*

$$\tau_J^i(s) = \inf\left(t\colon \sum_{j \in J} N_{j,i}(t) = s \right), \qquad s = 1, 2, \dots,$$

have the property that their corresponding likelihood functions $L(\pi, \lambda, \tau_J^i(s))$ belong to noncurved exponential families of order equal to the dimension of the parameter (π, λ).

Proof. For simplicity we shall consider the case when the dimension of the parameter (π, λ) is m^2, that is, $p_{i,j} \neq 0$ for each i, j. The other cases when $p_{i,j} = 0$ for some pairs (i, j) work similarly and are therefore omitted.

It is easy to see that the following linear dependencies for the components of $N(\tau_J^i(s))$ hold with probability 1:

$$\sum_{j \in J} N_{j,i}\big(\tau_J^i(s)\big) - s = 0,$$

$$N_{\cdot k}\big(\tau_J^i(s)\big) - N_{k \cdot}\big(\tau_J^i(s)\big) + \mathbf{1}_{(k)}\big(Z(0)\big) - \mathbf{1}_{(k)}\big(Z\big(\tau_J^i(s)\big)\big) = 0, \quad k = 1, \dots, m-1. \tag{19}$$

Of course $\mathbf{1}_{(k)}(Z(0))$ equals 1 if $k = 1$, and 0 otherwise, whereas $\mathbf{1}_{(k)}(Z(\tau_J^i(s)))$ equals 1 if $k = i$, and 0 otherwise. It is easy to see that the rank of the above linear system is m, and, furthermore, in view of Stefanov [(1991), Proposition 1] we have that there are $(m^2 - m)$ components of $(N_{i,j}(\tau_J^i(s)), i, j = 1, \dots, m)$ which are linearly independent. By the assumption that $A(t+1) - A(t)$ have conditional distributions, given $Z(t)$, which are absolutely continuous w.r.t. Lebesgue measure, one can easily see that $S_1(\tau_J^i(s)),\dots,S_m(\tau_J^i(s))$ are linearly independent with the above (m^2-m) linearly independent components of $N(\tau_J^i(s))$. Of course the components of the canonical parameter, that is, $(\ln p_{i,j}-f_i(\lambda_i))$ and λ_i, $i,j=1,\dots,m$, are linearly independent. Therefore, the minimal sufficient statistic of the model considered for $\tau_J^i(s)$ has dimension equal to m^2, and subsequently we get that the family (18) for $\tau_s=\tau_J^i(s)$ is a noncurved exponential one of order m^2, that is, of order which equals the dimension of the parameter (π,λ).

This completes the proof of Proposition 4.1. □

Next, as in Section 2, we define the random variables $R_1(s),\dots,R_m(s)$ as follows:

$$(20)\qquad \begin{aligned} R_i(s) &= \mathbf{1}_{(i)}(Z(\tau_s)) - \mathbf{1}_{(i)}(Z(0)), \qquad i=1,2,\dots,m-1,\\ R_m(s) &= s - \sum_{i,j=1}^m a_{i,j}N_{i,j}(\tau_s) - \sum_{i=1}^m b_iS_i(\tau_s). \end{aligned}$$

Likewise we have, for each $(\pi,\lambda)\in\Pi_1\times\Lambda_\cdot$, $P_{\pi,\lambda}$-a.s.,

$$(21)\qquad \begin{aligned} N_{\cdot i}(\tau_s) - N_{i\cdot}(\tau_s) &= R_i(s), \qquad i=1,2,\dots,m-1\\ \sum_{i,j=1}^m a_{i,j}N_{i,j}(\tau_s) + \sum_{i=1}^m b_iS_i(\tau_s) - s &= -R_m(s). \end{aligned}$$

In view of (21) the likelihood function given by (18) can be represented as follows:

$$(22)\qquad \exp\left(\sum_{i=1}^{m^2}\nu_i(\pi,\lambda)T_i(N(\tau_s),S(\tau_s)) + \sum_{i=1}^m \mu_i(\pi,\lambda)R_i(s) - \mu_{m+1}(\pi,\lambda)s\right)$$

where $\nu_i(\pi,\lambda)$, $i=1,\dots,m^2$ and $\mu_i(\pi,\lambda)$, $i=1,\dots,m$ are certain linear functions of $(\ln p_{i,j}-f_i(\lambda_i))$, $i,j=1,\dots,m$ and λ_i, $i=1,\dots,m$. $T(\cdot)=(T_1(\cdot),\dots,T_{m^2}(\cdot))$: $R^{m^2}\times R^m\to R^{m^2}$ is a linear transformation similar to that introduced in the previous sections, which is obtained using the following linear system:

$$(23)\qquad \begin{aligned} x_{\cdot i} - x_{i\cdot} &= 0, \qquad i=1,\dots,m-1,\\ \sum_{i,j=1,\,i\neq j}^m a_{i,j}x_{i,j} + \sum_{i=1}^m b_iy_i - s &= 0, \end{aligned}$$

where $((x_{i,j})_{i,j=1}^m,(y_i)_{i=1}^m)\in R^{m^2}\times R^m$ and $x_{\cdot i}=\sum_{j=1}^m x_{j,i}$, $x_{i\cdot}=\sum_{j=1}^m x_{i,j}$. Of course $\mu_{m+1}(\pi,\lambda)$ can be explicitly found for any fixed linear transformation T. Also, condition (2) is satisfied for $R_i(s)$, $i=1,\dots,m-1$, and it is easy to see that condition (3) is satisfied for R_m if we take

$$C_m(\cdot) = \sum_{i,j=1}^m |a_{i,j}| + \sum_{i=1}^m |b_iY_i|,$$

where $Y_1, \dots, Y_m$ are independent random variables such that for each i the density of Y_i is given by (15).

The proof that (22) is of the type given by (1) follows arguments similar to those used in Section 2 for the finite-state Markov chain with discrete time, and it is therefore omitted. Of course likewise there exists a linear transformation $G: R^{m^2} \to R^{m^2} \times R^m$ which is a generalized inverse to the linear transformation T and the rank of G is m^2. Also let the function $H: R^{m^2} \times R^m \to R^{m^2} \times R^m$ be defined as follows:

$$H_{i,j}\big((x_{k,l})_{k,l=1}^{m}, (y_k)_{k=1}^{m}\big) = \frac{x_{i,j}}{\sum_{k=1}^{m} x_{i,k}}, \qquad i,j = 1,\dots,m,$$

for the first m^2 components of H, and let

$$H_i\big((x_{k,l})_{k,l=1}^{m}, (y_k)_{k=1}^{m}\big) = \frac{y_i}{\sum_{k=1}^{m} x_{i,k}}, \qquad i = 1,\dots,m,$$

for the remaining m components of H. It is easy to see that $H(N(\tau_s), S(\tau_s))$ is the maximum likelihood estimator of $(\pi, f'(\lambda))$, where $f'(\lambda) = (f_1'(\lambda_1), \dots, f_m'(\lambda_m))$ and $f_i'(\lambda_i) = df(\lambda_i)/d\lambda_i$.

Then finally we get, as in Section 2, the following results about the asymptotic normality of $(N(\tau_s), S(\tau_s))$ and the maximum likelihood estimator of $(\pi, f(\lambda))$.

PROPOSITION 4.2. *As* $s \uparrow +\infty$,

$$\frac{(N(\tau_s), S(\tau_s)) - sG(\nabla\varphi(\theta))}{\sqrt{s}} \to N(0, G'\Sigma G'^T),$$

$P_{\pi,\lambda}$-*weakly*, $(\pi, \lambda) \in \Pi_1 \times \Lambda$, *where*

$$G' = \left(\frac{\partial G_i}{\partial x_j}(\nabla\varphi(\theta)) : i = 1,\dots,m^2 + m, \qquad j = 1,\dots,m^2\right),$$

$\varphi(\theta)$ *is the function in the representation* (1) *of* (22) *and* Σ *was defined in Section* 1.

PROPOSITION 4.3. *As* $s \uparrow +\infty$,

$$\frac{H(N(\tau_s), S(\tau_s)) - s(\pi, f'(\lambda))}{\sqrt{s}} \to N(0, H'G'\Sigma G'^T H'^T),$$

$P_{\pi,\lambda}$-*weakly*, $(\pi, \lambda) \in \Pi_1 \times \Lambda$, *where*

$$H' = \left(\frac{\partial H_i}{\partial x_j}(G(\nabla\varphi(\theta))) : i,j = 1,\dots,m^2 + m\right).$$

The asymptotic normality of the sequential maximum likelihood estimator of a smooth function of the parameters (π, λ) is obtained by applying Lemma 1.1 and Proposition 4.3 (cf. the last comments in Section 2). Also the results presented in this section are easily extendable to more general models, in which $A(t)$ does not satisfy the monotonicity condition given at the beginning of this section and the dominating measure in (15) is not necessarily Lebesgue.

5. Markov-additive processes. We shall be concerned with Markov-additive processes whose embedded Markov chains in continuous time have a finite number of states and are ergodic. Denote by I, $I = (1, 2, \dots, m)$, the space of the states of the chain. A Markov-additive process $(Z(t), A(t))_{t \geq 0}$ (the time parameter t is continuous) is a two-dimensional Markov process on the state space $I \times R$, satisfying the following properties [see Prabhu (1991)]:

1. For $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, $n > 2$, the increments $A(t_1) - A(0), A(t_2) - A(t_1), \dots, A(t_n) - A(t_{n-1})$ are conditionally independent given $Z(0), Z(t_1), \dots, Z(t_n)$.
2. The conditional distribution of $A(t_p) - A(t_{p-1})$, given $Z(t_{p-1}) = i$ and $Z(t_p) = j$, depends only on $t_p - t_{p-1}$ and i and j.

Let $\pi = (\lambda_{i,j})_{i,j=1}^m$ be the transition intensity matrix of the embedded m-state Markov chain $Z(t)$. We assume that the conditional density of $A(t) - A(s)$, given $Z(u) = i$ for all $u \in [s, t]$, is given by

$$\exp(\vartheta_i x - f_i(\vartheta_i)(t - s)), \tag{24}$$

with respect to a σ-finite measure which may depend on the state i in general, and ϑ_i is a real parameter, $\vartheta_i \in \Xi_i \subset R$. We assume that Ξ_i is the interior of the natural parameter space of the respective exponential family given by (24). Without loss of generality we assume $Z(0) = 1$ with probability 1.

Applying the general theory about absolute continuity of measures associated with random processes, developed in Gihman and Skorohod [(1974), Chapter 7, pages 440–443], it is easy to find the likelihood function corresponding to the observation of a single realization of the process up to time t. It has the following form:

$$\begin{aligned} L(\pi, \vartheta, t) = \exp\Bigg(& \sum_{i,j=1, i \neq j}^{m} N_{i,j}(t) \ln \lambda_{i,j} \\ & - \sum_{i=1}^{m} S_i(t)(\lambda_{i,i} + f_i(\vartheta_i)) + \sum_{i=1}^{m} A_i(t)\vartheta_i \Bigg), \end{aligned} \tag{25}$$

where $\pi = (\lambda_{i,j})_{i,j=1}^m$, $\vartheta = (\vartheta_1, \dots, \vartheta_m)'$, $N_{i,j}(t)$ is the number of the transitions from state i to state j of the Markov process $Z(s)$ in the time interval $[0, t]$, $S_i(t)$ is the sojourn time at state i of the process $Z(s)$ and

$$\begin{aligned} A_i(t) = \sum_{n=1}^{\infty} \Big(& (A(\eta_n^*(i)) - A(\eta_n(i)))\mathbf{1}_{[0, +\infty)}(t - \eta_n^*(i)) \\ & + (A(t) - A(\eta_n(i)))\mathbf{1}_{(\eta_n(i), \eta_n^*(i))}(t) \Big), \end{aligned}$$

where $\eta_n(i)$ is the nth consecutive time of first entrance of $Z(s)$ to state i, and $\eta_n^*(i)$ is the time of exit from state i after $\eta_n(i)$.

Once again, as in the preceding sections, the stopping times we consider are first crossing times of levels by linear combinations of the components of the canonical statistic $(N(t), S(t), \tilde{A}(t))$, where $S(t) = (S_1(t), \dots, S_m(t))$ and $\tilde{A}(t)) = (A_1(t), \dots, A_m(t))$, that is,

$$\tau_s := \inf\left(t\colon \sum_{i,j=1,\, i\neq j}^{m} a_{i,j}N_{i,j}(t) + \sum_{i=1}^{m} a_{i,i}S_i(t) + \sum_{i=1}^{m} b_i A_i(t) > s\right), \qquad s > 0,$$

where $a_{i,j}$ and b_i, $i, j = 1, \dots, m$, are real numbers not all equal to 0. Of course in the particular case when $a_{i,j} = 0$ for each i, j such that $i \neq j$ and $a_{i,i} = 1$, $b_i = 0$ for each i, $i = 1, \dots, m$, we get the case of fixed sample size, that is, $\tau_s = s$ with probability 1.

Also, it is easy to show (cf. Section 2) that, for each $(\pi, \vartheta) \in \Pi_1 \times \Xi$,

$$P_{\pi,\vartheta}(\tau_s < +\infty) = 1,$$

where

$$\Pi_1 \times \Xi = \left((\pi,\vartheta)\colon E_{\pi,\vartheta}\left(\sum_{i,j=1,\, i\neq j}^{m} a_{i,j}N_{i,j}(\eta_1) + \sum_{i=1}^{m} a_{i,i}S_i(\eta_1) + \sum_{i=1}^{m} b_i A_i(\eta_1)\right) > 0\right)$$

and η_1 is as used in the preceding sections. Criterion $\Pi_1 \times \Xi$ is practicable (cf. Remark 2.1 and bear in mind that η_1 is one of the stopping times described in Proposition 5.1). In view of Jacod and Shiryaev [(1987), Theorem 3.4, page 153], the sequential version of (25) has the following form:

$$L(\pi,\vartheta,\tau_s) = \exp\left(\sum_{i,j=1,\, i\neq j}^{m} N_{i,j}(\tau_s)\ln\lambda_{i,j} - \sum_{i=1}^{m} S_i(\tau_s)(\lambda_{i,i} + f_i(\vartheta_i)) + \sum_{i=1}^{m} A_i(\tau_s)\vartheta_i\right), \tag{26}$$
$$(\pi,\vartheta) \in \Pi_1 \times \Xi.$$

There are also stopping times, in the case of Markov-additive processes, which reduce the curved exponential family in general given by (26), to noncurved exponential ones of order equal to the dimension of the parameter (π, ϑ). The following proposition describes them.

PROPOSITION 5.1. *For each $i \in I$ and each $J \subseteq I$ such that $\sum_{j \in J} \lambda_{j,i} > 0$, the stopping times $\tau_J^i(s)$, $\tau_i^i(s)$, $\tilde{\tau}_i^i(s)$ and $\tau_{i,J}^i(s)$ given by*

$$\tau_J^i(s) = \inf\left(t\colon \sum_{j \in J} N_{j,i}(t) = s\right), \qquad s = 1, 2, \dots,$$

$$\tau_i^i(s) = \inf(t\colon S_i(t) = s), \qquad s > 0,$$

$$\tilde{\tau}_i^i(s) = \inf(t\colon A_i(t) = s),$$

$$\tau_{i,J}^i(s) = \inf\left(t\colon A_i(t) + \sum_{j \in J} N_{j,i}(t) = s\right), \qquad s = 1, 2, \dots,$$

where additionally we assume that $P_{\pi,\vartheta}(\tilde{\tau}_i^i(s) < +\infty) = 1$ and $P_{\pi,\vartheta}(\tau_{i,J}^i(s) < +\infty) = 1$ for each (π, ϑ), have the property that their corresponding likelihood functions $L(\pi, \vartheta, \tau_J^i(s))$, $L(\pi, \vartheta, \tau_i^i(s))$, $L(\pi, \vartheta, \tilde{\tau}_i^i(s))$ and $L(\pi, \vartheta, \tau_{i,J}^i(s))$ belong to noncurved exponential families of order equal to the dimension of the parameter (π, ϑ).

REMARK 5.1. The condition $P_{\pi,\vartheta}(\tilde{\tau}_i^i(s) < +\infty) = 1$ is satisfied, for example, if the process $A_i(t)$ is nonnegative with continuous trajectories and $A_i(t) \uparrow +\infty$ as $t \uparrow +\infty$, or if the process $A_i(t)$ is a Poisson process; in the latter case s must be a natural number. If $A_i(t)$, $i = 1, \dots, m$, are Poisson processes, then the Markov-additive process is called a Markov–Poisson process [see Prabhu (1991)]. The condition $P_{\pi,\vartheta}(\tau_{i,J}^i(s) < +\infty) = 1$ is satisfied, for example, in the case of a Markov–Poisson process. This follows from the fact that jumps in $Z(t)$ and $A(t)$ cannot occur simultaneously [see Prabhu (1991)].

PROOF OF PROPOSITION 5.1. For simplicity we shall consider the case when the dimension of the parameter (π, ϑ) is m^2, that is, $\lambda_{i,j} \neq 0$ for each i, j. The remaining cases when $\lambda_{i,j} = 0$ for some pairs of (i, j) work similarly and are therefore omitted. One should just replace $N_{i,j}(\cdot)$ by 0 if the respective $\lambda_{i,j} = 0$. The proof for $\tau_J^i(s)$ and $\tau_i^i(s)$ is an easy consequence (cf. the proof of Proposition 4.1) of Proposition 2 of Stefanov (1991) and is therefore omitted too.

Consider the stopping time $\tilde{\tau}_i^i(s)$. Of course, $Z(\tilde{\tau}_i^i(s)) = i$ with probability 1 and moreover the following equalities hold with probability 1:

$$A_i(\tilde{\tau}_i^i(s)) = s,$$

$$N_{\cdot k}(\tilde{\tau}_i^i(s)) - N_{k\cdot}(\tilde{\tau}_i^i(s)) + \mathbf{1}_{(k)}(Z(0)) - \mathbf{1}_{(k)}(Z(\tilde{\tau}_i^i(s))) = 0, \qquad k = 1, \dots, m-1.$$

Of course $\mathbf{1}_{(k)}(Z(0))$ equals 1 if $k = 1$, and 0 otherwise, whereas $\mathbf{1}_{(k)}(Z(\tilde{\tau}_i^i(s)))$ equals 1 if $k = i$, and 0 otherwise.

The rank of the sublinear system consisting of the last $(m-1)$ equations above is manifestly $(m-1)$. Thus, there are at least $(m-1)$ linear dependencies between the components of $N(\tau_i^i(s))$. We shall show that there are exactly $(m-1)$. Let

$$B_n = \left(N_{\cdot i}\left(\tilde{\tau}_i^i(s)\right) = n\right).$$

Of course $\sum_{k=0}^{\infty} P_{\pi,\vartheta}(B_k) = 1$ and, for each k and each pair (i, j) with probability 1,

$$N_{i,j}\left(\tilde{\tau}_i^i(s)\right)(\omega)\mathbf{1}_{B_n}(\omega) = N_{i,j}\left(\tau_I^i(n)\right)(\omega)\mathbf{1}_{B_n}(\omega), \tag{27}$$

where we recall that I means the state space of the embedded chain $Z(t)$. However, in view of Stefanov [(1991), Proposition 2] it follows that there are exactly m linear dependencies between the components of $N(\tau_I^i(n))$. Also the first linear dependence, that is,

$$N_{\cdot i}\left(\tau_I^i(n)\right) = n,$$

is different for each n, and the remaining $(m-1)$ linear dependencies are the same for each n. Thus, in view of these considerations and (27) we conclude that there are exactly $(m-1)$ linear dependencies between the components of $N(\tilde{\tau}_i^i(s))$.

It is easy to see that there is no linear dependence involving components from both $(N(\tilde{\tau}_i^i(s)), S(\tilde{\tau}_i^i(s)))$ and $\tilde{A}(\tilde{\tau}_i^i(s))$. Actually, it follows from the conditional independence of the increments of $A(t)$. This was defined precisely in property 1 at the beginning of this section. It implies enough control over the values assumed by the components of $\tilde{A}(\tilde{\tau}_i^i(s))$ to show that each linear dependence of the type mentioned above is violated with positive probability.

Of course the components of the canonical parameter $(\ln \lambda_{i,j}, (\lambda_{i,i} - f_i(\vartheta_i)), \vartheta_i,\ i \neq j,\ i, j = 1, \dots, m)$ in (26) are linearly independent. In view of the equality

$$A_i\left(\tilde{\tau}_i^i(s)\right) = s$$

and the easily seen fact that $A_1(\tilde{\tau}_i^i(s)), \dots, A_{i-1}(\tilde{\tau}_i^i(s)),\ A_{i+1}(\tilde{\tau}_i^i(s)), \dots, A_m(\tilde{\tau}_i^i(s))$ are linearly independent, we conclude that the dimension of the minimal sufficient statistic of the considered model for $\tilde{\tau}_i^i(s)$ equals m^2. Thus, we get that the family (26) for $\tau_s = \tilde{\tau}_i^i(s)$ is a noncurved exponential family of order m^2, that is, the order equals the dimension of the parameter (π, ϑ). The proof for $\tau_{i,J}^i(s)$ follows similar arguments to those above and is omitted.

The proof of Proposition 5.1 is complete. □

Next the random variables $R_1(s), \dots, R_m(s)$ and the linear transformations T and G are defined likewise (cf. the previous section). Likewise, condition (2) or (3) is satisfied for them. Also, the function H: $R^{m^2} \times R^m \to R^{m^2} \times R^m$ is defined as follows:

$$H_{i,j}\left((x_{k,l})_{k,l=1}^m, (y_k)_{k=1}^m\right) = \frac{x_{i,j}}{x_{i,i}}, \qquad i, j = 1, \dots, m,$$

for the first m^2 components of H and

$$H_i\big((x_{k,l})_{k,l=1}^{m},(y_k)_{k=1}^{m}\big)=\frac{y_i}{x_{i,i}},\qquad i=1,\dots,m,$$

for the remaining components of H. It is easy to see that $H(N(\tau_s), S(\tau_s), \tilde{A}(\tau_s))$ is the maximum likelihood estimator of $(\pi, f'(\vartheta))$, where $f'(\vartheta) = (f_1'(\vartheta_1), \dots, f_m'(\vartheta_m))$ and $f_i'(\vartheta_i) = df_i(\vartheta_i)/d\vartheta_i$.

Likewise the following propositions hold true.

PROPOSITION 5.2. *As* $s \uparrow +\infty$,

$$\frac{\big(N(\tau_s), S(\tau_s), \tilde{A}(\tau_s)\big) - sG(\nabla\varphi(\theta))}{\sqrt{s}} \to N(0, G'\Sigma G'^T),$$

$P_{\pi,\vartheta}$*-weakly*, $(\pi, \vartheta) \in \Pi_1 \times \Xi$, *where*

$$G'=\left(\frac{\partial G_i}{\partial x_j}(\nabla\varphi(\theta)) : i=1,\dots,m^2+m,\qquad j=1,\dots,m^2\right),$$

$\varphi(\theta)$ *is the one from representation* (1) *of* (26) *and* Σ *was defined in Section* 1.

PROPOSITION 5.3. *As* $s \uparrow +\infty$,

$$\frac{H\big(N(\tau_s), S(\tau_s)\tilde{A}(\tau_s)\big) - s(\pi, f'(\vartheta))}{\sqrt{s}} \to N(0, H'G'\Sigma G'^T H'^T),$$

$P_{\pi,\vartheta}$*-weakly*, $(\pi, \vartheta) \in \Pi_1 \times \Xi$, *where*

$$H'=\left(\frac{\partial H_i}{\partial x_j}(G(\nabla\varphi(\theta))) : i,j=1,\dots,m^2+m\right).$$

The asymptotic normality of the sequential maximum likelihood estimators of smooth functions of the parameter (π, ϑ) is obtained likewise by applying Lemma 1.1 (cf. Section 1).

6. Further remarks. Observe that, for τ_s as defined in Sections 2–5, we have

$$\tau_s/s \to \text{const}(\theta) > 0,\quad P_\theta\text{-in probability as } s \uparrow +\infty.$$

This follows from Proposition 1.1 and the fact that τ_s is a linear combination of the components of the corresponding canonical statistic in all models considered. Therefore, bearing in mind the results concerning random time changes in the central limit theorem for random processes [see Billingsley (1968)], we conclude that the limiting normal distributions of the respective quantities considered, for all stopping times τ_s, in any particular model, are the same up to a multiplication of the limiting random vectors by a constant. We may use this fact to obtain some useful information about the limiting normal distribution; for example, we may see whether the latter is a product of several independent normal distributions and if so which they are. It is possible to obtain such information by choosing the stopping time τ_s suitably. Note that if $\varphi(\theta)$ in the representation (1) can be split into m parts

$$\varphi(\theta) = \varphi_1(\theta_1, \dots, \theta_{k_1}) + \cdots + \varphi_m(\theta_{k_m+1}, \dots, \theta_n),$$

then the random vectors $(X_1, \dots, X_{k_1}), \dots, (X_{k_m+1}, \dots, X_n)$ are asymptotically independent.

7. Applications to functional limit theory. In this section we explain our method for deriving explicit solutions in functional limit theorems for the processes considered. For the sake of brevity, we exemplify it by two simple examples, namely, of a two-state Markov chain and a Poisson process modulated by a two-state Markov process.

First we discuss briefly the key insights from Serfozo (1975) and Stefanov (1986). Let $(\xi(t))_{t \geq 0}$ be a stochastic process whose trajectories are right continuous and have left limits. Let $(\xi(\tau_k))_{k \geq 0}$ be a process embedded in $\xi(t)$, where $(\tau_k)_{k \geq 0}$ is a sequence of stopping times such that $\tau_0 < \tau_1 < \cdots$ and $\tau_k \to +\infty$ a.s. Define $M_{[kt]}$ and the processes $S_k = (S_k(t))_{t \geq 0}$ and $\tilde{S}_k = (\tilde{S}_k(t))_{t \geq 0}$ as follows:

$$S_k(t) = \frac{(\xi(kt) - \mathscr{A}_k kt)}{\mathscr{B}_k},$$

$$\tilde{S}_k(t) = \frac{(\xi(\tau_{[kt]}) - \mathscr{A}_k \tau_{[kt]})}{\mathscr{B}_k},$$

$$M_{[kt]} = \sup\left(\left| \xi(s) - \xi(\tau_{[kt]}) - \mathscr{A}_k(s - \tau_{[kt]})\right| : \tau_{[kt]} \leq s < \tau_{[kt]+1}\right),$$

where $\mathscr{A}_k$ and $\mathscr{B}_k$ are constants. The following propositions are excerpts from Serfozo's main results [see Serfozo (1975), Theorems 2.1 and 2.2, Corollary 2.3 and Remark 2.4] and are used in deriving our explicit results. Let $\to_m$ mean any, but fixed, mode of convergence selected from a.s., in probability and in distribution. For all convergence results below we assume $k \uparrow +\infty$.

PROPOSITION 7.1. *Suppose $\tau_k/k \to_m a$, $a \neq 0$, and $M_{[k\cdot]}/\mathscr{B}_k \to_m 0$. If $\tilde{S}_k \to_m S(a\,\cdot)$ for some process S, then $S_k \to_m S$.*

PROPOSITION 7.2. *Suppose $\tau_k/k \to a$ a.s., $a \neq 0$, $M_{[k\cdot]}/\mathscr{B}_k \to 0$ a.s., and let K denote a compact subset of the Skorohod space $D = D[0, +\infty)$. If $(\tilde{S}_k(t))_{k \geq 0}$ is a.s. relatively compact with limit points K, then $(S_k(t))_{k \geq 0}$ is a.s. relatively compact with limit points $\{x(a^{-1}\,\cdot): x \in K\}$.*

In other words Serfozo (1975) provides conditions which ensure that limit results for the embedded process imply the same limit results for the original process. The results of Serfozo (1975) cover and unify almost all limit results available for the processes considered in this paper. However, deriving explicitly the limiting process, for example, the variance parameter of the limiting Wiener process in the functional central limit theorem, is a problem. Consequently, no general explicit limit results are available for these processes. Note that Bhattacharya and Waymire [(1990), pages 513–515] and Bhattacharya (1982) provide calculable expressions for the variance parameter in the functional central limit theorem for some Markov processes. The reason for the lack of such general explicit results is that *not enough* has been known about the properties of the sequences of stopping times used, beyond their useful regenerative property. However, the results of Stefanov (1986, 1991) and Propositions 4.1 and 5.1 given above fill that gap for the considered processes. Actually, from Stefanov [(1986), Theorem 1] it follows, in particular, that when using the sequences of stopping times [introduced in Stefanov (1991) and in Propositions 4.1 and 5.1] the embedded process has stationary independent increments, and, moreover, the moment generating function, and consequently the moments, of the increment can be given explicitly. The latter is due to the noncurved exponential structure of the distribution of the increment. This supplies enough information for obtaining explicitly the limiting laws and therefore making the results of Serfozo (1975) applicable in practice. A more detailed discussion follows.

Propositions 1 and 2 of Stefanov (1991) as well as Propositions 4.1 and 5.1 above say that, for suitably chosen stopping times, the respective likelihood functions belong to noncurved exponential families, that is, D_i's in the representation (1) are equal to 0. Of course, the following sequence of stopping times $(\tau(s))_{s\geq 0}$ is common for all these propositions:

$$\tau(s) = \inf\Big(t\colon \sum_{j\in I} N_{j,1}(t) = s\Big), \qquad s = 0, 1, \ldots,$$

if t is discrete, and

$$\tau(s) = \inf\Big(t\colon \sum_{j\in I,\, j\neq 1} N_{j,1}(t) = s\Big), \qquad s = 0, 1, \ldots,$$

if t is continuous, where I is the state space of the corresponding finite-state Markov chain. Without loss of generality we assume that $Z(0) = 1$ a.s. In view of the abovementioned propositions, we obtain the following representation for the sequential likelihood functions for $(\tau(s))_{s\geq 0}$:

$$\exp\Big(\sum_{i=1}^{n} \theta_i X_i(\tau(s)) + \varphi(\theta) s\Big), \tag{28}$$

which is derived explicitly in any particular case.

EXAMPLE 7.1 (Two-state Markov chain). It is straightforward to verify that (28) takes the following form:

$$\exp\bigg(\theta_1 N_{1,2}(\tau(s)) + \theta_2 N_{2,2}(\tau(s)) + s\ln\frac{1-\exp(\theta_1)}{1+\exp(\theta_1)-\exp(\theta_2)}\bigg),$$

where θ_1 and θ_2 are related to the transition probabilities as follows:

$$\theta_1 = \ln \frac{p_{1,2}\, p_{2,1}}{p_{1,1}}, \qquad \theta_2 = \ln p_{2,2}.$$

EXAMPLE 7.2 (Poisson process modulated by a two-state Markov process). This is a Markov-additive process for which [recall (24)]

$$f_1(\vartheta_1) = \exp(\vartheta_1), \qquad f_2(\vartheta_2) = \exp(\vartheta_2),$$

that is, $\vartheta_1 = \ln \mu_1$ and $\vartheta_2 = \ln \mu_2$, where μ_1 and μ_2 are the intensities of the two Poisson processes involved. It is easy to verify that (28) takes the following form:

$$\exp\big(\theta_1 S_1(\tau(s)) + \theta_2 S_2(\tau(s)) + \theta_3 A_1(\tau(s)) + \theta_4 A_2(\tau(s)) \\ + s \ln\big((\theta_1 + \exp(\theta_3))(\theta_2 + \exp(\theta_4))\big)\big),$$

where $S_1(t)$, $S_2(t)$, $A_1(t)$ and $A_2(t)$ were defined in (25), and θ_1, θ_2, θ_3 and θ_4 are related to μ_1, μ_2 and the transition intensity matrix $(\lambda_{i,j})_{i,j=1}^2$ of the embedded two-state Markov process as follows:

$$\begin{aligned} \theta_1 &= -\lambda_{1,2} - \mu_1, & \theta_2 &= -\lambda_{2,1} - \mu_2 \\ \theta_3 &= \ln \mu_1 & \theta_4 &= \ln \mu_2. \end{aligned}$$

From (28), in view of Stefanov [(1986), Theorem 1], it follows that the process $(X(\tau(s)))_{s \ge 0}$ is a process with stationary independent increments, and, moreover, we get explicitly the moment generating function of the increment.

REMARK 7.1. In particular, this presents also an alternative method for deriving the regeneration property of some sequences of stopping times, in this case those whose respective sequential likelihood functions belong to noncurved exponential families.

All known functional and nonfunctional limit results for processes with stationary independent increments are valid for the process $(X(\tau(s)))_{s \ge 0}$. Furthermore, well-known analytic properties of noncurved exponential families [cf. Barndorff-Nielsen (1978), page 114, or Brown (1986)] yield

$$\operatorname{Cov}\big(X_i(\tau(s)), X_j(\tau(s))\big) = -s \frac{\partial^2 \varphi(\theta)}{\partial \theta_i\, \partial \theta_j}. \tag{29}$$

Also, for each $s > 0$ and each $\alpha > 0$,

$$E\big(X_i(\tau(s)) - X_i(\tau(s-1))\big)^{\alpha} < +\infty, \qquad i = 1, 2, \dots, n, \tag{30}$$

$$\frac{\tau(s)}{s} \to E\tau(1) \overset{\text{def}}{=} \delta(\theta) > 0 \quad \text{a.s. as } s \uparrow +\infty. \tag{31}$$

The latter follows from the fact that $\tau(s)$ is a linear function of the components of $X(\tau(s))$ and thus $(\tau(s))_{s\geq 0}$ has stationary independent increments. Define for each k, $k > 1$, the processes $Y_k(t)$ and $\tilde{Y}_k(t)$ as follows:

$$(32)\qquad Y_k(t) = \frac{f(X(kt)) + kt\delta^{-1}(\theta)f(\nabla\varphi(\theta))}{\sqrt{k}},$$

$$(33)\qquad \tilde{Y}_k(t) = \frac{f(X(\tau([kt]))) + \tau([kt])\delta^{-1}(\theta)f(\nabla\varphi(\theta))}{\sqrt{k}},$$

where $X(\cdot)$ is as in (28), f: $R^n \to R$ is a linear function, that is,

$$f(x) = b_1x_1 + b_2x_2 + \cdots + b_nx_n,$$

and $\delta(\theta)$ is defined in (31). For the processes considered in this paper $\tau([kt])$ is a linear function of the components of the minimal sufficient statistic $X_1(\tau([kt])),\dots, X_n(\tau([kt]))$, say,

$$\tau([kt]) = \sum_{i=1}^{n} a_iX_i(\tau([kt])) + a_{n+1}(kt),$$

where $a_{n+1}(kt)$ depends on kt in general. Of course, $\tilde{Y}_k(t)$ can be represented also as follows:

$$(34)\quad \tilde{Y}_k(t) = \frac{\sum_{i=1}^{n}(b_i + a_i\delta^{-1}(\theta)f(\nabla\varphi(\theta)))X_i(\tau([kt])) + a_{n+1}(kt)\delta^{-1}(\theta)f(\nabla\varphi(\theta))}{\sqrt{k}}.$$

Bearing in mind that the expected value of $\tau(s)$ equals $s\delta(\theta)$, it is straightforward to see that the expected value of the numerator in (33) and consequently of the numerator in (34) is equal to 0. Of course, from the functional central limit theorem for processes with stationary independent increments we have that, as $k \to +\infty$, $\tilde{Y}_k$ converges in distribution to a Brownian motion starting at the origin with zero drift and diffusion coefficient σ^2, such that

$$\sigma^2 = -\sum_{i,j=1}^{n} c_ic_j\frac{\partial^2\varphi(\theta)}{\partial\theta_i\,\partial\theta_j},$$

where

$$c_i = b_i + a_i\delta^{-1}(\theta)f(\nabla\varphi(\theta)).$$

Condition (31) guarantees that the assumptions for τ_k in Propositions 7.1 and 7.2 are satisfied. In the case of finite-state Markov chains and Markov renewal processes the condition

$$(35)\qquad M_{[k\cdot]}/\sqrt{k} \to 0 \quad \text{in probability as } k \uparrow +\infty$$

follows from the following observations. For these processes we can assume without loss of generality that the components of $X(t)$ are nondecreasing functions of t. The latter is achieved by selecting a suitable canonical representation of the exponential family given by (28). Thus, as is easily seen,

$$(36)\qquad M_{[kt]} \leq \text{const}\sum_{i=1}^{n}|X_i(\tau([kt]+1)) - X_i(\tau([kt]))|.$$

Since $(X(\tau(s)))_{s\geq 0}$ is a process with stationary independent increments, the condition (36) can be written also as

$$M_{[kt]} \leq U_k, \qquad k = 1, 2, \ldots, \tag{37}$$

where $U_1, U_2, \ldots$ are i.i.d. r.v.'s whose moments are finite in view of (30). Thus, (35) holds true. In the case of Markov-additive processes we have to assume additionally that

$$\frac{\sup\left(\sum_{i=1}^{m} |A_i(s) - A_i(\tau([kt]))| : \tau([kt]) \leq s < \tau([kt]+1)\right)}{\sqrt{k}} \to 0 \tag{38}$$

in distribution as $k \uparrow +\infty$. Clearly condition (38) is satisfied if the trajectories of the components of $A(t)$ are monotone functions of t. Finally, in view of the above considerations and Proposition 7.1, and under assumption (38) in the case of Markov-additive processes, we get the following.

PROPOSITION 7.3. *The process* $(Y_k(t))_{t\geq 0}$, *defined above in* (32), *converges in distribution to a Brownian motion starting at the origin with zero drift and diffusion coefficient* $\sigma^2\delta^{-1}(\theta)$, *where* σ^2 *is given above.*

REMARK 7.2. For the processes considered in this paper the constant $\delta(\theta)$ is explicitly given because $\tau(s)$ is always a linear function of the components of the minimal sufficient statistic $X_1(\tau(s)), \ldots, X_n(\tau(s))$.

In particular, Proposition 7.3 covers all existing central limit results for finite-state Markov chains, with either discrete or continuous time parameter, for which *explicit* limit results are available [cf. Iosifescu (1980), page 138] and Bhattacharya and Waymire (1990), pages 313–315]. Moreover, our Proposition 7.3 supplies further explicit solutions, not previously available, in functional limit theorems for finite-state Markov chains, Markov renewal processes and Markov-additive processes.

Let $\mathscr{K}$ be the set of absolutely continuous functions x in $C[0,1]$ with $x(0) = 0$ and whose derivatives $\dot{x}$ are such that

$$\int_0^1 (\dot{x}(t))^2\, dt \leq 1.$$

In view of a slight extension of Strassen's (1964) invariance principle from the space $C[0,1]$ to the space $D[0,1]$ [cf., e.g., the general Theorem 1 of Maller (1988)], the process $(2\log\log k)^{-1/2}\sigma^{-1}\tilde{Y}_k(t)$ is relatively compact with a.s. limit set $\mathscr{K}$, where σ was given above. Considerations similar to those made prior to Proposition 7.3 imply that

$$M_{[k\cdot]}/\sqrt{2k\log\log k} \to 0 \quad \text{a.s.},$$

in the case of finite-state Markov chains and Markov renewal processes. For Markov-additive processes we assume additionally that

$$\frac{\sup(\sum_{i=1}^{m}|A_i(s) - A_i(\tau([kt]))|: \tau([kt]) \le s < \tau([kt]+1))}{\sqrt{2k \log\log k}} \tag{39}$$
$$\to 0 \quad \text{a.s.}$$

Once again it is clear that condition (39) is satisfied if the trajectories of the components of $A(t)$ are a.s. monotone functions of t. Thus, in view of Proposition 7.2 and under the additional assumption (39) in the case of Markov-additive processes, we get the following.

PROPOSITION 7.4. *The process* $(2\log\log k)^{-1/2}\sigma^{-1}Y_k(t)$ *is relatively compact with a.s. limit points* $\{x(\delta^{-1}(\theta)\cdot): x \in \mathscr{K}\}$.

Likewise, strong and weak laws of large numbers are easily obtainable. For example, in view of Proposition 7.1 and under the suitably changed assumption (39), just replace $\sqrt{2k \log\log k}$ by k, in the case of Markov-additive processes, we get the following.

PROPOSITION 7.5. *For each* t *the sequence* $Y_k(t)/\sqrt{k} \to 0$ *a.s. as* $k \to +\infty$.

EXAMPLE 7.1 (Two-state Markov chain, continued). Bearing in mind that $\tau(1) = 1 + N_{1,2}(\tau(1)) + N_{2,2}(\tau(1))$ it is easy to see that

$$\delta(\theta) = 1 - \frac{\partial\varphi(\theta)}{\partial\theta_1} - \frac{\partial\varphi(\theta)}{\partial\theta_2},$$

where

$$\varphi(\theta) = \ln\frac{1-\exp(\theta_1)}{1+\exp(\theta_1)-\exp(\theta_2)}.$$

Consider, for example, the process $U_k(t) = N_{2,1}(kt) + N_{2,2}(kt)$, that is, the number of occurrences of state 2 in the first kt steps. Observe that in this case we have $a_1 = a_2 = b_1 = b_2 = 1$. Applying Proposition 7.3 we get that, as $k \uparrow +\infty$, the process

$$\frac{U_k(t) + kt(\partial\varphi(\theta)/\partial\theta_1 + \partial\varphi(\theta)/\partial\theta_2)(1 - \partial\varphi(\theta)/\partial\theta_1 - \partial\varphi(\theta)/\partial\theta_2)^{-1}}{\sqrt{k}}$$

converges to a Brownian motion with zero drift and diffusion coefficient $\sigma^2\delta^{-1}(\theta)$, where

$$\sigma^2 = -\left(1 + \delta^{-1}(\theta)\left(\frac{\partial\varphi(\theta)}{\partial\theta_1} + \frac{\partial\varphi(\theta)}{\partial\theta_2}\right)\right)^2\left(\sum_{i,j=1}^{2}\frac{\partial^2\varphi(\theta)}{\partial\theta_i\,\partial\theta_j}\right)$$

and $\delta(\theta)$ and $\varphi(\theta)$ are given above.

REMARK 7.3. If the process in question $U_k(t)$ is such that $\tilde{U}_k(t) = \text{const}$, then one should select another stopping rule $\tau_1(s)$, among those whose corresponding likelihood functions are noncurved exponential families, such that $\tilde{U}_k(t)$ is random. Such stopping rules exist for any linear function of the components of the minimal sufficient statistic in all the processes considered in this paper.

EXAMPLE 7.2 (Poisson process modulated by a two-state Markov process, continued). Bearing in mind that $\tau(s) = S_1(\tau(s)) + S_2(\tau(s))$, it is easy to see that

$$\delta(\theta) = -\frac{\partial\varphi(\theta)}{\partial\theta_1} - \frac{\partial\varphi(\theta)}{\partial\theta_2},$$

where

$$\varphi(\theta) = \ln\big((\theta_1 + \exp(\theta_3))(\theta_2 + \exp(\theta_4))\big).$$

Consider, for example, the sequence of processes $(A_1^{(k)}(t))_{t\geq 0}$, $k = 1, 2, \ldots,$ where

$$A_1^{(k)}(t) := A_1(kt).$$

Observe that $b_1 = b_2 = b_4 = 0$, $b_3 = 1$ and $a_1 = a_2 = 1$, $a_3 = a_4 = 0$. From Proposition 7.3 we get that, as $k \uparrow +\infty$, the process

$$\frac{A_1^{(k)}(t) + kt(-\partial\varphi(\theta)/\partial\theta_1 - \partial\varphi(\theta)/\partial\theta_2)^{-1}\,\partial\varphi(\theta)/\partial\theta_3}{\sqrt{k}}$$

converges to a Brownian motion with zero drift and diffusion coefficient $\sigma^2\delta^{-1}(\theta)$, where

$$\sigma^2 = -\left(\left(\delta^{-1}(\theta)\left(\frac{\partial\varphi(\theta)}{\partial\theta_3}\right)\right)^2\left(\sum_{i,j=1}^{2}\frac{\partial^2\varphi(\theta)}{\partial\theta_i\,\partial\theta_j}\right)\right.$$
$$\left.+\left(\delta^{-1}(\theta)\left(\frac{\partial\varphi(\theta)}{\partial\theta_3}\right)\right)\left(\sum_{i=1}^{2}\frac{\partial^2\varphi(\theta)}{\partial\theta_i\,\partial\theta_3}\right) + \frac{\partial^2\varphi(\theta)}{\partial\theta_3^2}\right),$$

and $\delta(\theta)$ and $\varphi(\theta)$ are given above.

Acknowledgments. The author is grateful to the referees and an Associate Editor for their careful reading of the paper and suggestions concerning the exposition.

REFERENCES

ALBERT, A. (1962). Estimating the infinitesimal generator of a continuous time finite-state Markov chain. *Ann. Math. Statist.* **33** 727–753.

ANDERSEN, A. N. (1969). Asymptotic results for exponential families. *Bull. Inst. Internat. Statist.* **43** 241–242.

ANDERSON, T. W. and GOODMAN, L. A. (1957). Statistical inference above Markov chains. *Ann. Math. Statist.* **28** 89–110.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families*. Wiley, New York.
BARNDORFF-NIELSEN, O. (1980). Conditionality resolutions. *Biometrika* **67** 293–310.
BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, London.
BERK, R. (1972). Consistency and asymptotic normality of maximum likelihood estimates for exponential models. *Ann. Math. Statist.* **43** 193–204.
BHATTACHARYA, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete* **60** 185–201.
BHATTACHARYA, R. N. and WAYMIRE, E. C. (1990). *Stochastic Processes with Applications*. Wiley, New York.
BILLINGSLEY, P. (1961a). *Statistical Inference for Markov Processes*. Univ. Chicago Press.
BILLINGSLEY, P. (1961b). Statistical methods in Markov chains. *Ann. Math. Statist.* **32** 12–40.
BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
BROWN, L. (1986). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, CA.
CHUNG, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*, 2nd ed. Springer, New York.
ÇINLAR, (1972a). Markov additive processes I. *Z. Wahrsch. Verw. Gebiete* **24** 85–93.
ÇINLAR, (1972b). Markov additive processes II. *Z. Wahrsch. Verw. Gebiete* **24** 95–121.
FEIGIN, P. (1981). Conditional exponential families and a representation theorem for asymptotic inference. *Ann. Statist.* **9** 597–603.
FYGENSON, M. (1991). Optimal sequential estimation for semi-Markov and Markov renewal processes. *Comm. Statist. Theory Methods* **20** 1427–1444.
GIHMAN, I. I. and SKOROHOD, A. V. (1974). *The Theory of Stochastic Processes* **1**. Springer, Berlin.
GILL, R. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrsch. Verw. Gebiete* **53** 97–116.
GORDIN, M. I. and LIFŠIC, B. A. (1978). The central limit theorem for stationary Markov processes. *Dokl. Acad. Nauk SSSR* **19** 392–394.
GUT, A. (1988). *Stopped Random Walks*. Springer, Berlin.
HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
HEYDE, C. C. and FEIGIN, P. (1975). On efficiency and exponential families in stochastic process estimation. In *Statistical Distributions in Scientific Work* (G. P. Patil, S. Kotz and J. K. Ords, eds.) **1** 227–240. Reidel, Dordrecht.
HÖPFNER, R. (1987). A remark on stopping times associated with continuous-time ergodic Markov chains and an application to statistical inference in birth-and-death processes. *Scand. J. Statist.* **14** 211–220.
IOSIFESCU, M. (1980). *Finite Markov Processes and their Applications*. Wiley, New York.
JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.
JOHANSEN, S. (1979). *Introduction to the Theory of Regular Exponential Families*. Univ. Copenhagen, Inst. Math. Statist.
KARR, A. (1991). *Point Processes and their Statistical Inference*, 2nd ed. Dekker, New York.
KÜCHLER, U. and SORENSEN, M. (1989). Exponential families of stochastic processes: a unifying semimartingale approach. *Internat. Statist. Rev.* **57** 123–160.
KÜCHLER, U. and SORENSEN, M. (1994). Exponential families of stochastic processes and Lévy processes. *J. Statist. Plann. Inference* **39** 211–237.
LEHMANN, E. (1986). *Testing Statistical Hypothesis*, 2nd ed. Wiley, New York.
MALLER, R. (1988). A functional law of the iterated logarithm for distributions in the domain of partial attraction of the normal distribution. *Stochastic Process. Appl.* **27** 179–194.
MOORE, E. H. and PYKE, R. (1968). Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Statist. Math.* **20** 411–424.
PHELAN, M. (1990a). Estimating the transition probabilities from censored Markov renewal processes. *Statist. Probab. Lett.* **10** 43–47.
PHELAN, M. (1990b). Bayes estimation from a Markov renewal process. *Ann. Statist.* **18** 603–616.

PRABHU, N. U. (1991). Markov renewal and Markov-additive processes. Technical Report 984, College of Engineering, Cornell Univ.

PYKE, R. and SCHAUFELE, R. (1964). Limit theorems for Markov renewal processes. *Ann. Math. Statist.* **35** 1746–1764.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

SERFOZO, R. F. (1975). Functional limit theorems for stochastic processes based on embedded processes. *Adv. in Appl. Probab.* **7** 123–139.

SILVEY, S. D. (1961). A note on maximum likelihood in the case of dependent random variables. *J. Roy. Statist. Soc. Ser. B* **23** 444–452.

SORENSEN, M. (1986). On sequential maximum likelihood estimation for exponential families of stochastic processes. *Internat. Statist. Rev.* **54** 191–210.

SORENSEN, M. (1991). Likelihood methods for diffusions with jumps. In *Statistical Inference in Stochastic Processes* (N. U. Prabhu and I. V. Basawa, eds.) 67–105. Dekker, New York.

STEFANOV, V. T. (1986). Efficient sequential estimation in exponential type processes. *Ann. Statist.* **14** 1606–1611.

STEFANOV, V. T. (1988a). A sequential approach for reducing curved exponential families of stochastic processes to noncurved exponential ones. *Contemp. Math.* **80** 323–330.

STEFANOV, V. T. (1988b). On some stopping times for dependent Bernoulli trials. *Scand. J. Statist.* **15** 39–50.

STEFANOV, V. T. (1991). Noncurved exponential families associated with observations over finite-state Markov chains. *Scand. J. Statist.* **18** 353–356.

STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS 6009, W.A.
AUSTRALIA