Empirical estimators for semi-Markov processes

Priscilla E. Greenwood
University of British Columbia

Wolfgang Wefelmeyer
University of Siegen

Abstract

A semi-Markov process stays in state x for a time s and then jumps to state y according to a transition distribution $Q(x, dy, ds)$. A statistical model is described by a family of such transition distributions. We give conditions for a nonparametric version of local asymptotic normality as the observation time tends to infinity. Then we introduce 'empirical' estimators for linear functionals of the distribution $\pi(dx)Q(x, dy, ds)$, with π denoting the invariant distribution of the embedded Markov chain, and characterize the empirical estimators which are efficient for a given model. We discuss efficiency of several classical estimators, in particular the jump frequency, the proportion of visits to a given set, the proportion of time spent in a set, and an estimator for $Q(x, \{y\} \times [0, t])$ suggested by Moore and Pyke (1968) for countable state space.

1 Introduction

A semi-Markov process $Y = (Y_t)_{t \geq 0}$ is a process on the time interval $[0, \infty)$, with values in some arbitrary state space E, which stays in state x for a time s and then jumps to state y according to a transition distribution $Q(x, dy, ds)$. For a review of estimation in semi-Markov models see Jain (1990). Applications are discussed in Janssen (1986) and Andersen et al. (1993).

Let T_i denote the i-th jump time. Set $T_0 = 0$. Write $X_i = Y_{T_i}$ for the state of Y at time T_i, and $S_i = T_i - T_{i-1}$ for the sojourn time of Y in state X_{i-1}. Then (X_i, S_i), $i \geq 0$, is a Markov chain with transition distribution $Q(x, dy, ds)$. Suppose the transition distribution $Q(x, dy \times [0, \infty))$ of the Markov chain X_i, $i \geq 0$, has an invariant distribution $\pi(dx)$. We want to estimate functionals of the form

$$\pi Q f = \iiint \pi(dx) Q(x, dy, ds) f(x, y, s), \tag{1.1}$$

[1]Work supported by NSERC, Canada.
[2]*AMS 1991 subject classifications.* Primary: 62M09; secondary: 60K15, 62G05.
[3]*Key words and phrases.* Efficient estimator, nonparametric estimation, semi-Markov process.

with $f(x, y, s)$ a $\pi(dx)Q(x, dy, ds)$-square integrable function on $E \times E \times [0, \infty)$. We observe Y on a time interval $[0, n]$ and write N_n for the observed number of jumps. If Y is recurrent, a natural estimator for πQf is the *'empirical' estimator*

$$E_n f = N_n^{-1} \sum_{i=1}^{N_n} f(X_{i-1}, X_i, S_i). \tag{1.2}$$

For countable state space, Pyke and Schaufele (1964) show that $n^{1/2}(E_n f - \pi Qf)$ is asymptotically normal. Our main result, Theorem 1, characterizes the functions f for which $E_n f$ is efficient in a given model.

Here are two applications for the model in which Q is completely unknown. By Theorem 2 the proportion $\#\{X_i \in B\}/N_n$ of visits to B is an efficient estimator for the probability πB of B under the invariant distribution of the embedded chain. By Theorem 3 the proportion $n^{-1}\int_0^n 1(Y_t \in B)dt$ of time spent by Y in B is an efficient estimator for the probability of B under the invariant distribution of the semi-Markov process Y.

Let us indicate two further applications when the state space is countable. Then the transition distribution Q is determined by the numbers $Q_{xy}(t) = Q\,(x, \{y\} \times [0, t])$. In the time interval $[0, n]$, let $N_n^{xy}(t)$ count the transitions from x to y after a sojourn time not longer than t, and $N_n^x(t)$ the visits to x of duration not longer than t. Write $N_n^{xy} = N_n^{xy}(\infty)$ and $N_n^x = N_n^x(\infty)$ for the corresponding counts with arbitrary sojourn times. Theorem 6 says that the estimator $N_n^{xy}(t)/N_n^x$ is efficient for $Q_{xy}(t)$ if Q is completely unknown. Theorem 8 says that the estimator $N_n^{xy}N_n^x(t)/(N_n^x)^2$ introduced by Moore and Pyke (1968) is efficient for $Q_{xy}(t)$ if the sojourn time and next state are independent given the present state.

If the estimator (1.2) is based on a function of other arguments besides X_{i-1}, X_i, S_i, then it is never efficient, not even if Q is completely unknown. This can be seen by an argument similar to Theorem 1. We also refer to the comment on this point in Greenwood and Wefelmeyer (1995a). A simple example for countable state space is the estimator $N_n^{x(2)y}/(N_n^x - 1)$ for the two-step transition probability $Q^2\,(x, \{y\} \times [0, \infty))$, with $N_n^{x(2)y}$ counting the two-step transitions from x to y. The estimator can be written as a ratio of two empirical estimators, with denominator $(N_n^x - 1)/(N_n - 1)$ and numerator

$$(N_n - 1)^{-1} \sum_{i=2}^{N_n} 1(X_{i-2} = x,\ X_i = y).$$

The latter estimator is not asymptotically equivalent to an estimator of the form (1.2), and one can show that it is not efficient. A better estimator for $Q^2\,(x, \{y\} \times [0, \infty))$ would make use of the representation of Q^2 in terms of the one-step transition distribution Q.

When the state space E is countable, we may describe the semi-Markov process through the counting processes $N^x = (N_t^x)_{t \geq 0}$, $x \in E$, with N_t^x the number of visits to state x in the time interval $[0,t]$. The vector of processes N^x, $x \in E$, was introduced and studied in detail by Pyke (1961a,b) under the name 'Markov renewal process'. (This name is often used in the sense introduced by Çinlar (1969), namely for the Markov chain (X_i, T_i), $i \geq 0$.) The counting processes N^x are used in applications to survival analysis. Typically, one considers a large number of i.i.d. copies of the semi-Markov process on a *fixed* time interval. This leads to a different asymptotic theory. For asymptotic normality of estimators for $Q_{xy}(t)$ on the basis of *censored* observations in such a setting we refer to Lagakos et al. (1978), Gill (1980), Voelkel and Crowley (1984) and Phelan (1990b). Bayes estimators are considered by Phelan (1990a). We will not study efficiency questions for these estimators here.

The paper is organized as follows. In Section 2 we formulate a nonparametric version of local asymptotic normality for semi-Markov models, and recall a Hájek–LeCam convolution theorem for differentiable functionals on such models. We show that the linear functional $\pi Q f$ defined in (1.1) is differentiable, and determine its gradient. Our main result, Theorem 1, describes a condition on the function f which is necessary and sufficient for the empirical estimator $E_n f$ defined in (1.2) to be efficient in a given model. We apply the result to two specific models: one in which nothing is assumed about the transition distribution, and one in which the sojourn time and next state are assumed independent given the present state. Section 3 gives applications for countable state space. The lemmas are proved in Section 4.

2 Characterizing efficient empirical estimators

Let E be an arbitrary set with countably generated σ-field $\mathcal{E}$. Let (X_i, S_i), $i \geq 0$, be a Markov chain with values in $E \times [0, \infty)$, transition distribution $Q(x, dy, ds)$ and initial distribution $\eta(dx)\varepsilon_0(ds)$. In particular, $S_0 = 0$. Set

$$\begin{aligned} T_i &= \sum_{j=0}^{i} S_j, \qquad i \geq 0, \\ N_t &= \max\{i \geq 0 : T_i \leq t\}, \qquad t \geq 0. \end{aligned}$$

Then $Y_t = X_{N_t}$, $t \geq 0$, is called a *semi-Markov process*. The process stays in state X_{i-1} for a *sojourn time* S_i. At *jump time* T_i it jumps to state X_i.

As usual, write

$$\begin{aligned} \pi \otimes Q(dx, dy, ds) &= \pi(dx) Q(x, dy, ds), \\ \pi Q(dy, ds) &= \int \pi(dx) Q(x, dy, ds). \end{aligned}$$

For a suitably integrable function $f(x, y, s)$ write

$$Qf(x) = Q_x f = \iint Q(x, dy, ds) f(x, y, s),$$

and for a function $f(x)$ set

$$\pi f = \int \pi(dx) f(x).$$

In particular, for a function $f(x, y, s)$,

$$\pi Q f = (\pi \otimes Q) f = \iiint \pi(dx) Q(x, dy, ds) f(x, y, s).$$

Assumption 1. *The Markov chain (X_i, S_i), $i \geq 0$, is positive Harris recurrent.*

Introduce the marginal distributions

$$Q_1(x, dy) = Q\left(x, dy \times [0, \infty)\right), \quad Q_2(x, ds) = Q(x, E \times ds).$$

Let π denote the invariant distribution of Q_1. We note that πQ is the invariant distribution of Q.

The *mean sojourn time* in state x is

$$m(x) = \int Q_2(x, ds) s.$$

Assumption 2. *We have $Q_2\left(x, \{0\}\right) = 0$ for $x \in E$, and $\pi m < \infty$.*

Let $\|f\| = (\pi f^2)^{1/2}$ denote the norm of $L_2(\pi)$, and $\|R\| = \sup\{\|Rf\| : \|f\| = 1\}$ the corresponding operator norm of a transition kernel $R(x, dy)$. Set $\Pi(x, dy) = \pi(dy)$.

Assumption 3. *We have $\|Q_1^j - \Pi\| \to 0$ for $j \to \infty$.*

We use the notations

$$F = L_2(\pi \otimes Q), \quad F_0 = L_2(\pi), \quad F_1 = L_2(\pi \otimes Q_1), \quad F_2 = L_2(\pi \otimes Q_2).$$

Note that for $f_1 \in F_1$ and $f_2 \in F_2$,

$$Q f_1 = Q_1 f_1, \quad Q f_2 = Q_2 f_2.$$

Let H denote the subspace of F consisting of all functions $h(x, y, s)$ with $Q_x h = 0$ for all $x \in E$. Write bH for the subspace of bounded functions in H. Define H_1 and H_2, bH_1 and bH_2 correspondingly.

The semi-Markov process Y is observed on the time interval $[0, n]$. Write P_n for the distribution of Y_t, $t \in [0, n]$. We begin with three essentially probabilistic lemmas. The proofs of Lemmas 1, 4 and 5 are in Section 4.

Lemma 1. *Under Assumptions 1 and 2, $n/N_n \to \pi m$ a.s.*

By Assumption 2, πm is positive. Lemma 1 and a central limit theorem for Markov chains (Meyn and Tweedie, 1994, p. 411, Theorem 17.0.1(i)) imply the following.

Lemma 2. *Under Assumptions 1 and 2, for* $h \in H$,

$$n^{-1/2} \sum_{i=1}^{N_n} h(X_{i-1}, X_i, S_i) \Rightarrow N_h \quad \textit{under } P_n,$$

where N_h *is normal with mean* 0 *and variance* $(\pi m)^{-1} \pi Q h^2$.

From Lemma 1 and a martingale approximation for Markov chains of Gordin and Lifšic (1978), the idea of which goes back to Gordin (1969), we obtain the following stochastic approximation. For a stronger version of the martingale approximation with a more detailed proof we refer to Greenwood and Wefelmeyer (1995a, Lemma 1).

Lemma 3. *Under Assumptions 1 to 3, for* $f \in F$,

$$n^{-1/2} \sum_{i=1}^{N_n} (f(X_{i-1}, X_i, S_i) - \pi Q f - (Af)(X_{i-1}, X_i, S_i)) = o_{P_n}(1)$$

with

$$(Af)(x, y, s) = f(x, y, s) - Q_x f + (A_0 Q f)(x, y). \tag{2.1}$$

Here A_0 *is defined for* $f_0 \in F_0$ *by*

$$(A_0 f_0)(x, y) = \sum_{j=0}^{\infty} (Q_{1y}^j f_0 - Q_{1x}^{j+1} f_0). \tag{2.2}$$

The operator A_0 is linear and maps F_0 into H_1. The operator A is linear and maps F into H; it is the identity on H. Lemmas 1 and 3 imply that for $f \in F$ the empirical estimator

$$E_n f = N_n^{-1} \sum_{i=1}^{N_n} f(X_{i-1}, X_i, S_i)$$

introduced in (1.2) admits a stochastic approximation

$$n^{1/2}(E_n f - \pi Q f) = \pi m \cdot n^{-1/2} \sum_{i=1}^{N_n} (Af)(X_{i-1}, X_i, S_i) + o_{P_n}(1). \tag{2.3}$$

Then Af is called the *influence function* of $E_n f$ in H. Lemma 2 implies that $n^{1/2}(E_n f - \pi Q f)$ is asymptotically normal with variance $\pi m \cdot \pi Q (Af)^2$. For finite or countable state space, different proofs of asymptotic normality have been given by Taga (1963) for $f(x, y, s)$ equal to $\delta_u(y)$ and $s\delta_u(x)$ with $u \in E$, by Pyke and Schaufele (1964) and

Hatori (1966) for general $f(x,y,s)$, and by McLean and Neuts (1967) for $sf_0(x)$. For arbitrary state space, refinements of this central limit theorem are given by Malinovskii (1985) for $f(x,y,s)$ equal to s, (1986) and (1987) for $f(y)$ and (1991) for $f(x,s)$.

We turn to two statistical lemmas. First we show that a semi-Markov model is locally asymptotically normal under our assumptions. Such a model is indexed by a family of transition distributions. Examples which we consider later are the 'full' model, indexed by the set of all transition distributions on E, and the model indexed by the set of transition distributions for which the sojourn time and next state are independent given the present state. We describe a local model around the fixed transition distribution Q as follows. Fix a closed linear subspace $G \subset H$. Choose $G' \subset bG$ such that G' is dense in G. The set G' will play the role of *local parameter space*. For $h \in G'$ choose a sequence Q_{nh} of transition distributions in the index set with *derivative* h in the following sense:

$$Q_{nh}(x,dy,ds) = Q(x,dy,ds)\left(1 + n^{-1/2}h_n(x,y,s)\right) \tag{2.4}$$

with $h_n \to h$ in sup-norm. We write P_n and P_{nh} for the distribution of Y_t, $t \in [0,n]$, if Q and Q_{nh}, respectively, are true, and the initial distribution is η.

Lemma 4. *Under Assumptions 1 and 2, for $h \in G'$,*

$$\log dP_{nh}/dP_n = n^{-1/2}\sum_{i=1}^{N_n} h(X_{i-1},X_i,S_i) - \frac{1}{2}(\pi m)^{-1}\pi Q h^2 + o_{P_n}(1).$$

Together with Lemma 2, this is a nonparametric version of local asymptotic normality. For finite state space and finite-dimensional parameter, and when the sojourn time and next state are independent given the present state, the lemma is basically due to Akritas and Roussas (1980). For arbitrary state space, a parametric version is implicit in Malinovskii (1992). The lemma remains true if differentiability (2.4) is replaced by an appropriate variant of Hellinger differentiability. For Markov step processes, with transition distribution

$$Q(x,dy,ds) = Q_1(x,dy)\lambda(x)\exp\left(-s\lambda(x)\right)ds,$$

compare condition H1$''$ in Höpfner et al. (1990). Versions of Lemma 4 for Markov step processes are contained in Höpfner (1993a, 1993b). He writes the approximation to the likelihood ratio in a different way. Local asymptotic normality for Markov step processes is also implicit in Malinovskii (1989). The strong differentiability condition (2.4) suffices for our purposes.

The efficiency concept we use is asymptotic. From Lemmas 4 and 2 we see that the underlying family of distributions is locally approximated by a Gaussian shift family indexed by functions h in the local parameter space G'. Given a smooth functional of the transition distribution, we can describe a lower bound on the asymptotic risk for a large class of estimators of the functional. Smoothness is defined as follows. A functional $k(Q)$ is said to be *differentiable* on G' with *gradient* g if $g \in F$ and

$$n^{1/2}\left(k(Q_{nh}) - k(Q)\right) \to \pi Qhg, \qquad h \in G'.$$

The projection g_G of g into G is also a gradient, the *canonical gradient.* It is uniquely determined. We consider the following class of estimators. An estimator $\hat{k}$ is called *regular* on G' for k with *limit* L if

$$n^{1/2}\left(\hat{k} - k(Q_{nh})\right) \Rightarrow L \quad \text{under} \quad P_{nh}, \qquad h \in G'.$$

The point here is that L does not depend on the local parameter h. In other words: the distribution of the standardized error is asymptotically equivariant under local changes of measure.

The convolution theorem states that $L = M + N$ with M independent of N, and N normal with mean 0 and variance $\pi m \cdot \pi Q g_G^2$. A convenient reference is Greenwood and Wefelmeyer (1990). Among estimators $\hat{k}$ having such a representation of the limit L, those with $M = 0$ have standardized error $n^{1/2}\left(\hat{k} - k(Q)\right)$ most concentrated, asymptotically, in any symmetric interval. Eqivalently, they have minimal asymptotic risk with respect to any bounded symmetric bowl-shaped loss function. This justifies calling $\hat{k}$ *efficient* on G' for k if

$$n^{1/2}\left(\hat{k} - k(Q)\right) \Rightarrow N \quad \text{under } P_n.$$

The local parameter space G' does not enter the definition of efficiency except that it restricts the competing estimators to those which are regular on G'.

The next lemma says that πQf is a differentiable functional of Q. It generalizes a result for Markov chains and functions $f(y) = 1(y \leq t)$ in Penev (1991).

Lemma 5. *Under Assumption 3, for $f \in F$, the functional πQf is differentiable on G' with gradient Af.*

We can now characterize the functions f for which the empirical estimator $E_n f$ is efficient in a given model, with local parameter space G'.

Theorem 1. *Under Assumptions 1 to 3, for $f \in F$, the estimator $E_n f$ is regular and efficient on G' for πQf if and only if $Af \in G$.*

Proof. By Lemma 5, the functional πQf is differentiable on G' with gradient $g = Af$. The estimator $E_n f$ is regular and efficient on G' for πQf if and only if it has the canonical gradient g_G as influence function in H:

$$n^{1/2}(E_n f - \pi Qf) = \pi m \cdot n^{-1/2} \sum_{i=1}^{N_n} g_G(X_{i-1}, X_i, S_i) + o_{P_n}(1). \tag{2.5}$$

This is a straightforward consequence of local asymptotic normality, Lemmas 2 and 4. For a convenient reference see Greenwood and Wefelmeyer (1990). The factor πm appears because the gradient is defined in terms of the norm $(\pi Q h^2)^{1/2}$, not in terms of the norm $((\pi m)^{-1} \pi Q h^2)^{1/2}$ induced by local asymptotic normality. — On the other hand, by relation (2.3) the estimator $E_n f$ has the stochastic approximation

$$n^{1/2}(E_n f - \pi Q f) = \pi m \cdot n^{-1/2} \sum_{i=1}^{N_n} (Af)(X_{i-1}, X_i, S_i) + o_{P_n}(1).$$

If $Af = g_G$, the estimator $E_n f$ must be regular and efficient by the characterization (2.5). Conversely, if $E_n f$ is regular and efficient, then, again by characterization (2.5),

$$n^{-1/2} \sum_{i=1}^{N_n} (g_G - Af)(X_{i-1}, X_i, S_i) = o_{P_n}(1). \tag{2.6}$$

By Lemma 2, the asymptotic variance of this random variable is $(\pi m)^{-1} \pi Q (g_G - Af)^2$. By (2.6) we must have $\pi Q(g_G - Af)^2 = 0$ and hence $Af = g_G \;\; \pi \otimes Q$-a.s. □

Let us apply Theorem 1 to the *full model* with arbitrary Q, and to the model in which the sojourn time and next state are independent given the present state.

1. Full model. In this subsection we consider the model in which no structural assumptions on the transition distribution are made. Then there is a particularly simple local model, with $G = H$ and local parameter space $G' = bH$: For $h \in bH$, set

$$Q_{nh}(x, dy, ds) = Q(x, dy, ds) \left(1 + n^{-1/2} h(x, y, s)\right).$$

Differentiability (2.4) holds trivially with $h_n = h$. Since the operator A maps F into H, Theorem 1 implies the following.

Theorem 2. *Under Assumptions 1 to 3, for all $f \in F$, the estimator $E_n f$ is regular and efficient on bH for $\pi Q f$. Its asymptotic variance is $\pi m \cdot \pi Q (Af)^2$.*

For $f(x, y, s) = 1(y \in B)$ we obtain that the proportion $\#\{X_i \in B\}/N_n$ of visits to B among the observed jumps is an efficient estimator for the probability πB of B under the invariant distribution of the embedded Markov chain.

The proportion $n^{-1} \int_0^n 1(Y_t \in B) dt$ of time spent by the semi-Markov process Y in B is an estimator for $\pi(m 1_B)/\pi m$. This is the probability of B under the invariant distribution of Y; see e.g. Tomko (1989). The following theorem shows that the estimator is efficient if Q is completely unknown. We make an additional assumption on Q_2.

Assumption 4. *The family of distributions $Q_2(x, ds)$ is tight.*

The assumption can be avoided in Theorem 3 if one knows that Y_n converges in distribution. Tomko (1989) gives sufficient conditions.

Theorem 3. *Let Assumptions 1 to 4 hold, and let πQ_2 have a finite second moment. Then n/N_n is regular and efficient on bH for πm. For $B \in \mathcal{E}$, the estimator $n^{-1}\int_0^n 1(Y_t \in B)dt$ is regular and efficient on bH for $\pi(m1_B)/\pi m$.*

Proof. Consider first the estimator

$$\hat{k}_B = N_n^{-1}\sum_{i=1}^{N_n} S_i 1(X_{i-1} \in B).$$

It is of the form $E_n f$ with $f(x,y,s) = s1(x \in B)$. We have

$$\pi Q f^2 \le \iint \pi(dx)Q_2(x,ds)s^2 < \infty.$$

Hence $f \in F$, and $\hat{k}_B$ is regular and efficient on bH for $\pi(m1_B)$ by Theorem 1. In particular, $\hat{k}_E = T_{N_n}/N_n$ is regular and efficient on bH for πm. It remains to show that the estimator $\hat{k}_B$ is asymptotically equivalent to the estimator

$$N_n^{-1}\int_0^n 1(Y_t \in B)dt.$$

Then $\hat{k}_E$ is asymptotically equivalent to n/N_n, and $\hat{k}_B/\hat{k}_E$ is asymptotically equivalent to the proportion of time spent by Y in B, and Theorem 3 is proved. To prove the required asymptotic equivalence, write

$$\hat{k}_B = N_n^{-1}\int_0^{T_{N_n}} 1(Y_t \in B)dt.$$

Since n/N_n is bounded in probability by Lemma 1, it remains to show that $n^{-1/2}(n - T_{N_n}) = o_{P_n}(1)$. It suffices to see that $T_{N_n+1} - T_{N_n}$ is bounded in probability. By Assumption 4, for each $\varepsilon > 0$ there exists c such that for all $x \in E$,

$$Q_2(x,[c,\infty)) \le \varepsilon.$$

Hence for all n,

$$P\{T_{N_n+1} - T_{N_n} \ge c\} = \int P(Y_n \in dx)Q_2(x,[c,\infty)) \le \varepsilon.$$

□

From the proof of Theorem 3 and Lemmas 2 and 3 we obtain in particular that N_n/n is asymptotically normal. For finite state space see Taga (1963).

2. Conditionally independent time and state. In this subsection we consider the model given by all transition distributions for which the sojourn time and next state are independent given the present state,

$$Q(x, dy, ds) = Q_1(x, dy)Q_2(x, ds).$$

A local model is introduced as follows. For $h_1 \in bH_1$ and $h_2 \in bH_2$ set

$$\begin{aligned} Q_{1nh_1}(x, dy) &= Q_1(x, dy)\left(1 + n^{-1/2}h_1(x, y)\right), \\ Q_{2nh_2}(x, ds) &= Q_2(x, ds)\left(1 + n^{-1/2}h_2(x, s)\right). \end{aligned}$$

Choose $G = H_1 + H_2$. For

$$h = h_1 + h_2 \in bH_1 + bH_2 = G'$$

set

$$Q_{nh}(x, dy, ds) = Q_{1nh_1}(x, dy)Q_{2nh_2}(x, ds).$$

Then differentiability (2.4) holds with

$$h_n(x, y, s) = h_1(x, y) + h_2(x, s) + n^{-1/2}h_1(x, y)h_2(x, s).$$

Theorem 4. *Under Assumptions 1 to 3, for $f \in F$, the estimator $E_n f$ is regular and efficient on $bH_1 + bH_2$ for πQf if and only if $f \in F_1 + F_2$.*

Proof. By Theorem 1, the estimator $E_n f$ is regular and efficient on $bH_1 + bH_2$ if and only if $Af \in H_1 + H_2$. By definition (2.1) of Af,

$$(Af)(x, y, s) = f(x, y, s) - Q_x f + (A_0 Qf)(x, y).$$

We have $A_0 Qf \in H_1$. Furthermore, $f - Qf \in H_1 + H_2$ if and only if $f \in F_1 + F_2$. □

The spaces H_1 and H_2 are orthogonal with respect to the norm $(\pi Qf^2)^{1/2}$. In particular, for $h = h_1 + h_2$ and $h' = h_1' + h_2'$ in $H_1 + H_2$,

$$\pi Qhh' = \pi Q_1 h_1 h_1' + \pi Q_2 h_2 h_2'.$$

By Theorem 4, for $f_1 \in F_1$ the estimator

$$E_n f_1 = N_n^{-1} \sum_{i=1}^{N_n} f_1(X_{i-1}, X_i)$$

is regular and efficient on $bH_1 + bH_2$ for $\pi Q_1 f_1$. By Lemma 5, the functional $\pi Q_1 f_1$ has canonical gradient

$$f_1(x, y) - Q_{1x} f_1 + (A_0 Q_1 f_1)(x, y).$$

The canonical gradient is in H_1. This means that $E_n f_1$ remains efficient under arbitrary restrictions on Q_2. For example, the semi-Markov process may be a Markov step process, with

$$Q_2(x, ds) = \lambda(x) \exp\left(-s\lambda(x)\right) ds,$$

or a Markov chain, with $Q_2\left(x, \{1\}\right) = 1$ for $x \in E$. Efficiency of $E_n f_1$ for Markov step processes was already shown in Greenwood and Wefelmeyer (1994), for Markov chains in Greenwood and Wefelmeyer (1995a) and Bickel (1993).

While $E_n f_1$ remains efficient under restrictions on Q_2, it is not true that $E_n f_2$ remains efficient under restrictions on Q_1, in general. By (2.3) the influence function of $E_n f_2$ in H is

$$f_2(x, s) - Q_{2x} f_2 + (A_0 Q_2 f_2)(x, y).$$

If H_1 is replaced by a smaller space, say H_1', then $A_0 Q_2 f_2$ may fall outside H_1', and the influence function will not equal the canonical gradient in H_1'. An example is the estimator n/N_n. By the proof of Theorem 3, it is asymptotically equivalent to $E_n f_2$ with $f_2(x, s) = s$. We have $Q_2 f_2 = m$ and

$$(A_0 Q_2 f_2)(x, y) = \sum_{j=0}^{\infty} (Q_{1y}^j m - Q_{1x}^{j+1} m),$$

which may not lie in H_1'. Hence it may be possible to improve the estimator n/N_n if we know more about Q_1. This is not surprising in view of the fact that n/N_n estimates πm which depends, through π, on Q_1.

3 Countable state space

In this section we apply the results of Section 2 to some functionals of interest when the state space E is countable, with discrete σ-field. Fix a transition distribution $Q(x, dy, ds)$ and an initial distribution $\eta(dx)\varepsilon_0(ds)$. Introduce

$$q_{xy} = Q_1\left(x, \{y\}\right), \quad H_x(t) = Q_2\left(x, [0, t]\right).$$

The transition distribution is determined by the numbers

$$Q_{xy}(t) = Q\left(x, \{y\} \times [0, t]\right) = q_{xy} F_{xy}(t),$$

with $F_{xy}(t)$ the conditional distribution function of the sojourn time in state x, given that the next state will be y. Note that

$$H_x(t) = \sum_{y \in E} q_{xy} F_{xy}(t).$$

The mean sojourn time in state x is

$$m_x = \int t \, dH_x(t).$$

Write $q_{xy}^i = Q_1^i(x, \{y\})$ for the i-step transition probability from x to y.

Assumptions 1 to 3 of Section 2 are implied by the following assumption. It will be in force throughout Section 3.

Assumption. *The Markov chain (X_i, S_i), $i \geq 0$, is irreducible and positive recurrent, and $\|Q_1^j - \Pi\| \to 0$ for $j \to \infty$. Further, $H_x(0) = 0$ for $x \in E$, and $\sum_{x \in E} \pi_x m_x < \infty$, where the π_x denote the invariant probabilities of (q_{xy}).*

The estimators below will be expressed in terms of the following four statistics. In the time interval $[0, n]$, and for $u, v \in E$, let $N_n^{uv}(t)$ count the transitions from u to v after a sojourn time not longer than t, and let $N_n^u(t)$ count the visits to u of duration not longer than t. Write $N_n^{uv} = N_n^{uv}(\infty)$ and $N_n^u = N_n^u(\infty)$ for the corresponding counts with arbitrary sojourn time.

1. Full model. In this subsection we consider the model given by all transition distributions, and construct a local model with local parameter space $G' = bH$ as in Subsection 1 of Section 2.

Theorem 5. *For $u, v \in E$ and $t \geq 0$, the estimators N_n^u/N_n, N_n^{uv}/N_n^u, $N_n^u(t)/N_n^u$ and $N_n^{uv}(t)/N_n^u$ are regular and efficient in bH for π_u, q_{uv}, $H_u(t)$ and $F_{uv}(t)$, respectively.*

Proof. Apply Theorem 2 with $f(x, y, s)$ replaced by $\delta_u(y)$, $\delta_u(x)\delta_v(y)$, $\delta_u(x)1(s \leq t)$ and $\delta_u(x)\delta_v(y)1(s \leq t)$. These functions are in bF. Hence the estimators N_n^u/N_n, N_n^{uv}/N_n, $N_n^u(t)/N_n$ and $N_n^{uv}(t)/N_n$ are regular and efficient on bH for π_u, $\pi_u q_{uv}$, $\pi_u H_u(t)$ and $\pi_u q_{uv} F_{uv}(t)$, respectively. The assertion follows by combining these estimators appropriately. □

Theorem 5 implies the following.

Theorem 6. *For $u, v \in E$ and $t \geq 0$, the estimator $N_n^{uv}(t)/N_n^u$ is regular and efficient on bH for $Q_{uv}(t)$.*

We will see that the estimator in Theorem 6 can be improved if the sojourn time and next state are known to be independent given the present state.

2. Conditionally independent time and state. In this subsection we consider the model given by all transition distributions for which the sojourn time and next state are independent given the present state. This means that $F_{xy}(t)$ does not depend on y. Hence

$$F_{xy}(t) = H_x(t), \quad Q_{xy}(t) = q_{xy} H_x(t).$$

We show that the first three estimators in Theorem 5 remain efficient in this smaller model. Construct a local model with local parameter space $G' = bH_1 + bH_2$ as in Subsection 2 of Section 2.

Theorem 7. *For $u, v \in E$ and $t \geq 0$, the estimators N_n^u/N_n, N_n^{uv}/N_n^u and $N_n^u(t)/N_n^u$ are regular and efficient on $bH_1 + bH_2$ for π_u, q_{uv} and $H_u(t)$, respectively.*

Proof. Apply Theorem 4 with $f(x, y, s)$ replaced by $\delta_u(y)$, $\delta_u(x)\delta_v(y)$ and $\delta_u(x)1(s \leq t)$. These functions are in $bF_1 + bF_2$. Hence N_n^u/N_n, N_n^{uv}/N_n and $N_n^u(t)/N_n$ are regular and efficient on $bH_1 + bH_2$ for π_u, $\pi_u q_{uv}$ and $\pi_u H_u(t)$, respectively. The assertion follows by combining these estimators appropriately. □

By Lemmas 2 and 3, the estimators in Theorem 7 are asymptotically normal. For finite state space see Moore and Pyke (1968). Theorem 7 implies the following.

Theorem 8. *For $u, v \in E$ and $t \geq 0$, the Moore–Pyke estimator $N_n^{uv}N_n^u(t)/(N_n^u)^2$ is regular and efficient on $bH_1 + bH_2$ for $Q_{uv}(t)$.*

Moore and Pyke (1968) introduce the estimator described in Theorem 8 and show that it is asymptotically normal if the sojourn time and next state are independent given the present state. They remark that, according to Pyke and Schaufele (1964), this assumption incurs no loss of generality. The argument is also in Pyke and Schaufele (1966). Note, however, that in the larger model the Moore–Pyke estimator does not estimate $Q_{uv}(t)$ but $q_{uv}H_u(t)$.

The estimator $N_n^{uv}(t)/N_n^u$ in Theorem 6 is not efficient when the sojourn time and next state are independent given the present state. It is the ratio of the two estimators $N_n^{uv}(t)/N_n$ and N_n^u/N_n. The second estimator is efficient on $bH_1 + bH_2$ by Theorem 7; the first is of the form $E_n f$ with $f(x, y, s) = \delta_u(x)\delta_v(y)1(s \leq t)$. Since f is not in $F_1 + F_2$, the first estimator is inefficient by Theorem 4. Hence $N_n^{uv}(t)/N_n^u$ is inefficient.

4 Proofs of the lemmas

Proof of Lemma 1. By Assumption 2, the expectation πm of S_i under the stationary law $\pi \otimes Q_2$ of (X_i, S_i) is finite. By Assumption 1, the Markov chain (X_i, S_i), $i \geq 0$, is positive Harris recurrent, and the ergodic theorem (e.g., Meyn and Tweedie, 1994, p. 411, Theorem 17.0.1(iv)) implies

$$T_n/n = n^{-1} \sum_{i=1}^{n} S_i \to \pi m \quad \text{a.s.}$$

We have $T_{N_n}/n \leq 1 < T_{N_n+1}/n$ and $N_n \to \infty$ a.s. Hence $T_{N_n}/n \to 1$ a.s. and therefore $T_{N_n}/N_n \to \pi m$ a.s. This implies $n/N_n \to \pi m$ a.s. □

Proof of Lemma 4. To prove local asymptotic normality, we need a representation of the likelihood ratio in terms of the transition distributions Q and Q_{nh}. It is obtained by specializing the representation for multivariate point processes due to Jacod (1975).

Associate with the semi-Markov process Y a multivariate point process, i.e. a random measure on $(0,\infty)\times E$,

$$\mu(dt,dy)=\sum_{i\geq 1}\varepsilon_{(T_i,X_i)}(dt,dy).$$

If the transition distribution is Q, the random measure has compensator

$$\nu(dt,dy)=Q(X_{t-},dy,dt-T_{N_{t-}})\big/Q_2\left(X_{t-},[t-T_{N_{t-}},\infty)\right).$$

See Jacod (1975, Proposition (3.1)) or Jacod and Shiryaev (1987, p. 136, Theorem 1.33). Set

$$\begin{aligned} a_t &= \nu\left(\{t\}\times E\right), \\ \nu^c(dt,dy) &= 1(a_t=0)\nu(dt,dy). \end{aligned}$$

Write ν_{nh} and a_t^{nh} if the transition distribution is Q_{nh}, and define Y_{nh} by

$$\nu_{nh}(dt,dy)=Y_{nh}(t,y)\nu(dt,dy).$$

Since the initial distribution η of the semi-Markov process Y does not change with Q, we obtain from Jacod (1975, Theorem (5.1)) or Jacod and Shiryaev (1987), p. 190, Theorem 5.43) the following representation of the likelihood ratio between P_{nh} and P_n,

$$\begin{aligned} dP_{nh}/dP_n &= \left(\prod_{i=1}^{N_n} Y_{nh}(X_i,S_i)\right)\prod_{t\leq n,t\neq T_i}\left(1-\frac{a_t^{nh}-a_t}{1-a_t}\right) \\ &\quad \exp\int_0^n\int_E\left(1-Y_{nh}(t,y)\right)\nu^c(dt,dy). \end{aligned}$$

We rewrite the likelihood ratio using the hazard measure

$$H(x,ds)=Q_2(x,ds)/Q_2\left(x,[s,\infty)\right).$$

Let $H^c(x,\cdot)$ denote the continuous part of $H(x,\cdot)$. Then

$$\begin{aligned} a_t &= H(X_{t-},\{t-T_{N_{t-}}\}), \\ \nu^c(dt\times E) &= H^c(X_{t-},dt-T_{N_{t-}}). \end{aligned}$$

Analogous relations hold for a_t^{nh} and ν_{nh}^c. Note that $a_t^{nh}=1$ iff $a_t=1$. By definition (2.4) of Q_{nh} the ν-density of ν_{nh} is

$$\begin{aligned} Y_{nh}(t,y) &= \left(1+n^{-1/2}h_n(X_{i-1},X_i,S_i)\right) \\ &\quad Q_2\left(X_{t-},[t-T_{N_{t-}},\infty)\right)\big/Q_{2nh}\left(X_{t-},[t-T_{N_{t-}},\infty)\right). \end{aligned}$$

Hence the likelihood ratio is

$$\begin{aligned} dP_{nh}/dP_n &= \prod_{i=1}^{N_n} \left(1+n^{-1/2}h_n(X_{i-1},X_i,S_i)\right) \\ &\quad Q_2\left(X_{i-1},[S_i,\infty)\right) \Big/ Q_{2nh}\left(X_{i-1},[S_i,\infty)\right) \\ &\quad \prod_{t\le n, t\ne T_i} \left(1-\frac{H_{nh}(X_{t-},\{t-T_{N_{t-}}\})-H(X_{t-},\{t-T_{N_{t-}}\})}{1-H(X_{t-},\{t-T_{N_{t-}}\})}\right) \\ &\quad \exp\int_0^n \left(H^c(X_{t-},dt-T_{N_{t-}})-H^c_{nh}(X_{t-},dt-T_{N_{t-}})\right). \end{aligned}$$

We show that most of the factors other than $1+n^{-1/2}h_n$ cancel. By the deterministic version of the Doléans–Dade exponential formula given by Jacod (1975, Lemma (3.5)),

$$Q_2\left(x,(t,\infty)\right) = \exp\left(-H^c\left(x,(0,t]\right)\right)\prod_{s\le t}\left(1-H(x,\{s\})\right).$$

Furthermore,

$$Q_2\left(x,(t,\infty)\right) = Q_2\left(x,[t,\infty)\right)\left(1-H(x,\{t\})\right).$$

Corresponding relations hold for Q_{2nh} and its hazard measure H_{nh}. Hence the likelihood ratio is

$$\begin{aligned} dP_{nh}/dP_n &= \prod_{i=1}^{N_n} \left(1+n^{-1/2}h_n(X_{i-1},X_i,S_i)\right) \\ &\quad Q_{2nh}\left(X_{N_n},(n-T_{N_n},\infty)\right) \Big/ Q_2\left(X_{N_n},(n-T_{N_n},\infty)\right). \end{aligned}$$

For finite state space, and when the sojourn time and next state are independent given the present state, similar representations are in Moore and Pyke (1968) and Akritas and Roussas (1980).

We show that the last displayed line tends to 1. Define $R_1(x,dy,s)$ by

$$Q(x,dy,ds) = R_1(x,dy,s)Q_2(x,ds).$$

Set $h_{n2}(x,s) = \int R_1(x,dy,s)h_n(x,y,s)$. Then

$$Q_{2nh}(x,ds) = Q_2(x,ds)\left(1+n^{-1/2}h_{n2}(x,s)\right).$$

Since h is bounded and $h_n \to h$ in sup-norm, there exists d such that $|h_{2n}| \le d$ for n sufficiently large. Hence for measurable sets $D \subset [0,\infty)$,

$$\begin{aligned} |Q_{2nh}(x,D)-Q_2(x,D)| &= n^{-1/2}\left|\int Q_2(x,ds)h_{n2}(x,s)1(s\in D)\right| \\ &\le n^{-1/2}dQ_2(x,D). \end{aligned}$$

Therefore, the likelihood ratio has the approximation

$$dP_{nh}/dP_n = \left(\prod_{i=1}^{N_n} \left(1 + n^{-1/2} h_n(X_{i-1}, X_i, S_i)\right)\right)(1 + o_{P_n}(1)).$$

If N_n were nonrandom, the right-hand term would be a representation for the likelihood ratio of the Markov chain (X_i, S_i) observed for $i = 0, \dots, N_n$. Compare also Gill (1983) and Andersen et al. (1993, p. 680). Since $N_n/n \to (\pi m)^{-1}$ a.s. by Lemma 1, the assertion now follows by a similar expansion as for Markov chains; see Penev (1991). □

Proof of Lemma 5. Greenwood and Wefelmeyer (1995b), following Kartashov (1985), who uses other norms, obtain a perturbation expansion for invariant distributions: Uniformly for $f_0 \in F_0$ with $\|f_0\| \le 1$,

$$(\pi' - \pi) f_0 = \pi(Q_1' - Q_1) A_0 f_0 + o(\|Q_1' - Q_1\|), \tag{4.1}$$

with A_0 defined in (2.2).

We apply (4.1) for $Q_1' = Q_{1nh}$. Define $R_2(x, y, ds)$ by

$$Q(x, dy, ds) = Q_1(x, dy) R_2(x, y, ds).$$

Set $h_{n1}(x, y) = \int R_2(x, y, ds) h_n(x, y, s)$. Then

$$Q_{1nh}(x, dy) = Q_1(x, dy)\left(1 + n^{-1/2} h_{n1}(x, y)\right).$$

Since h is bounded and $h_n \to h$ in sup-norm, there exists d such that $|h_n| \le d$. Therefore, $|h_{n1}| \le d$ for n sufficiently large, and

$$\|Q_{1nh} - Q_1\| \le n^{-1/2} d.$$

Hence (4.1) implies uniformly for $f_0 \in F_0$ with $\|f_0\| \le 1$,

$$n^{1/2}(\pi_{nh} - \pi) f_0 = \pi Q_1(h_{n1} A_0 f_0) + o(1). \tag{4.2}$$

Let $f \in F$. Use the definition of Q_{nh} to write

$$\begin{aligned} &n^{1/2}(\pi_{nh} Q_{nh} f - \pi Q f) \\ &= \pi Q h_n f + n^{1/2}(\pi_{nh} - \pi) Q f + (\pi_{nh} - \pi) Q h_n f. \end{aligned} \tag{4.3}$$

We consider the three right-hand terms, beginning with the last. Since the sequence $\|Q h_n f\|$ is bounded,

$$(\pi_{nh} - \pi) Q h_n f \to 0. \tag{4.4}$$

Applying (4.2) for $f_0 = Qf$ and using the definition of h_{n1},

$$
\begin{aligned}
(4.5) \qquad n^{1/2}(\pi_{nh} - \pi)Qf &= \pi Q(hA_0Qf) + \pi Q\left((h_n - h)A_0Qf\right) + o(1) \\
&\to \pi Q(hA_0Qf).
\end{aligned}
$$

Finally, using $Qh = 0$,

$$
(4.6) \qquad \pi Q h_n f \to \pi Q h f = \pi Q\left(h(f - Qf)\right).
$$

Applying (4.4)–(4.6) to (4.3),

$$
n^{1/2}(\pi_{nh}Q_{nh}f - \pi Qf) \to \pi Q\left(h(f - Qf)\right) + \pi Q(hA_0Qf) = \pi Q(hAf).
$$

This is the assertion. □

Acknowledgment. We thank Ron Pyke for suggesting the main problem of this paper and for the idea of Lemma 1, and Vsevolod Malinovskii for observing that our conditions could be weakened using a result of Gordin and Lifšic.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Akritas, M. G. and Roussas, G. G. (1980). Asymptotic inference for continuous time semi-Markov processes. *Scand. J. Statist.* 7, 73–79.

Bickel, P. J. (1993). Estimation in semiparametric models. In: *Multivariate Analysis: Future Directions* (C. R. Rao, ed.), 55–73, North-Holland, Amsterdam.

Çinlar, E. (1969). Markov renewal theory. *Adv. in Appl. Probab.* 1, 123–187.

Gill, R. D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 53, 97–116.

Gill, R. D. (1983). Discussion of the papers by Helland and Kurtz. *Bull. Inst. Internat. Statist.* 50 (3), 239–243.

Gordin, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl.* 10, 1174–1176.

Gordin, M. I. and Lifšic, B. A. (1978). The central limit theorem for stationary Markov processes. *Soviet Math. Dokl.* 19, 392–394.

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* 36, 353–370.

Greenwood, P. E. and Wefelmeyer, W. (1994). Nonparametric estimators for Markov step processes. *Stochastic Process. Appl.* 52, 1–16.

Greenwood, P. E. and Wefelmeyer, W. (1995a). Efficiency of empirical estimators for Markov chains. *Ann. Statist.* 23, 132–143.

Greenwood, P. E. and Wefelmeyer, W. (1995b). Maximum likelihood estimator and Kullback–Leibler information in misspecified Markov chain models. Preprint.

Hatori, H. (1966). A limit theorem on (J, X)-processes. *Kōdai Math. Sem. Reports* 18, 317–321.

Höpfner, R. (1993a). On statistics of Markov step processes: representation of log-likelihood ratio processes in filtered local models. *Probab. Theory Related Fields* 94, 375–398.

Höpfner, R. (1993b). Asymptotic inference for Markov step processes: observation up to a random time. *Stochastic Process. Appl.* 48, 295–310.

Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* 86, 105–129.

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 31, 235–253.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

Jain, R. K. (1990). Statistical estimation for semi-Markov models. *Metron* 48, 223–242.

Janssen, J., ed. (1986). *Semi-Markov Models. Theory and Applications.* Plenum Press, New York.

Kartashov, N. V. (1985). Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theory Probab. Math. Statist.* 30, 71–89.

Lagakos, S. W., Sommer, C. J. and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika* 65, 311–317.

Malinovskii, V. K. (1985). On limit theorems for the number of Markov renewals. In: *Stability Problems for Stochastic Models* (V. V. Kalashnikov, V. M. Zolotarev, eds.), 190–222, Lecture Notes in Mathematics 1155, Springer-Verlag, Berlin.

Malinovskii, V. K. (1986). Asymptotic expansions in the central limit theorem for recurrent Markov renewal processes. *Theory Probab. Appl.* 31, 523–526.

Malinovskii, V. K. (1987). Limit theorems for recurrent semi-Markov processes and Markov renewal processes. *J. Soviet Math.* 36, 493–502.

Malinovskii, V. K. (1989). Asymptotic optimality of tests in a hypothesis testing problem for recurrent jump Markov processes. *J. Soviet Math.* 44, 503–510.

Malinovskii, V. K. (1991). On integral and local limit theorems for recursive Markov renewal processes. *J. Soviet Math.* 57, 3286–3301.

Malinovskii, V. K. (1992). Asymptotic optimality of criteria in the problem of testing hypotheses for a recurrent semi-Markov process. *J. Soviet Math.* 59, 955–959.

McLean, R. A. and Neuts, M. F. (1967). The integral of a step function defined on a semi-Markov process. *SIAM J. Appl. Math.* 15, 726–737.

Meyn, S. P. and Tweedie, R. L. (1994). *Markov Chains and Stochastic Stability.* Second Printing. Springer-Verlag, London.

Moore, E. H. and Pyke, R. (1968). Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Statist. Math.* 20, 411–424.

Penev, S. (1991). Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plann. Inference* 27, 105–123.

Phelan, M. J. (1990a). Bayes estimation from a Markov renewal process. *Ann. Statist.* 18, 603–616.

Phelan, M. J. (1990b). Estimating the transition probability from censored Markov renewal processes. *Statist. Probab. Lett.* 10, 43–47.

Pyke, R. (1961a). Markov renewal processes: definitions and preliminary properties. *Ann. Math. Statist.* 32, 1231–1242.

Pyke, R. (1961b). Markov renewal processes with finitely many states. *Ann. Math. Statist.* 32, 1243–1259.

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

Pyke, R. and Schaufele, R. (1966). The existence and uniqueness of stationary measures for Markov renewal processes. *Ann. Math. Statist.* 37, 1439–1462.

Taga, Y. (1963). On the limiting distributions in Markov renewal processes with finitely many states. *Ann. Inst. Statist. Math.* 15, 1–10.

Tomko, J. (1989). On the theory of semi–Markov processes with common phase space. *Theory Probab. Appl.* 34, 275–289.

Voelkel, J. G. and Crowley, J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Ann. Statist.* 12, 142–160.