Efficient estimation for semiparametric semi-Markov processes

Priscilla E. Greenwood
Arizona State University

Ursula U. Müller
Universität Bremen

Wolfgang Wefelmeyer
Universität Siegen

Abstract

We consider semiparametric models of semi-Markov processes with arbitrary state space. Assuming that the process is geometrically ergodic, we characterize efficient estimators, in the sense of Hájek and Le Cam, for arbitrary real-valued smooth functionals of the distribution of the embedded Markov renewal process. We construct efficient estimators of the parameter and of linear functionals of the distribution. In particular we treat the two cases in which we have a parametric model for the transition distribution of the embedded Markov chain and an arbitrary conditional distribution of the inter-jump times, and vice versa.

1 Introduction

Suppose we observe a semi-Markov process Z_t, $t \geq 0$, with embedded Markov renewal process $(X_0, T_0), (X_1, T_1), \dots$, on a time interval $0 \leq t \leq n$. The transition distribution of the Markov renewal process factors as

$$D(x, dy, ds) = Q(x, dy)R(x, y, ds),$$

where $Q(x, dy)$ is the transition distribution of the embedded Markov chain $X_0, X_1, \dots$, and $R(x, y, ds)$ is the conditional distribution of the inter-jump times $S_j = T_j - T_{j-1}$ given $X_{j-1} = x$ and $X_j = y$. We assume that the embedded Markov chain is geometrically ergodic. We write $P(dx, dy, ds)$ for the joint stationary law of (X_{j-1}, X_j, S_j), and $P_1(dx)$ and $P_2(dx, dy)$ for its marginals. We are interested in estimation of functionals of Q and R. Our results hold also for observations $(X_0, T_0), \dots, (X_n, T_n)$ of the embedded Markov renewal process. For discrete state space and the fully parametric or nonparametric cases, the asymptotic distribution of maximum likelihood estimators, Bayes estimators, and empirical estimators has been studied by Taga [31], Pyke and Schaufele [27], Hatori [13], McLean and Neuts [20], Moore and Pyke [22], Ouhbi and Limnios [23, 24] and, with censoring, by Lagakos, Sommer and Zelen [18], Gill [5], Voelkel and Crowley [32] and Phelan [25, 26].

We focus primarily on *semiparametric* models and on the construction of *efficient* estimators. The simplest semiparametric models are obtained by specifying a parametric form for one of the factors of $D(x, dy, ds) = Q(x, dy)R(x, y, ds)$. In one case, we assume a parametric model Q_ϑ for the transition distribution of the embedded Markov chain and leave the conditional distribution of the inter-jump times unspecified (model Q). In a second case, we assume a parametric model R_ϑ for the conditional distribution of the inter-jump times and leave the transition distribution of the embedded Markov chain unspecified (model R).

The estimating problems connected with these two models are specific to the semi-Markov setting; in particular, they have no non-trivial counterpart for Markov chains. To keep the paper readable and short, we concentrate on the two simple models above. More general models, involving possibly infinite-dimensional parameters, perhaps on both factors simultaneously, could be treated along the same lines.

We want to estimate ϑ and linear functionals of the form

$$Ef(X_{j-1}, X_j, S_j) = \iint P_1(dx)Q(x, dy) \int R(x, y, ds)f(x, y, s) = P_1QRf,$$

with $Q = Q_\vartheta$ or $R = R_\vartheta$ parametric. Interesting applications are estimation of probabilities $P(X_{j-1} \in A, X_j \in B, S_j \le c)$, $P(X_{j-1} \in A, X_j \in B)$ and $P(X_{j-1} \in A)$, and of ratios $P(S_j \le c \mid X_{j-1} \in A, X_j \in B)$ and $P(X_j \in B \mid X_{j-1} \in A)$. We can also treat expectations ES_j and conditional expectations $E(S_j \mid X_{j-1} \in A, X_j \in B)$ and $E(X_j \mid X_{j-1} \in A)$.

Natural estimators for ϑ are the maximum likelihood estimators based on the conditional distributions Q_ϑ or R_ϑ. We show that they are efficient in our two models. In particular, they are adaptive in the sense that knowing the nonparametric factor of $Q(x, dy)R(x, y, ds)$ cannot give estimators with smaller asymptotic variance. A natural estimator for a linear functional $Ef(X_{j-1}, X_j, S_j)$ is the empirical estimator

$$\frac{1}{N_n} \sum_{j=1}^{N_n} f(X_{j-1}, X_j, S_j),$$

where $N_n = \max\{j : T_j \le n\}$. Greenwood and Wefelmeyer [8] have shown that this estimator is efficient in the fully nonparametric semi-Markov model; see also Greenwood and Wefelmeyer [7] for Markov step processes. We construct better, efficient, estimators for our two semiparametric models Q and R.

For our first model, the functional $Ef(X_{j-1}, X_j, S_j)$ can be written

$$Ef(X_{j-1}, X_j, S_j) = P_{1\vartheta} Q_\vartheta R f = \iint P_{1\vartheta}(dx) Q_\vartheta(x, dy) R_{xy} f$$

with $R_{xy} f = \int R(x, y, ds) f(x, y, s)$. To exploit the structure of the model, we use a plug-in estimator, i.e. we replace the conditional expectation Rf by a kernel estimator $\hat{R}f$. By what we refer to as the plug-in principle, we expect that $P_{1\vartheta} Q_\vartheta \hat{R} f$ will converge at the parametric rate $n^{-1/2}$ under appropriate conditions on the kernel and the bandwidth, even though the kernel estimator has a slower rate of convergence. In a second step, we replace the parameter ϑ by an estimator $\hat{\vartheta}$. This results in the estimator $P_{1\hat{\vartheta}} Q_{\hat{\vartheta}} \hat{R} f$ for $Ef(X_{j-1}, X_j, S_j)$. It is efficient if an efficient estimator $\hat{\vartheta}$ is used for ϑ. Related plug-in estimators have been used in other, mainly nonparametric, contexts before. For quadratic functionals of densities with i.i.d. observations see Hall and Marron [12], Bickel and Ritov [2], Eggermont and LaRiccia [4] and the references there. Similar results exist for regression models; see e.g. Goldstein and Messer [6] and Efromovich and Samarov [3]. In semiparametric time series models with independent innovations, the stationary density can be written as a smooth functional of the innovation density and the parameters; $n^{1/2}$-consistent and efficient plug-in estimators are constructed in Saavedra and Cao [28] and Schick and Wefelmeyer [29, 30].

For our second model, the functional $Ef(X_{j-1}, X_j, S_j)$ can be written

$$Ef(X_{j-1}, X_j, S_j) = P_2 R_\vartheta f = \iint P_2(dx, dy) \int R_\vartheta(x, y, ds) f(x, y, s).$$

Here we can estimate the nonparametric part P_2 by the empirical distribution based on the embedded Markov chain. Again we replace ϑ by an estimator $\hat{\vartheta}$. We show that the resulting estimator

$$\frac{1}{N_n} \sum_{j=1}^{N_n} \int R_{\hat{\vartheta}}(X_{j-1}, X_j, ds) f(X_{j-1}, X_j, s)$$

is efficient if $\hat{\vartheta}$ is efficient.

The paper is organized as follows. In Section 2 we state local asymptotic normality for arbitrary semi-Markov models and characterize efficient estimators for smooth functionals on such models. In Section 3 we construct efficient estimators of ϑ and $Ef(X_{j-1}, X_j, S_j)$ for model Q, and in Section 4 for model R. Throughout the paper, the discussion will be informal.

2 Characterization of efficient estimators

In this section we consider general semi-Markov models described by families of distributions $Q(x, dy)$ and $R(x, y, dz)$. To calculate asymptotic variance bounds and characterize efficient estimators, we fix Q and R and introduce a local model at (Q, R) by perturbing Q as $Q_{nu}(x, dy) \doteq Q(x, dy)(1 + n^{-1/2}u(x, y))$ and R as $R_{nv}(x, y, ds) \doteq R(x, y, ds)(1 + n^{-1/2}v(x, y, s))$. Since Q_{nu} and R_{nv} are again conditional distributions, the function u will vary in some subset U_0 of

$$U = \{u \in L_2(P_2) : Q_x u = 0\},$$

and the function v will vary in some subset V_0 of

$$V = \{v \in L_2(P) : R_{xy} v = 0\}.$$

Here $Q_x u = \int Q(x, dy)u(x, y)$ and $R_{xy} v = \int R(x, y, ds)v(x, y, s)$. Similarly, we will write $D_x v = \iint D(x, dy, ds)v(x, y, s)$. The sets U_0 and V_0 are called the *tangent spaces* for Q and R. For simplicity we take them linear and closed. Note that U_0 and V_0 are *orthogonal* subspaces of $L_2(P)$. The perturbations $Q_{nu} \doteq Q(1 + n^{-1/2}u)$ and $R_{nv} \doteq R(1 + n^{-1/2}v)$ are meant in the sense that Q_{nu} and R_{nv} are Hellinger differentiable with derivatives u and v. For appropriate versions in arbitrary Markov step models and in nonparametric semi-Markov models see Höpfner, Jacod and Ladelli [14] and Greenwood and Wefelmeyer [8].

We assume that $\int D(x, dy, \{0\}) = 0$, that the mean inter-jump time $m = ES_j$ is finite, and that the embedded Markov chain is positive Harris recurrent. Then

$$\frac{n}{N_n} \to m \quad \text{a.s.} \tag{2.1}$$

Furthermore, the following law of large numbers and martingale central limit theorem hold. For $f \in L_2(P)$ we have

$$\frac{1}{N_n} \sum_{j=1}^{N_n} f(X_{j-1}, X_j, S_j) \to Pf \quad \text{a.s.,} \tag{2.2}$$

and for $w \in L_2(P)$ with $D_x w = 0$ we have

$$n^{-1/2} \sum_{j=1}^{N_n} w(X_{j-1}, X_j, S_j) \Rightarrow m^{-1/2} L, \tag{2.3}$$

where L is a normal random variable with mean zero and variance Pw^2.

Now write $M^{(n)}$ for the distribution of Z_t, $0 \le t \le n$, if Q and R are in effect, and $M_{uv}^{(n)}$ if Q_{nu} and R_{nv} are. Similarly as in Höpfner, Jacod and Ladelli [14] and Greenwood and Wefelmeyer [8], and using orthogonality of U_0 and V_0, we obtain *local asymptotic normality*: For $u \in U_0$ and $v \in V_0$,

$$\log \frac{dM_{uv}^{(n)}}{dM^{(n)}} = H_n - \frac{1}{2}\sigma^2(u, v) + o_p(1), \tag{2.4}$$

where

$$\begin{aligned} H_n &= n^{-1/2} \sum_{j=1}^{N_n} \big(u(X_{j-1}, X_j) + v(X_{j-1}, X_j, S_j) \big), \\ \sigma^2(u, v) &= m^{-1}(P_2 u^2 + P v^2), \end{aligned}$$

and H_n is asymptotically normal with mean zero and variance σ^2.

We want to estimate functionals of (Q, R). A real-valued functional $\varphi(Q, R)$ is said to be *differentiable* at (Q, R) with *gradient* (g, h) if $g \in U$, $h \in V$, and the functional has a linear approximation in terms of the inner product from the LAN-norm,

$$n^{1/2}(\varphi(Q_{nu}, R_{nv}) - \varphi(Q, R)) \to m^{-1}(P_2(ug) + P(vh)), \quad u \in U_0, v \in V_0. \tag{2.5}$$

The projection (g_0, h_0) of (g, h) onto $U_0 \times V_0$ is called the *canonical* gradient of φ. An estimator $\hat{\varphi}$ is called *regular* for φ at (Q, R) with *limit* L if L is a random variable such that

$$n^{1/2}(\hat{\varphi} - \varphi(Q_{nu}, R_{nv})) \Rightarrow L \quad \text{under } M_{uv}^{(n)}, \quad u \in U_0, v \in V_0. \tag{2.6}$$

The convolution theorem of Hájek [11] and Le Cam [19] says that L is distributed as the convolution of a normal random variable with mean zero and variance $\sigma^2(g_0, h_0) = m^{-1}(P_2 g_0^2 + P h_0^2)$ with another random variable. This justifies calling $\hat{\varphi}$ *efficient* if it has this asymptotic variance.

An estimator $\hat{\varphi}$ is called *asymptotically linear* with *influence function* (a, b) if $a \in U$, $b \in V$, and

$$n^{1/2}(\hat{\varphi} - \varphi(Q, R)) = n^{-1/2} \sum_{j=1}^{N_n} \big(a(X_{j-1}, X_j) + b(X_{j-1}, X_j, S_j) \big) + o_p(1). \tag{2.7}$$

With these definitions, $\hat{\varphi}$ is regular and efficient if and only if it is asymptotically linear with influence function equal to the canonical gradient:

$$n^{1/2}(\hat{\varphi} - \varphi(Q, R)) = n^{-1/2} \sum_{j=1}^{N_n} \big(g_0(X_{j-1}, X_j) + h_0(X_{j-1}, X_j, S_j) \big) + o_p(1). \tag{2.8}$$

A reference for this characterization in the i.i.d. case is in Bickel, Klaassen, Ritov and Wellner [1]; for semi-Markov processes parametrized by D see Greenwood and Wefelmeyer [8].

We point out that the orthogonality of U_0 and V_0 implies that functionals of one of the factors of $Q(x, dy)R(x, y, ds)$ can be estimated adaptively with respect to the other factor in the following sense. Suppose $\varphi(Q, R)$ depends only on Q. Then (2.5) holds with $h = 0$, and the canonical gradient is of the form $(g_0, 0)$. Suppose now that $\hat{\varphi}$ is efficient in a model with R completely unspecified. Then it will remain efficient for any submodel for R, in particular when R is known. The same holds with interchanged roles of Q and R. We apply this observation to estimation of ϑ in models Q and R, Sections 3 and 4.

We will also need a version of the central limit theorem (2.3) for functions that are not conditionally centered. Suppose that the embedded Markov chain is geometrically ergodic in the L_2 sense. For $k \in L_2(P_2)$ define

$$(Ak)(x, y) = \sum_{i=0}^{\infty} (Q_y^i k - Q_x^{i+1}).$$

Set $f_0(x, y, s) = f(x, y, s) - R_{xy} f$. Then

$$\begin{aligned} & n^{-1/2} \sum_{j=1}^{N_n} \big(f(X_{j-1}, X_j, S_j) - P_1 Q R f \big) \\ &= n^{-1/2} \sum_{j=1}^{N_n} \big(ARf(X_{j-1}, X_j) + f_0(X_{j-1}, X_j, S_j) \big) + o_p(1). \end{aligned} \tag{2.9}$$

Note that $Q_x Ak = 0$ for $k \in L_2(P_2)$. For Markov chains, the above martingale approximation goes back to Gordin [9] and Gordin and Lifšic [10]; see Meyn and Tweedie [21], Section 17.4. For semi-Markov processes we refer to Greenwood and Wefelmeyer [8]. From (2.3) we obtain that the above standardized sum is asymptotically normal with variance $m^{-1}(P_2(ARf)^2 + P f_0^2)$.

To calculate gradients of linear functionals $Ef(X_{j-1}, X_j, S_j)$, we need the following perturbation expansion due to Kartashov [15, 16, 17]:

$$n^{1/2}(P_{1nu}Q_{nu}k - P_1 Qk) \to P_2(kBu) = P_2(uAk), \quad k \in L_2(P_2), \tag{2.10}$$

where B is the adjoint of A. We will not need the explicit form of B. The perturbation expansion implies that $\varphi(Q, R) = Pf = P_1 QRf$ is differentiable for $f \in L_2(P)$,

$$n^{1/2}(P_{1nu}Q_{nu}R_{nv}f - P_1 QRf) \to P_2(uARf) + P(vf_0), \quad u \in U_0, v \in V_0. \tag{2.11}$$

Here we have used that U and V are orthogonal. For a proof of (2.11) we refer to Greenwood and Wefelmeyer [8]. Note that there we do not factor D and have local parameters $h(x, y, s)$ which here are written $u(x, y) + v(x, y, s)$.

3 Model Q

In this section we consider model Q, in which we have a parametric family Q_ϑ for Q and leave R unspecified. For simplicity we assume that ϑ is one-dimensional. A natural estimator for ϑ is the maximum likelihood estimator based on Q_ϑ. Suppose $Q_\vartheta(x, dy)$ has density $q_\vartheta(x, y)$ with respect to some dominating measure $\nu_Q(x, dy)$, and that q_ϑ has derivative $\dot{q}_\vartheta$ with respect to ϑ. Write $\lambda_\vartheta = \dot{q}_\vartheta / q_\vartheta$ for the score function. The maximum likelihood estimator $\hat{\vartheta}$ solves the estimating equation

$$\sum_{j=1}^{N_n} \lambda_\vartheta(X_{j-1}, X_j) = 0.$$

A stochastic expansion of $\hat{\vartheta}$ is now obtained by the usual arguments. First we recall two well-known relations for λ_ϑ and $\dot{\lambda}_\vartheta$, namely

$$\begin{aligned} 0 &= \partial_\vartheta(\nu_Q q_\vartheta) = \nu_Q \dot{q}_\vartheta = Q_\vartheta \lambda_\vartheta, \\ 0 &= \partial_\vartheta(Q_\vartheta \lambda_\vartheta) = \partial_\vartheta \nu_Q(\lambda_\vartheta q_\vartheta) = \nu_Q(\lambda_\vartheta \dot{q}_\vartheta + \dot{\lambda}_\vartheta q_\vartheta) = Q_\vartheta(\lambda_\vartheta^2 + \dot{\lambda}_\vartheta). \end{aligned}$$

Write $P_{2\vartheta} = P_{1\vartheta} \otimes Q_\vartheta$ and let $I_\vartheta = P_{2\vartheta}\lambda_\vartheta^2$ denote *Fisher information.* We obtain from the second relation that $I_\vartheta = -P_{2\vartheta}\dot{\lambda}_\vartheta$. From the law of large numbers (2.2) and (2.1) we obtain by Taylor expansion that $\hat{\vartheta}$ is asymptotically linear with influence function $(mI_\vartheta^{-1}\lambda_\vartheta, 0)$:

$$n^{1/2}(\hat{\vartheta} - \vartheta) = mI_\vartheta^{-1} n^{-1/2} \sum_{j=1}^{N_n} \lambda_\vartheta(X_{j-1}, X_j) + o_p(1). \tag{3.1}$$

From the martingale central limit theorem (2.3) we conclude that $n^{1/2}(\hat\vartheta - \vartheta)$ is asymptotically normal with variance mI_ϑ^{-1}.

To prove semiparametric efficiency of $\hat\vartheta$, we must interpret ϑ as a functional of (Q, R) through $\varphi(Q_\vartheta, R) = \vartheta$. The local model for Q_ϑ is obtained by perturbing ϑ as $\vartheta_{na} = \vartheta + n^{-1/2}a$ and Q_ϑ as $Q_{\vartheta_{na}} \doteq Q_\vartheta(1 + n^{-1/2}a\lambda_\vartheta)$. Hence the tangent space U_0 for Q consists of all functions of the form $a\lambda_\vartheta$, $a \in \mathbf{R}$. The canonical gradient (g_0, h_0) of ϑ is therefore of the form $(a_0\lambda_\vartheta, 0)$, where a_0 is determined from (2.5) by

$$a = m^{-1}P_{2\vartheta}(a\lambda_\vartheta a_0\lambda_\vartheta) = aa_0m^{-1}I_\vartheta, \quad a \in \mathbf{R}.$$

This gives $a_0 = mI_\vartheta^{-1}$ and $g_0 = mI_\vartheta^{-1}\lambda_\vartheta$. Since $\hat\vartheta$ has influence function $(mI_\vartheta^{-1}\lambda_\vartheta, 0)$ by (3.1), it is efficient by characterization (2.8). Note that $\hat\vartheta$ is *adaptive* with respect to R in the sense that it remains efficient even if we know R.

Now we consider estimation of a linear functional $Ef(X_{j-1}, X_j, S_j) = P_{2\vartheta}Rf$ with $f \in L_2(P_{2\vartheta} \otimes R)$. A natural estimator is the empirical estimator

$$\frac{1}{N_n}\sum_{j=1}^{N_n} f(X_{j-1}, X_j, S_j).$$

We have $ARf \in U$ and $f_0 \in V$. From (2.9) we obtain that the empirical estimator is asymptotically linear with influence function $(mARf(x,y), mf_0(x,y,s))$ and asymptotic variance $m(P_{2\vartheta}(ARf)^2 + P_{2\vartheta}Rf_0^2)$. If nothing were known about Q, the empirical estimator would be efficient; see Greenwood and Wefelmeyer [8]. Since we have assumed a parametric model Q_ϑ, we can construct better estimators exploiting the structure of the model. We assume that the state space is the real line, and that P has Lebesgue density p. Let $p_{1\vartheta}$ and q_ϑ be the densities of $P_{1\vartheta}$ and Q_ϑ. Then $p_{2\vartheta}(x,y) = p_{1\vartheta}(x)q_\vartheta(x,y)$ is the density of $P_{2\vartheta}$. We write $Rf = a/p_{2\vartheta}$ with

$$a(x,y) = \int p(x,y,s)ds\, f(x,y,s)$$

and estimate Rf by $\hat{R}f = \hat{a}/\hat{p}_2$ with kernel estimators

$$\begin{aligned}
\hat{a}(x,y) &= \frac{1}{N_n}\sum_{j=1}^{N_n}\frac{1}{b^2}k\Big(\frac{x - X_{j-1}}{b}, \frac{y - X_j}{b}\Big)f(x,y,S_j),\\
\hat{p}_2(x,y) &= \frac{1}{N_n}\sum_{j=1}^{N_n}\frac{1}{b^2}k\Big(\frac{x - X_{j-1}}{b}, \frac{y - X_j}{b}\Big),
\end{aligned}$$

where k is a mean zero density and $b = b_n$ is a bandwidth that tends to zero at a rate to be determined later. Our estimator for $P_{2\vartheta}Rf$ is $P_{2\hat{\vartheta}}\hat{R}f$ with $\hat{\vartheta}$ a $n^{1/2}$-consistent estimator of ϑ. We prove that it is asymptotically linear if f is differentiable. Under appropriate smoothness assumptions on p, a modified proof will cover discontinuous f, in particular indicator functions. To calculate the influence function of $P_{2\hat{\vartheta}}\hat{R}f$, we write

$$\hat{R}f = Rf + \frac{\hat{a} - a}{\hat{p}_2} - \frac{\hat{p}_2 - p_{2\vartheta}}{\hat{p}_2} Rf.$$

Then our estimator is approximated as

$$P_{2\hat{\vartheta}}\hat{R}f \doteq P_{2\hat{\vartheta}}Rf + \iint dxdy(\hat{a}(x,y) - a(x,y)) - \iint dxdy\,(\hat{p}_2(x,y) - p_{2\vartheta}(x,y))R_{xy}f.$$

Let $b = n^{-1/4}$. Since the kernel k integrates to one and has mean zero, a change of variables $u = (x - X_{j-1})/b$ and $v = (x - X_j)/b$ and a Taylor expansion give

$$\begin{aligned} \iint dxdy\,\hat{a}(x,y) &= \frac{1}{N_n}\sum_{j=1}^{N_n}\iint dudv\,k(u,v)f(X_{j-1} + bu, X_j + bv, S_j) \\ &= \frac{1}{N_n}\sum_{j=1}^{N_n} f(X_{j-1}, X_j, S_j) + o_p(n^{-1/2}). \end{aligned} \tag{3.2}$$

Similarly,

$$\begin{aligned} &\iint dxdy\,\hat{p}_2(x,y)R_{xy}f \\ &= \frac{1}{N_n}\sum_{j=1}^{N_n}\iint dudv\,k(u,v)\int R(X_{j-1} + bu, X_j + bv, ds)f(X_{j-1} + bu, X_j + bv, s) \\ &= \frac{1}{N_n}\sum_{j=1}^{N_n} R_{X_{j-1},X_j}f + o_p(n^{-1/2}). \end{aligned} \tag{3.3}$$

With the notation $f_0(x,y,s) = f(x,y,s) - R_{xy}f$, these two expansions lead to

$$P_{2\hat{\vartheta}}\hat{R}f = P_{2\hat{\vartheta}}Rf + \frac{1}{N_n}\sum_{j=1}^{N_n} f_0(X_{j-1}, X_j, S_j) + o_p(n^{-1/2}). \tag{3.4}$$

It remains to expand $P_{2\hat{\vartheta}}Rf$. With $Q_{\vartheta_{na}} \doteq Q_\vartheta(1 + n^{-1/2}a\lambda_\vartheta)$ and the perturbation expansion (2.10) for $u = a\lambda_\vartheta$ and $a = n^{1/2}(\hat{\vartheta} - \vartheta)$, a Taylor expansion gives

$$P_{2\hat{\vartheta}}Rf = P_{2\vartheta}Rf + P_{2\vartheta}(\lambda_\vartheta ARf)(\hat{\vartheta} - \vartheta) + o_p(n^{-1/2}). \tag{3.5}$$

Suppose now that $\hat\vartheta$ is efficient. Then it has influence function $(mI_\vartheta^{-1}\lambda_\vartheta, 0)$ by (3.1). Together with (3.4) and (3.5) we obtain that

$$
\begin{aligned}
&n^{1/2}(P_{2\hat\vartheta}\hat R f - P_{2\vartheta} R f)\\
&= mn^{-1/2}\sum_{j=1}^{N_n}\big(I_\vartheta^{-1}P_{2\vartheta}(\lambda_\vartheta ARf)\lambda_\vartheta(X_{j-1},X_j) + f_0(X_{j-1},X_j,S_j)\big) + o_p(1).
\end{aligned}
$$

Hence by (2.3) our estimator is asymptotically normal with variance

$$m\big(I_\vartheta^{-1}(P_{2\vartheta}(\lambda_\vartheta ARf))^2 + P_{2\vartheta}Rf_0^2\big).$$

Note that by the Cauchy–Schwarz inequality,

$$I_\vartheta^{-1}(P_{2\vartheta}(\lambda_\vartheta ARf))^2 \le P_{2\vartheta}(ARf)^2.$$

Since the empirical estimator has asymptotic variance $m(P_{2\vartheta}(ARf)^2 + P_{2\vartheta}Rf_0^2)$, our estimator is better unless ARf is proportional to λ_ϑ.

Now we prove that our estimator $P_{2\hat\vartheta}\hat R f$ is efficient. By the characterization (2.8) of efficient estimators, we must show that the influence function of $P_{2\hat\vartheta}\hat R f$ equals the canonical gradient of the functional $\varphi(Q,R) = P_{2\vartheta}Rf$. Let $\vartheta_{na} = \vartheta + n^{-1/2}a$ and $R_{nv} \doteq R(1+n^{-1/2}v)$. Then $Q_{\vartheta_{na}} \doteq Q_\vartheta(1+n^{-1/2}a\lambda_\vartheta)$, and the perturbation expansion (2.11) implies

$$n^{1/2}(P_{2\vartheta_{na}}R_{nv}f - P_{2\vartheta}Rf) \to aP_{2\vartheta}(\lambda_\vartheta ARf) + P_{2\vartheta}R(vf_0), \quad a\in\mathbf{R}, v\in V.$$

Since R is unspecified and hence the tangent space V_0 for R is V, the canonical gradient of $P_{2\vartheta}Rf$ is of the form $(a_0\lambda_\vartheta, mf_0)$, where a_0 is determined from (2.5) by

$$aP_{2\vartheta}(\lambda_\vartheta ARf) = aa_0m^{-1}I_\vartheta, \quad a\in\mathbf{R}.$$

This gives $a_0 = mI_\vartheta^{-1}P_{2\vartheta}(\lambda_\vartheta ARf)$. Hence $P_{2\hat\vartheta}\hat R f$ is efficient by characterization (2.8).

We end this section with some comments. If we set $f(x,y,s) = s$, we obtain an efficient estimator for $P_{2\vartheta}Rf = ES_j = m$, the mean inter-jump time. If the inter-jump time distribution does not depend on the states, then our estimator is asymptotically equivalent to the empirical estimator $\frac{1}{N_n}\sum_{j=1}^{N_n} S_j$.

If the state space is discrete, we can replace $\hat R = \hat a/\hat p_2$ by the simpler estimator

$\overline{R}f = \overline{a}/\overline{p}_2$ with

$$\overline{a}(x,y) = \frac{1}{N_n}\sum_{j=1}^{N_n} 1(X_{j-1}=x, X_j=y) f(x,y,S_j),$$
$$\overline{p}_2(x,y) = \frac{1}{N_n}\sum_{j=1}^{N_n} 1(X_{j-1}=x, X_j=y).$$

The analysis of $P_{2\hat{\vartheta}}\overline{R}f$ then simplifies in (3.2) and (3.3). Some examples would be estimation of $P(a,b,(-\infty,c])$, $P_{2\vartheta}(a,b)$, $P_1(a)$ and of ratios $R(a,b,(-\infty,c])$ and $Q(a,b)$.

4 Model R

In this section we consider model R, in which we have a parametric family R_ϑ for R and leave Q unspecified. Again we assume that ϑ is one-dimensional. We proceed as in Section 3. A natural estimator for ϑ is the maximum likelihood estimator based on R_ϑ. We assume that $R_\vartheta(x,y,ds)$ has density $r_\vartheta(x,y,s)$ with respect to some dominating measure $\nu_R(x,y,ds)$, and write $\mu_\vartheta = \dot{r}_\vartheta/r_\vartheta$ for the score function. We have $R_\vartheta\mu_\vartheta = 0$ and $R_\vartheta(\mu_\vartheta^2 + \dot{\mu}_\vartheta) = 0$. In particular, the *Fisher information* $J_\vartheta = P_2 R_\vartheta \mu_\vartheta^2$ equals $-P_2 R_\vartheta \dot{\mu}_\vartheta$. The maximum likelihood estimator solves the estimating equation

$$\sum_{j=1}^{N_n} \mu_\vartheta(X_{j-1}, X_j, S_j) = 0.$$

As in Section 3 we obtain that $\hat{\vartheta}$ is asymptotically linear, now with influence function $(0, mJ_\vartheta^{-1}\mu_\vartheta)$:

$$n^{1/2}(\hat{\vartheta} - \vartheta) = mJ_\vartheta^{-1} n^{-1/2} \sum_{j=1}^{N_n} \mu_\vartheta(X_{j-1}, X_j, S_j) + o_p(1). \tag{4.1}$$

Hence $n^{1/2}(\hat{\vartheta} - \vartheta)$ is asymptotically normal with variance mJ_ϑ^{-1}.

To prove efficiency of $\hat{\vartheta}$, we interpret ϑ as a functional of (Q,R) through $\varphi(Q, R_\vartheta) = \vartheta$. The local model for R_ϑ is described by perturbing ϑ as $\vartheta_{na} = \vartheta + n^{-1/2}a$ and R_ϑ as $R_{\vartheta_{na}} \doteq R_\vartheta(1 + n^{-1/2}a\mu_\vartheta)$. So the tangent space for R consists of all functions of the form $a\mu_\vartheta$, $a \in \mathbf{R}$, and the canonical gradient (g_0, h_0) of ϑ is of the form $(0, a_0\mu_\vartheta)$, where a_0 is determined from (2.5) by

$$a = m^{-1} P_2 R_\vartheta (a\mu_\vartheta a_0 \mu_\vartheta) = a a_0 m^{-1} J_\vartheta, \quad a \in \mathbf{R}.$$

This gives $a_0 = mJ_\vartheta^{-1}$ and $h_0 = mJ_\vartheta^{-1}\mu_\vartheta$. Since $\hat\vartheta$ is asymptotically linear with influence function $(0, mJ_\vartheta^{-1}\mu_\vartheta)$, it is efficient by characterization (2.8) and adaptive with respect to Q.

To estimate $Ef(X_{j-1}, X_j, S_j) = P_2R_\vartheta f$, we can again use the empirical estimator. However, a better estimator is

$$\hat P_2 R_{\hat\vartheta} f = \frac{1}{N_n}\sum_{j=1}^{N_n}\int R_{\hat\vartheta}(X_{j-1}, X_j, ds) f(X_{j-1}, X_j, s).$$

Here $\hat P_2$ stands for the empirical distribution

$$\frac{1}{N_n}\sum_{j=1}^{N_n}\delta_{(X_{j-1},X_j)}(dx, dy),$$

where $\delta_{(X_{j-1},X_j)}$ is the one-point distribution on (X_{j-1}, X_j). With $R_{\vartheta_{na}} \doteq R_\vartheta(1 + n^{-1/2}a\mu_\vartheta)$ and $a = n^{1/2}(\hat\vartheta - \vartheta)$, a Taylor expansion gives

$$\begin{aligned}
\hat P_2 R_{\hat\vartheta} f &= P_2R_\vartheta f + \hat P_2 R_{\hat\vartheta} f - P_2 R_{\hat\vartheta} f + P_2 R_{\hat\vartheta} f - P_2 R_\vartheta f \\
&= P_2R_\vartheta f + \frac{1}{N_n}\sum_{j=1}^{N_n}\int R_\vartheta(X_{j-1}, X_j, ds) f(X_{j-1}, X_j, s) - P_2R_\vartheta f \\
&\quad + P_2R_\vartheta(\mu_\vartheta f)(\hat\vartheta - \vartheta) + o_p(n^{-1/2}).
\end{aligned}$$

Since $R_\vartheta\mu_\vartheta = 0$, we have $P_2R_\vartheta(\mu_\vartheta f) = P_2R_\vartheta(\mu_\vartheta f_0)$ and hence

$$\begin{aligned}
&n^{1/2}(\hat P_2 R_{\hat\vartheta} f - P_2R_\vartheta f) \\
&= mn^{-1/2}\sum_{j=1}^{N_n}\Big(\int R_\vartheta(X_{j-1}, X_j, ds) f(X_{j-1}, X_j, s) - P_2R_\vartheta f\Big) \\
&\quad + P_2R_\vartheta(\mu_\vartheta f_0)(\hat\vartheta - \vartheta) + o_p(1).
\end{aligned}$$

Suppose that $\hat\vartheta$ is efficient for ϑ. Then $\hat\vartheta$ is asymptotically linear with influence function $(0, mJ_\vartheta^{-1}\mu_\vartheta)$; see (4.1). From the martingale approximation (2.9) we see that $\hat P_2 R_{\hat\vartheta} f$ then has influence function $(mAR_\vartheta f, mJ_\vartheta^{-1}P_2R_\vartheta(\mu_\vartheta f_0)\mu_\vartheta)$. Hence $\hat P_2 R_{\hat\vartheta} f$ is asymptotically normal with variance

$$m\big(P_2(AR_\vartheta f)^2 + J_\vartheta^{-1}(P_2R_\vartheta(\mu_\vartheta f_0))^2\big).$$

Note that by the Cauchy–Schwarz inequality,

$$J_\vartheta^{-1}(P_2R_\vartheta(\mu_\vartheta f_0))^2 \le P_2R_\vartheta f_0^2.$$

Hence our estimator is better than the empirical estimator unless f_0 is proportional to μ_ϑ.

Now we prove that $\hat{P}_2 R_{\hat{\vartheta}} f$ is efficient. Let $\vartheta_{na} = \vartheta + n^{-1/2} a$ and $Q_{nu} \doteq Q(1+n^{-1/2}u)$. Then $R_{\vartheta_{na}} \doteq R_\vartheta(1 + n^{-1/2} a \mu_\vartheta)$, and (2.11) implies

$$n^{1/2}(P_{2nu} R_{\vartheta_{na}} f - P_2 R_\vartheta f) \to P_2(uAR_\vartheta f) + aP_2 R_\vartheta(\mu_\vartheta f_0), \quad u \in U, a \in \mathbf{R}.$$

Since Q is unspecified and hence the tangent space U_0 for Q is U, the canonical gradient of $P_2 R_\vartheta f$ is of the form $(mAR_\vartheta f, a_0 \mu_\vartheta)$, where a_0 is determined by

$$aP_2 R_\vartheta(\mu_\vartheta f_0) = a a_0 m^{-1} J_\vartheta, \quad a \in \mathbf{R}.$$

This gives $a_0 = m J_\vartheta^{-1} P_2 R_\vartheta(\mu_\vartheta f_0)$. Hence $\hat{P}_2 R_{\hat{\vartheta}} f$ is efficient by characterization (2.8).

For example, if we set $f(x, y, s) = s$, we obtain an efficient estimator

$$\frac{1}{N_n} \sum_{j=1}^{N_n} \int R_{\hat{\vartheta}}(X_{j-1}, X_j, ds)s$$

of the mean inter-jump time $m = ES_j$. It is better than the empirical estimator $\frac{1}{N_n} \sum_{j=1}^{N_n} S_j$ unless $s - \int R_\vartheta(x, y, ds)s$ is proportional to $\mu_\vartheta(x, y, s)$. If the inter-jump time distribution does not depend on the states, i.e. $R_\vartheta(x, y, ds) = R_\vartheta(ds)$, then our estimator is equivalent to the simpler estimator $\int R_{\hat{\vartheta}}(ds)s$, which is better than the empirical estimator $\frac{1}{N_n} \sum_{j=1}^{N_n} S_j$ unless $s - \int R_\vartheta(x, y, ds)s$ is proportional to $\mu_\vartheta(s)$, i.e. if the inter-jump time distribution R_ϑ is exponential with scale parameter ϑ.

Acknowledgment. Work supported by NSERC, Canada.

References

[1] Bickel, P. J.; Klaassen, C. A. J.; Ritov, Y.; Wellner, J. A., *Efficient and Adaptive Estimation for Semiparametric Models*. Springer: New York, 1998.

[2] Bickel, P. J.; Ritov, Y., Estimating integrated squared density derivatives: Sharp best order of convergence estimates. Sankhya Ser. A, 50 (1988) 381–393.

[3] Efromovich, S.; Samarov, A., Adaptive estimation of the integral of squared regression derivatives. Scand. J. Statist., 27 (2000) 335–351.

[4] Eggermont, P. P. B.; LaRiccia, V. N., *Maximum Penalized Likelihood Estimation, Vol I, Density Estimation*. Springer: New York, 2001.

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

[6] Goldstein, L.; Messer, K., Optimal plug-in estimators for nonparametric functional estimation. Ann. Statist., 20 (1992) 1306–1328.

[7] Greenwood P. E.; Wefelmeyer, W., Nonparametric estimators for Markov step processes. Stochastic Process. Appl., 52 (1994) 1-16.

[8] Greenwood P. E.; Wefelmeyer, W., Empirical estimators for semi-Markov processes. Math. Meth. Statist., 5 (1996) 299-315.

[9] Gordin, M. I., The central limit theorem for stationary processes. Soviet Math. Dokl., 10 (1969) 1174-1176.

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

[11] Hájek, J., A characterization of limiting distributions of regular estimates. Z. Wahrsch. Verw. Gebiete, 14 (1970) 323–330.

[12] Hall, P.; Marron, J. S., Estimation of integrated squared density derivatives. Statist. Probab. Lett., 6 (1987) 109–115.

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

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

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

[16] Kartashov, N. V., Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. Theory Probab. Appl., 30 (1985b) 247–259.

[17] Kartashov, N. V., *Strong Stable Markov Chains*. VSP: Utrecht, 1996.

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

[19] Le Cam, L., Limits of experiments. Proc. Sixth Berkeley Symp. Math. Statist. Probab., l (1971) 245–261.

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

[21] Meyn, S. P.; Tweedie, R. L., *Markov Chains and Stochastic Stability*. Springer: London, 1993.

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

[23] Ouhbi, L.; Limnios N., Nonparametric estimation for semi-Markov kernels with applications to reliability analysis. Appl. Stochastic Models Data Anal., 12 (1996) 209–220.

[24] Ouhbi, L.; Limnios N., Nonparametric estimation for semi-Markov processes based on its hazard rate functions. Stat. Inference Stoch. Process., 2 (1999) 151–173.

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

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

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

[28] Saavedra, A.; Cao, R., On the estimation of the marginal density of a moving average process. Canad. J. Statist., 28 (2000) 799–815.

[29] Schick, A.; Wefelmeyer, W., Root n consistent and optimal density estimators for moving average processes. Technical Report, Department of Mathematical Sciences, Binghamton University, 2002a. http://math.binghamton.edu/anton/preprint.html.

[30] Schick, A.; Wefelmeyer, W., Functional convergence and optimality of plug-in estimators for stationary densities of moving average processes. Technical Report, Department of Mathematical Sciences, Binghamton University, 2002b. http://math.binghamton.edu/anton/preprint.html.

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

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