ESTIMATING JOINT DISTRIBUTIONS OF MARKOV CHAINS

Anton Schick and Wolfgang Wefelmeyer

Binghamton University and Universität-GH Siegen

Abstract. Suppose we observe a stationary Markov chain with unknown transition distribution. The empirical estimator for the expectation of a function of two successive observations is known to be efficient. For reversible Markov chains, an appropriate symmetrization is efficient. For functions of more than two arguments, these estimators cease to be efficient. We determine the influence function of efficient estimators of expectations of functions of several observations, both for completely unknown and for reversible Markov chains. We construct simple efficient estimators in both cases.

AMS 1991 subject classification: Primary 62G05, 62G20; Secondary 62M05, 62G07

Keywords and phrases: Stationary Markov chain; drift condition; V-uniform ergodicity; efficient estimation; reversibility; empirical estimator.

Version: May 24, 1999

1. Introduction

To begin let $X_0, \dots, X_n$ be observations from a stationary time series. We want to estimate the expectation $E[f(X_1, \dots, X_m)]$ of a function f of m arguments. The usual estimator is the empirical estimator

$$E_n f = \frac{1}{n+2-m} \sum_{j=m-1}^{n} f(X_{j-m+1}, \dots, X_j).$$

This estimator is efficient if nothing is known about the distribution of the time series, or if the observations come from a Markov chain of order $m-1$ or higher with unknown transition distribution. This follows by a straightforward extension of Greenwood and Wefelmeyer (1995), who consider first-order Markov chains and show that $E_n f$ is efficient for functions f of two arguments. (For functions of *one* argument, different efficiency proofs for $E_n f$ are in Penev, 1991, and Bickel, 1993.)

The empirical estimator $E_n f$ may cease to be efficient if more is known abot the time series. For example, if the time series is known to be reversible, we can improve $E_n f$ by symmetrization, using

$$E_n f^{sym} = \frac{1}{2(n+2-m)} \sum_{j=m-1}^{n} f^{sym}(X_{j-m+1}, \dots, X_j),$$

where f^{sym} is obtained by symmetrizing with respect to time reversal,

$$f^{sym}(x_1, \dots, x_m) = \tfrac{1}{2}[f(x_1, \dots, x_m) + f(x_m, \dots, x_1)].$$

For first-order Markov chains and functions of two arguments, Greenwood and Wefelmeyer (1999) show that the symmetrized empirical estimator is efficient. (In particular, reversibility carries no information on the expectation of a function of *one* argument.)

If f is a function of m arguments and the order of the chain is known to be less than $m-1$, the empirical estimator is not efficient. For *discrete* state space, it is easy to improve $E_n f$. Suppose, for simplicity, that the chain is of order one, with transition probabilities $Q(x, y)$ and stationary probabilities $\pi(x)$. Then

$$E[f(X_1, \dots, X_m)] = \sum_{x_1, \dots, x_m} \pi(x_1) \left(\prod_{i=2}^{m} Q(x_{i-1}, x_i) \right) f(x_1, \dots, x_m),$$

and an efficient estimator is obtained by replacing $\pi(x)$ and $Q(x, y)$ by empirical estimators $E_n(x)$ and $E_n(x, y)/E_n(x)$ with

$$E_n(x) = \frac{1}{n+1} \#\{j : X_j = x\}, \quad E_n(x, y) = \frac{1}{n} \#\{j : X_{j-1} = x, X_j = y\}.$$

The resulting estimator for $E[f(X_1, \dots, X_m)]$ is efficient because $E_n(x)$ and $E_n(x, y)$ are by the result of Greenwood and Wefelmeyer (1995) mentioned above. If the chain is known to be reversible, an efficient estimator is obtained by symmetrizing $E_n(x, y)$.

We mention in passing that such constructions do not carry over to random *fields*. However, Greenwood, McKeague and Wefelmeyer (1999) construct estimators which exploit knowledge of the range of interactions and are better than empirical estimators but not efficient.

We also recall that the empirical estimator $E_n f$ is inefficient for functions of *more than one* argument if the observations are known to be *independent*. Then the joint law of $X_1, \dots, X_m$ is the product of the marginal laws, and $E[f(X_1, \dots, X_m)]$ can be estimated by the von Mises statistic

$$\frac{1}{n^m} \sum_{i_1, \dots, i_m = 1}^{n} f(X_{i_1}, \dots, X_{i_m}).$$

In the nonparametric model, i.e., when the law of X_i is unknown, the von Mises statistic is efficient; see Levit (1974) and Koshevnik and Levit (1976).

Suppose now that $X_0, \dots, X_n$ come from a first-order Markov chain with *arbitrary* state space. We want to find an efficient estimator for an expectation $E[f(X_1, \dots, X_m)]$ with $m > 2$. The above construction for discrete state space does not carry over to continuous state space. For arbitrary state space, we base the construction on a characterization of efficient estimators which is based on Hájek's (1970) convolution theorem. We show in Section 3 that for the full nonparametric model, with nothing known about the transition distribution of the Markov chain, the characterization reads as follows: An estimator T_n is regular and efficient for $E[f(X_1, \dots, X_m)]$ if (and only if) it admits a stochastic approximation of the form

$$n^{1/2}(T_n - E[f(X_1, \dots, X_m)]) = n^{-1/2} \sum_{j=1}^{n} Tf(X_{j-1}, X_j) + o_{P_n}(1), \tag{1.1}$$

where $Tf = Af_1 + f_2 + \cdots + f_m$ with $f_1(x) = E(f(X_1, \dots, X_m)|X_1 = x)$,

$$f_i(x, y) = E(f(X_1, \dots, X_m)|X_{i-1} = x, X_i = y) - E(f(X_1, \dots, X_m)|X_{i-1} = x),$$

for $i = 2, \dots, m$, and where A is the operator which maps a function ϕ on the state space into the function $A\phi$ defined by

$$A\phi(x, y) = \sum_{k=0}^{\infty} [E(\phi(X_k)|X_0 = y) - E(\phi(X_{k+1})|X_0 = x)].$$

The operator A appears in the well-known martingale approximation

$$n^{-1/2} \sum_{j=1}^{n} (\phi(X_j) - E[\phi(X_1)]) = n^{-1/2} \sum_{j=1}^{n} A\phi(X_{j-1}, X_j) + o_{P_n}(1),$$

which goes back to Gordin (1969).

For a reversible Markov chain we have

$$E[f(X_1, \dots, X_m)] = E[f^{sym}(X_1, \dots, X_m)].$$

This is used in Section 4 to show that an estimator T_n is efficient in the submodel of all reversible chains if the stochastic approximation (1.1) holds with f replaced by f^{sym}. In Section 5 we discuss the construction of efficient estimators for $E[f(X_1, \dots, X_m)]$ in the full nonparametric Markov chain model with $m \geq 3$. Relying on the sample splitting technique of Schick (1998) we show that efficient estimators can be constructed if we have appropriate estimators of the functions $E[f(X_1,\dots,X_m) \mid X_{i-1} = x, X_i = y]$, $i = 2,\dots,m$, and $E[f(X_1,\dots,X_m) \mid X_{i-1} = x]$, $i = 3,\dots,m$ out of which the functions $f_1,\dots,f_m$ above are made up. Of special interest are functions of the form $f(X_1,\dots,X_m) = u_1(X_1)\cdot\ldots\cdot u_m(X_m)$. They are treated in detail in Section 6. Efficient estimators in the model of all *reversible* chains are obtained by replacing f by f^{sym}. The proofs of the main results are in Sections 7 and 8.

2. Notation

In this section we collect notation that is used throughout the paper. Let S be a state space with countably generated σ-field $\mathcal{S}$. By a signed kernel we mean a linear combination of Markov kernels. Let ν be a signed measure on $\mathcal{S}$ and K be a signed kernel on $S \times \mathcal{S}$. Let ϕ be a measurable function on S. Then $\nu(\phi)$ denotes the integral $\int \nu(dx)\phi(x)$, and $K\phi$ denotes the function on S defined by

$$K\phi(x) = \int K(x,dy)\phi(y), \quad x \in S,$$

provided the integrals make sense. Further, νK denotes the signed measure on $\mathcal{S}$ defined by

$$\nu K(A) = \int \nu(dx) \int K(x,A), \quad A \in \mathcal{S}.$$

If K_1 and K_2 are signed kernels, then K_1K_2 denotes the signed kernel defined by

$$K_1K_2(x,A) = \int K_1(x,dy)K_2(y,A), \quad A \in \mathcal{S}.$$

We define K^r iteratively by $K^r = K^{r-1}K$, starting with $K^1 = K$.

We will assume the Markov chain to be V-uniformly ergodic; see Meyn and Tweedie (1993, Chapter 16). In connection with this concept, we need the following notation. Let V be a measurable function from S to $[1,\infty)$. Then we let $\mathcal{L}_V$ denote the set of all measurable functions l from S to $\mathbb{R}$ for which

$$|l|_V = \sup_{x\in S} \frac{|l(x)|}{V(x)} < \infty,$$

and let $\mathcal{M}_V$ denote the set of all signed measures μ on $\mathcal{S}$ such that

$$\|\mu\|_V = \sup_{|l|\le V} |\mu(l)| := \sup\{|\mu(l)| : l \in \mathcal{L}_V, |l|_V \le 1\} < \infty.$$

The spaces $\mathcal{L}_V$ and $\mathcal{M}_V$ are Banach spaces for the norms $|\cdot|_V$ and $\|\cdot\|_V$, respectively. Moreover,

$$|l|_V = \sup_{\|\mu\|_V \le 1} |\mu(l)| := \sup\{|\mu(l)| : \mu \in \mathcal{M}_V, \|\mu\|_V \le 1\}, \quad l \in \mathcal{L}_V.$$

In particular, if $V = 1$, then $\mathcal{L}_V$ is the set of all bounded measurable functions endowed with the sup-norm, and $\mathcal{M}_V$ is the set of all signed measures with finite total variation norm. For the signed kernel K we set

$$|||K|||_V = \sup_{|l| \le V} |Kl|_V = \sup_{x \in S} \sup_{|l| \le V} \frac{|Kl(x)|}{V(x)} = \sup_{x \in S} \sup_{|l| \le V} \frac{|\int K(x, dy) l(y)|}{V(x)}.$$

It is easy to check that

$$|||K|||_V = \sup_{\|\mu\|_V \le 1} \|\mu K\|_V.$$

Thus, if $|||K|||_V < \infty$, then the map $l \mapsto Kl$ defines a bounded linear operator on $\mathcal{L}_V$, and the map $\mu \mapsto \mu K$ defines a bounded linear operator on $\mathcal{M}_V$.

3. Characterization of efficient estimators

In this section we describe a characterization of efficient estimators for differentiable functionals in Markov chain models. Let $X_0, \dots, X_n$ be observations from a stationary Markov chain on S with transition distribution Q and stationary distribution π. We write $\pi \otimes Q$ for the joint distribution of (X_0, X_1) and $\pi \otimes Q^{\otimes j}$ for the joint distribution of $(X_0, \dots, X_j)$, $j \ge 2$. Let V be a measurable function from S to $[1, \infty)$. Throughout we impose the following assumptions on the chain.

Assumption 1. *The function V is π-square integrable:*

$$\pi(V^2) < \infty. \tag{3.1}$$

The kernel Q satisfies

$$|||Q|||_{V^2} = \sup_{x \in S} \frac{QV^2(x)}{V^2(x)} < \infty \tag{3.2}$$

and is V-uniformly ergodic:

$$\lim_{j \to \infty} |||Q^j - \Pi|||_V = 0. \tag{3.3}$$

We also need the following assumption on the function f of m arguments whose expectation we want to estimate.

Assumption 2. *There are a finite constant C_f and nonnegative numbers $\alpha_1, \dots, \alpha_m$ satisfying $\alpha_1 + \dots + \alpha_m = 1$ such that*

$$|f(x_1, \dots, x_m)| \le C_f V^{\alpha_1}(x_1) \cdot \ldots \cdot V^{\alpha_m}(x_m), \quad x_1, \dots, x_m \in S.$$

The following remarks give sufficient conditions for, and consequences of, these assumptions.

Remark 1. A sufficient condition for Assumption 1 is the V^2-uniform ergodicity of Q, i.e.,

$$\lim_{j \to \infty} |||Q^j - \Pi|||_{V^2} = 0. \tag{3.4}$$

It follows from Meyn and Tweedie (1993, Section 16) that for an aperiodic chain, V-uniform ergodicity is implied by the following drift condition: For some small set Γ and positive constants $\lambda < 1$ and $b < \infty$,

$$\int Q(x, dy)V(y) \le \lambda V(x) + b\mathbf{1}_\Gamma(x), \quad x \in S.$$

An event Γ in $\mathcal{S}$ is called small if there exist a measure ν on $\mathcal{S}$ with $\nu(\Gamma) > 0$ and a positive integer j such that $Q^j(x, B) \ge \nu(B)$ for all $x \in \Gamma$ and $B \in \mathcal{S}$. In particular, if Q is aperiodic and satisfies the above drift condition with V^2 instead of V, then Q is V^2-uniformly ergodic and thus satisfies Assumption 1.

Remark 2. It follows from the (conditional) moment inequality and (3.2) that

$$QV^\alpha \le MV^\alpha$$

for all $\alpha \in [0, 2]$ and some finite constant M. In view of this, Assumptions 1 and 2 guarantee the square-integrability of $f(X_1, \dots, X_m)$. Indeed, one calculates

$$E[f(X_1, \dots, X_m)^2] = \pi \otimes Q^{\otimes(m-1)}(f^2) \le C_f^2 M^{m-1} \pi(V^2).$$

Remark 3. We need the following stability results for the stationary distribution. We have $|||Q^j - \Pi|||_V = |||(Q - \Pi)^j|||_V$. Hence V-uniform ergodicity (3.3) of Q implies *geometric V-uniform ergodicity* of Q: For positive constants D and $\rho < 1$,

$$|||Q^j - \Pi|||_V \le D\rho^j \quad \text{for all } j. \tag{3.5}$$

Since

$$|||Q - \Pi|||_V = \sup_{\mu \in \mathcal{M}_V, \|\mu\|_V \le 1} \|\mu(Q - \Pi)\|_V,$$

$|||\cdot|||_V$ is an operator norm of the type considered by Kartashov (1985, 1996). It follows from (3.5) that the operator

$$U = I + \sum_{j=1}^{\infty} (Q^j - \Pi) = \sum_{j=0}^{\infty} (Q - \Pi)^j$$

is well defined on $\mathcal{L}_V$ and bounded, and the inverse of $I - Q + \Pi$. Thus Q is uniformly ergodic in the norm $|||\cdot|||_V$ (Kartashov, 1985, Theorem 1) and hence *strongly stable* in this norm (Kartashov, 1985, Theorem 4): Each Markov kernel Q_* in some neighborhood of Q has a unique invariant measure π_*, and $\|\pi_* - \pi\|_V \to 0$ as $|||Q_* - Q|||_V \to 0$.

One can even show the following perturbation expansion of π_* in terms of Q_*,

$$\pi_* - \pi = \pi \sum_{j=1}^{\infty} ((Q_* - Q)U)^j \tag{3.6}$$

if Q_* is close enough to Q. Indeed, since $(\nu - \mu)\Pi = 0$ for any two probability measures ν and μ, we find that $\pi_* - \pi = \pi_*(Q_* - Q) + (\pi_* - \pi)(Q - \Pi)$. Iterating this we arrive at

$$\pi_* - \pi = \pi_*(Q_* - Q) + \sum_{j=1}^{n} \pi_*(Q_* - Q)(Q^j - \Pi) + (\pi_* - \pi)(Q^{n+1} - \Pi), \quad n = 1, 2, \dots.$$

In view of (3.5), we see that the right hand side converges in the $|||\cdot|||_V$-norm to $\pi_*(Q_* - Q)U$. This shows that

$$\pi_* - \pi = \pi_*(Q_* - Q)U.$$

Iterating this yields

$$\pi_* - \pi = \pi \sum_{j=1}^{n} ((Q_* - Q)U)^j + \pi_*((Q_* - Q)U)^{n+1}.$$

Thus, if $|||(Q_* - Q)U|||_V < 1$, we get the desired (3.6). A version of the representation (3.6) for a different norm was already given by Greenwood and Wefelmeyer (1997).

The characterization of regular and efficient estimators is based on a nonparametric version of Hájek's (1970) convolution theorem. We recall that the nonparametric Markov chain model is locally asymptotically normal. As in Greenwood and Wefelmeyer (1999) we take as local parameter space the set

$$H = \{h \in L_2(\pi \otimes Q) : Qh = 0\}.$$

For a function $k(x,y)$ of two arguments, we introduce the conditional centering

$$\Delta k(x,y) = k(x,y) - \int Q(x,dz)k(x,z). \tag{3.7}$$

If $k \in L_2(\pi \otimes Q)$ then $\Delta k \in H$. For each local parameter $h \in H$, we let $\langle Q_{n,h} \rangle$ denote the sequence of transition probabilities defined by

$$Q_{n,h}(x,dy) = Q(x,dy)(1 + n^{-1/2}h_n(x,y)),$$

where

$$h_n(x,y) = \Delta \bar{h}_n(x,y) = \bar{h}_n(x,y) - \int Q(x,dz)\bar{h}_n(x,z), \quad x,y \in S,$$

with $\bar{h}_n = h\mathbf{1}_{\{|h| \le n^{1/8}\}}$. Since $|h_n| \le 2n^{1/8}$ and $Qh_n = 0$, the kernel $Q_{n,h}$ is indeed a probability kernel, and

$$|||Q_{n,h} - Q|||_V \le 2|||Q|||_V \, n^{-3/8}. \tag{3.8}$$

It follows from this inequality and the results of Kartashov (1985) that for large n, each $Q_{n,h}$ has an invariant measure $\pi_{n,h}$, and that

$$\sup_{h \in H} \|\pi_{n,h} - \pi - \pi(Q_{n,h} - Q)U\|_V = O(n^{-3/4}). \tag{3.9}$$

Moreover, it is easy to check that

$$n^{1/2}((1 + n^{-1/2}h_n)^{1/2} - 1) \to \tfrac{1}{2}h \quad \text{in } L_2(\pi \otimes Q).$$

The following nonparametric version of *local asymptotic normality* can now be derived along the lines of Roussas (1965, 1972). For minimal assumptions see Höpfner (1993).

Theorem 1. *Let $P_n = \pi \otimes Q^{\otimes n}$ and $P_{n,h} = \pi_{n,h} \otimes Q_{n,h}^{\otimes n}$ denote the joint distribution of $X_0, \dots, X_n$ under Q and $Q_{n,h}$, respectively. Then*

$$\log \frac{dP_{n,h}}{dP_n}(X_0, \dots, X_n) = n^{-1/2} \sum_{j=1}^{n} h(X_{j-1}, X_j) - \tfrac{1}{2}\pi \otimes Q(h^2) + o_{P_n}(1)$$

and

$$\mathfrak{L}(n^{-1/2} \sum_{j=1}^{n} h(X_{j-1}, X_j) \mid P_n) \implies \mathcal{N}(0, \pi \otimes Q(h^2)),$$

with $\mathcal{N}(a,b)$ the normal distribution with mean a and variance b.

Consider now a submodel of the full nonparametric Markov chain model. The local parameter space H_0 of the submodel is a subset of H. We assume that H_0 is linear. Suppose we want to estimate the value $t(Q)$ of a functional t at Q. Assume that the functional is differentiable in the submodel with *gradient* g at Q, i.e., $g \in L_2(\pi \otimes Q)$ and

$$n^{1/2}\Big(t(Q_{n,h}) - t(Q)\Big) \to \pi \otimes Q(hg) \quad \text{for all } h \in H_0.$$

Of course, g is not uniquely determined, but its projection g_0 onto H_0 is. The projection g_0 is called the *canonical gradient.*

Now consider an estimator (sequence) $\langle T_n \rangle$ of $t(Q)$. The estimator $\langle T_n \rangle$ is said to have an *influence function* ψ *at* Q if $\psi \in H$ and

$$n^{1/2}(T_n - t(Q)) = n^{-1/2} \sum_{j=1}^{n} \psi(X_{j-1}, X_j) + o_{P_n}(1).$$

The estimator $\langle T_n \rangle$ is said to be *regular at* Q in the submodel if there exists a distribution L such that

$$\mathfrak{L}(n^{1/2}(T_n - t(Q_{n,h})) \mid P_{n,h}) \implies L \quad \text{for all } h \in H_0.$$

The convolution theorem implies that the limiting distribution L is a convolution

$$L = \mathcal{N}(0, \pi \otimes Q(g_0^2)) * K,$$

that a regular estimator has least dispersed limiting distribution $\mathcal{N}(0, \pi \otimes Q(g_0^2))$ only if it has an influence function which equals the canonical gradient, and that an estimator with influence function g_0 is regular. In view of this, we call the estimator $\langle T_n \rangle$ *efficient* if it has influence function g_0:

$$n^{1/2}(T_n - t(Q)) = n^{-1/2} \sum_{j=1}^{n} g_0(X_{j-1}, X_j) + o_{P_n}(1). \tag{3.10}$$

Now let f be a function of m arguments fulfilling Assumption 2. We are interested in the functional $t = t_f$ defined by

$$t_f(Q) = E[f(X_1, \dots, X_m)] = \pi \otimes Q^{\otimes(m-1)}(f).$$

Note that π depends on Q. Our first goal is to show that that this functional is differentiable, and to calculate its canonical gradient in the full nonparametric Markov chain model. We first treat the case of a function $f = \phi$ of *one* argument, $m = 1$. For this we need the following additional notation. For a function $\phi \in \mathcal{L}_V$, let $A\phi$ denote the map on S^2 defined by

$$A\phi(x, y) = U\phi(y) - QU\phi(x) = U\phi(y) - \int Q(x, dz) U\phi(z), \quad x, y \in S.$$

Lemma 1. *For $\phi \in \mathcal{L}_V$, the functional $t_\phi(Q) = \pi(\phi)$ is differentiable in the full non-parametric model, with canonical gradient $A\phi$.*

PROOF: From (3.9) we obtain

$$n^{1/2}(t_\phi(Q_{n,h}) - t_\phi(Q)) - \iint \pi(dx)Q(x,dy)h_n(x,y)U\phi(y) \to 0.$$

Since $h_n \to h$ in $L_2(\pi \otimes Q)$ and $U\phi \in L_2(\pi \otimes Q)$, we have

$$\iint \pi(dx)Q(x,dy)h_n(x,y)U\phi(y) \to \iint \pi(dx)Q(x,dy)h(x,y)U\phi(y).$$

Since $Qh = 0$, we can replace $U\phi(y)$ on the right side by $U\phi(y) - QU\phi(x) = A\phi(x,y)$. □

The efficient influence function will involve conditional expectations of functions of more than one argument, going both forward and backward in time. This requires new notation. Write $\bar{Q}$ for the transition distribution of the *reversed* chain, i.e., the unique kernel $\bar{Q}$ such that

$$\pi(dx)Q(x,dy) = \pi(dy)\bar{Q}(y,dx).$$

The operator Q will always act *forward* in time; the operator $\bar{Q}$ will always act *backward* in time. Conditional expectations of functions of several arguments will always be taken on the *rightmost* argument under Q, and on the *leftmost* argument under $\bar{Q}$. In particular, for the function f of m arguments,

$$\begin{aligned} Q^{\otimes(m-i)}f(X_1,\dots,X_i) &= E[f(X_1,\dots,X_m) \mid X_1,\dots,X_i], \quad i = 1,\dots,m, \\ \bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)} &= E[f(X_1,\dots,X_m) \mid X_{i-1},X_i], \quad i = 2,\dots,m. \end{aligned}$$

We use $\odot$ rather than $\otimes$ as an additional reminder that for $\bar{Q}$ the time order is reversed.

Theorem 2. *The functional $t_f(Q) = \pi \otimes Q^{\otimes(m-1)}(f)$ is differentiable in the full non-parametric model, with canonical gradient*

$$g_0 = Tf = AQ^{\otimes(m-1)}f + \sum_{i=2}^{m} \Delta\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f. \tag{3.11}$$

The operator T is a bounded linear operator from $L_2(\pi \otimes Q^{\otimes(m-1)})$ to $L_2(\pi \otimes Q)$.

The proof is in Section 7.

Remark 4. For a function $f = \phi$ of *one* argument, we recover the result of Penev (1991) that the empirical estimator $E_n\phi$ is efficient for $E[\phi(X_1)]$. In this case $g_0 = A\phi$. It follows from the martingale representation of Gordin (1969; see also Meyn and Tweedie, 1993, Section 17.4) that

$$\frac{1}{n}\sum_{j=1}^{n}\phi(X_j) = \pi(\phi) + \frac{1}{n}\sum_{j=1}^{n}A\phi(X_{j-1},X_j) - \frac{1}{n}(U\phi(X_n) - U\phi(X_0)). \tag{3.12}$$

Hence

$$n^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\phi(X_j) - E[\phi(X_1)]\right) = n^{-1/2}\sum_{j=1}^{n}A\phi(X_{j-1},X_j) + o_{P_n}(1).$$

Thus this average has influence function g_0 and is therefore efficient, and so is the empirical estimator

$$E_n\phi = \frac{1}{n+1}\sum_{j=0}^{n}\phi(X_j).$$

Remark 5. For $m = 2$ we recover the result of Greenwood and Wefelmeyer (1995) that the empirical estimator $E_n f$ is efficient. Assume first that $m \geq 2$ and consider the empirical estimator

$$E_n f = \frac{1}{n-m+2}\sum_{j=m-1}^{n} f(X_{j-m+1},\dots,X_j).$$

The martingale representation (3.12) extends immediately to functions of more than one argument,

$$n^{1/2}(E_n f - E[f(X_1,\dots,X_m)]) = n^{-1/2}\sum_{j=m-1}^{n} A_m f(X_{j-m+1},\dots,X_j) + o_{P_n}(1),$$

where

$$\begin{aligned} A_m f(x_1,\dots,x_m) &= AQ^{\otimes(m-1)}f(x_{m-1},x_m) \\ &\quad + \sum_{i=2}^{m}[Q^{\otimes(m-i)}f(x_{m-i+1},\dots,x_m) - Q^{\otimes(m-i+1)}f(x_{m-i+1},\dots,x_{m-1})] \end{aligned}$$

for $x_1,\dots,x_m \in S$. Consequently,

$$\mathfrak{L}(n^{1/2}(E_n f - t_f(Q)) \mid P_n) \implies \mathcal{N}(0, \pi\otimes Q^{\otimes(m-1)}((A_m f)^2)).$$

It is easy to verify that the canonical gradient is the conditional expectation

$$g_0(X_{m-1},X_m) = E[A_m f(X_1,\dots,X_m) \mid X_{m-1},X_m]. \tag{3.13}$$

Thus, for $m = 2$, we have $A_m f = g_0$, and the empirical estimator is efficient in the full model in this case. For $m \geq 3$, the asymptotic variance of the empirical estimator exceeds the information bound $\pi \otimes Q^{\otimes(m-1)}(g_0^2)$ by the amount

$$E[(A_m f(X_1, \dots, X_m) - g_0(X_{m-1}, X_m))^2].$$

Remark 6. The empirical estimator is regular in the full model. To see this, note first that (3.13) implies

$$E[A_m f(X_1, \dots, X_m) h(X_{m-1}, X_m)] = \pi \otimes Q(g_0 h), \quad h \in H.$$

With the aid of Le Cam's Third Lemma, one can now show that

$$\mathfrak{L}(n^{1/2}(E_n f - t_f(Q_{n,h})) \mid P_{n,h}) \implies \mathcal{N}(0, \pi \otimes Q^{\otimes(m-1)}((A_m f)^2)), \quad h \in H.$$

Hence the empirical estimator is regular in the full model.

4. Reversible Markov chains

Let us now treat the case when the Markov chain is known to be reversible, $\bar{Q} = Q$. It follows from Theorem 2.1 of Roberts and Rosenthal (1997) and the V-uniform ergodicity of Q that Q is $L_2(\pi)$-geometrically ergodic:

$$\|Q^n - \varPi\| = \sup\{\pi((Q^n \phi - \varPi\phi)^2) : \phi \in L_2(\pi), \pi(\phi^2) \leq 1\} \leq D\rho^n$$

for some finite D and some $\rho < 1$. This allows us to extend the operators U and A from $\mathcal{L}_V$ to $L_2(\pi)$. For $m = 2$ the operator A_m is

$$A_2 k(x, y) = AQk(x, y) + k(x, y) - Qk(x), \quad x, y \in S,$$

with $Qk(x) = \int Q(x, dz)k(x, z)$. Our A_2 corresponds to the operator A used by Greenwood and Wefelmeyer (1999). Note that $A_2 k = A\phi$ if $k(x, y) = \phi(x)$ or $k(x, y) = \phi(y)$ or $k(x, y) = \frac{1}{2}(\phi(x) + \phi(y))$.

Let B denote the adjoint of A_2, and let H^{rev} be the subset of all $h \in H$ for which Bh is symmetric, i.e., $Bh(x, y) = Bh(y, x)$. As was shown by Greenwood and Wefelmeyer (1999) in their proof of Lemma 3, for every $h \in H^{rev}$ there exist transition distributions $Q_{n,h}$ with invariant measures $\pi_{n,h}$ such that the corresponding Markov chain is reversible, i.e., $\bar{Q}_{n,h} = Q_{n,h}$, or equivalently,

$$\pi_{n,h}(dx) Q_{n,h}(x, dy) = \pi_{n,h}(dy) Q_{n,h}(y, dx),$$

and such that $Q_{n,h}$ has local parameter h, i.e.,

$$Q_{n,h}(x, dy) = Q(x, dy)(1 + n^{-1/2} h_n(x, y))$$

with $|h_n| \leq C n^{1/8}$ for some finite constant C and

$$n^{1/2}((1 + n^{-1/2} h_n)^{1/2} - 1) \to \tfrac{1}{2} h \quad \text{in } L_2(\pi \otimes Q).$$

Hence H^{rev} is the local parameter space in the model of all *reversible* chains. To calculate the canonical gradient of $t_f(Q) = \pi \otimes Q^{\otimes(m-1)}(f)$ in this model, we note first that the canonical gradient g_0 of $t_f(Q)$ in the full nonparametric model, given in (3.11), can be written as $g = A_2 k$ with

$$k = Q^{\otimes(m-2)} f + \sum_{i=3}^{m} \Delta \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i)} f. \tag{4.1}$$

This follows from $\Delta k \in H$ for $k \in L_2(\pi \otimes Q)$ and $A_2 h = h$ for $h \in H$, and from $A_2 Q^{\otimes(m-2)} f = A Q^{\otimes(m-1)} f + \Delta Q^{\otimes(m-2)} f$. The canonical gradient g_0^{rev} of the functional $t_f(Q)$ in the model of all reversible chains is the projection of g_0 onto H^{rev}. As in the proof of Theorem 2 of Greenwood and Wefelmeyer (1999) we obtain the projection by symmetrizing k with respect to time reversal,

$$g_0^{rev} = A_2 k^{sym}$$

with

$$k^{sym}(x, y) = \tfrac{1}{2}(k(x, y) + k(y, x)).$$

Easy calculations show that

$$A_2 k^{sym} = \tfrac{1}{2}(A Q^{\otimes(m-1)} f + A \bar{Q}^{\odot(m-1)} f) + \sum_{i=2}^{m} \Delta [Q^{\odot(i-2)} Q^{\otimes(m-i)} f]^{sym}.$$

Alternatively, we can write

$$g_0^{rev} = T f^{sym}.$$

Recall from the Introduction that

$$f^{sym}(x_1, \ldots, x_m) = \tfrac{1}{2}[f(x_1, \ldots, x_m) + (x_m, \ldots, x_1)].$$

Note that $T f^{sym}$ is the canonical gradient of the functional $t_{f^{sym}}(Q) = \pi \otimes Q^{\otimes(m-1)}(f^{sym})$ in the full nonparametric model. This functional coincides with $t_f(Q)$ on the set of reversible Markov kernels.

Remark 7. For a function $f = \phi$ of *one* argument we recover the result of Greenwood and Wefelmeyer (1999) that reversibility does not help estimating $E[\phi(X_1)]$. By (4.1) the canonical gradient of $t_\phi(Q) = E[\phi(X_1)]$ in the full nonparametric model can be written $g_0 = A_2 k$ with $k(x, y) = \phi(y)$. Hence $g_0^{rev} = A_2 \phi^{sym}$ with

$$\phi^{sym}(x, y) = \tfrac{1}{2}(\phi(y) + \phi(x)).$$

But $A_2 \phi^{sym} = A\phi$. Hence the empirical estimator $E_n \phi$ is efficient for $E[\phi(X_1)]$ in the model of all reversible Markov chains.

Remark 8. For $m = 2$ we recover the result of Greenwood and Wefelmeyer (1999) that the symmetrized empirical estimator is efficient in the model of all reversible chains. Assume first that $m \geq 2$ and consider the symmetrized empirical estimator

$$E_n f^{sym} = \frac{1}{(n - m + 2)} \sum_{j=m-1}^{n} f^{sym}(X_{j-m+1}, \ldots, X_j).$$

As in (3.12), we obtain that

$$n^{1/2}(E_n f^{sym} - E[\phi(X_1)]) = n^{-1/2} \sum_{j=m-1}^{n} A_m f^{sym}(X_{j-m+1}, X_j) + o_{P_n}(1).$$

As $t_{f^{sym}}$ coincides with t_f on the set of reversible kernels, we see that

$$\mathfrak{L}(n^{1/2}(E_n f^{sym} - t_f(Q)) \mid P_n) \implies \mathcal{N}(0, \pi \otimes Q^{\otimes(m-1)}((A_m f^{sym})^2)).$$

It is easy to verify that

$$g_0^{rev}(X_{m-1}, X_m) = E[A_m f^{sym}(X_1, \ldots, X_m) \mid X_{m-1}, X_m].$$

Thus for $m = 2$ we have $A_2 f^{sym} = g_0^{rev}$, and the symmetrized empirical estimator is efficient in this case. For $m \geq 3$, the asymptotic variance of the empirical estimator exceeds the information bound $\pi \otimes Q^{\otimes(m-1)}((g_0^{rev})^2)$ by the amount

$$E[(A_m f^{sym}(X_1, \ldots, X_m) - g_0^{rev}(X_{m-1}, X_m))^2].$$

5. Construction of efficient estimators

We have already seen that the empirical estimator for $E[f(X_1, \ldots, X_m)]$ is efficient in the full nonparametric model if $m \leq 2$, but is typically not efficient if $m > 2$. Let us now study how to construct an efficient estimator in this case. Assume throughout this section that $m > 2$.

First consider the "estimator"

$$\begin{aligned}\tau_n &= \frac{1}{n}\sum_{j=1}^{n}\left(Q^{\otimes(m-1)}f(X_{j-1}) + \sum_{i=2}^{m}\Delta\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f(X_{j-1},X_j)\right)\\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{i=2}^{m}\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f(X_{j-1},X_j) - \sum_{i=3}^{m}\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i+1)}f(X_{j-1})\right)\end{aligned}$$

which still depends on the unknown transition distribution. By our assumptions we have $Q^{\otimes(m-1)}f \in \mathcal{L}_V$. Thus it follows from the martingale representation (3.12) that

$$\begin{aligned}n^{1/2}&\left(\frac{1}{n}\sum_{j=1}^{n}Q^{\otimes(m-1)}(X_{j-1}) - \pi(Q^{\otimes(m-1)}f)\right)\\ &\qquad\qquad = n^{-1/2}\sum_{j=1}^{n}AQ^{\otimes(m-1)}f(X_{j-1},X_j) + o_{P_n}(1).\end{aligned}$$

As $\pi(Q^{\otimes(m-1)}f) = \pi\otimes Q^{\otimes(m-1)}(f) = t_f(Q)$, we obtain that τ_n has influence function g_0. This suggests a plug-in estimator that replaces the unknown Markov kernels Q and $\bar{Q}$ in the "estimator" τ_n by estimators, say Q_n and $\bar{Q}_n$.

To avoid technical difficulties we adopt the sample splitting technique of Schick (1998). For simplicity we use his two-split. For this let $\langle N_n\rangle$ be a sequence of positive integers such that

$$\frac{2N_n}{n}\to 1, \quad n-2N_n\to\infty \quad\text{and}\quad \frac{n-2N_n}{\sqrt{n}}\to 0.$$

Let $Q_{n,1}$ and $\bar{Q}_{n,1}$ be the estimators of Q and $\bar{Q}$ based on the first N_n+1 observations $X_0,\dots,X_{N_n}$ only, and let $Q_{n,2}$ and $\bar{Q}_{n,2}$ be the same estimators based on the last N_n+1 observations $X_{n-N_n},\dots,X_n$. Let now

$$\hat{\tau}_n = \tfrac{1}{2}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}f_{n,2}(X_{j-1},X_j) + \frac{1}{N_n}\sum_{j=n-N_n+1}^{n}f_{n,1}(X_{j-1},X_j)\right), \tag{5.1}$$

where, for $j=1,2$,

$$f_{n,j}(x,y) = \sum_{i=2}^{m}\bar{Q}_{n,j}^{\odot(i-2)}Q_{n,j}^{\otimes(m-i)}f(x,y) - \sum_{i=3}^{m}\bar{Q}_{n,j}^{\odot(i-2)}Q_{n,j}^{\otimes(m-i+1)}f(x), \quad x,y\in S.$$

In the next theorem we state conditions on the estimators Q_n and $\bar{Q}_n$ under which $\hat{\tau}_n$ is efficient.

Theorem 3. *In addition to $m \geq 3$, assume that*

$$\|\|\bar{Q}\|\|_V + \|\|Q_n\|\|_V + \|\|\bar{Q}_n\|\|_V = O_{P_n}(1), \tag{5.2}$$

$$\sum_{i=2}^{m} \pi \otimes Q(|\bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i)} f - \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i)} f|) = o_{P_n}(1), \tag{5.3}$$

$$\sum_{i=3}^{m} \pi(|\bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i+1)} f - \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i+1)} f|) = o_{P_n}(1), \tag{5.4}$$

$$\sum_{i=3}^{m} (\bar{Q}_n^{\odot(i-2)} - \bar{Q}^{\odot(i-2)}) \odot \pi \otimes (Q_n - Q) \otimes Q_n^{\otimes(m-i)}(f) = o_{P_n}(n^{-1/2}). \tag{5.5}$$

Then the estimator sequence $\langle \hat{\tau}_n \rangle$ defined in (5.1) is asymptotically equivalent to the "estimator" sequence $\langle \tau_n \rangle$,

$$n^{1/2}(\hat{\tau}_n - \tau_n) = o_{P_n}(1). \tag{5.6}$$

Thus $\langle \hat{\tau}_n \rangle$ has influence function g_0 and hence is regular and efficient for $E[f(X_1, \ldots, X_m)]$ in the full model.

The proof is in Section 8.

Remark 9. Suppose the state space S is finite. Take Q_n and $\bar{Q}_n$ to be the empirical estimators defined by

$$Q_n(x, \{y\}) = \frac{\sum_{j=1}^{n} \mathbf{1}[X_{j-1} = x, X_j = y]}{\sum_{j=1}^{n} \mathbf{1}[X_{j-1} = x]}$$

and

$$\bar{Q}_n(x, \{y\}) = \frac{\sum_{j=1}^{n} \mathbf{1}[X_j = x, X_{j-1} = y]}{\sum_{j=1}^{n} \mathbf{1}[X_j = x]}$$

for $x, y \in S$. Then

$$n^{1/2}|Q_n(x, \{y\}) - Q(x, \{y\}| + n^{1/2}|\bar{Q}_n(x, \{y\}) - \bar{Q}(x, \{y\})| = O_{P_n}(1)$$

for all x and y in S. From this it is easy to see that (5.2)–(5.5) hold. In this simple case, one does not have to split the sample. One can show directly that

$$\frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=2}^{m} \bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i)} f(X_{j-1}, X_j) - \sum_{i=3}^{m} \bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i+1)} f(X_{j-1}) \right)$$

has influence function g_0. Of course, for discrete state space we do not need this construction. A simple efficient estimator for $E[f(X_1,\dots,X_m)]$ in this case is described in the Introduction.

Remark 10. Let

$$\nu_{n,a} = a(Q_n - Q) + (1-a)Q \quad \text{and} \quad \bar\nu_{n,a} = a(\bar Q_n - \bar Q) + (1-a)\bar Q, \quad a = 0,1.$$

Then we have, for $j = 1,\dots,m$,

$$Q_n^{\otimes j} = (Q_n - Q + Q)^{\otimes j} = (\nu_{n,0} + \nu_{n,1})^{\otimes j} = \sum_{a_1,\dots,a_j=0,1} \nu_{n,a_1} \otimes \cdots \otimes \nu_{n,a_j}$$

and

$$\bar Q_n^{\odot j} - \bar Q^{\odot j} = \sum_{a_1,\dots,a_j=0,1:a_1+\cdots+a_j\ge 1} \bar\nu_{n,a_1} \odot \cdots \odot \bar\nu_{n,a_j}.$$

Using these relations we can conclude that the left-hand side of (5.5) equals

$$(5.7) \qquad \sum_{i=1}^{m-2} \sum_{A_i} \bar\nu_{n,a_1} \odot \cdots \odot \bar\nu_{n,a_i} \odot \pi \otimes \nu_{n,a_{i+1}} \otimes \cdots \otimes \nu_{n,a_{m-1}}(f) = o_{P_n}(n^{-1/2}),$$

where $A_i = \{(a_1,\dots,a_{m-1}) \in \{0,1\}^{m-1} : a_1 + \cdots + a_i \ge 1, a_{i+1} = 1\}$.

Remark 11. Covariances and correlations of time series involve expectations of functions of the form

$$f(x_1,\dots,x_m) = u(x_1)\,w(x_m), \quad x_1,\dots,x_m \in S.$$

For such f,

$$\bar Q_n^{\odot(i-2)} Q_n^{\otimes(m-i)} f(x,y) = \bar Q_n^{i-2} u(x)\; Q_n^{m-i} w(y), \quad x,y \in S,$$

and

$$\bar Q^{\odot(i-2)} Q^{\otimes(m-i)} f(x,y) = \bar Q^{i-2} u(x)\; Q^{m-i} w(y), \quad x,y \in S.$$

Thus the left-hand sides of (5.3) and (5.4) simplify to

$$\sum_{i=2}^{m} \iint \pi(dx) Q(x,dy) |\bar Q_n^{i-2} u(x) Q_n^{m-i} w(y) - \bar Q^{i-2} u(x) Q^{m-i} w(y)|$$

and

$$\sum_{i=3}^{m} \int \pi(dx) |\bar Q_n^{i-2} u(x) Q_n^{m-i+1} w(x) - \bar Q^{i-2} u(x) Q^{m-i+1} w(x)|.$$

Hence (5.3) and (5.4) are implied by

$$\pi(|\bar{Q}_n^i u - \bar{Q}^i u|^2 + |Q_n^i w - Q^i w|^2) = o_{P_n}(1), \quad i = 1, \dots, m-2.$$

Furthermore, one can check that the left-hand side of (5.5) simplifies to

$$\sum_{i=1}^{m-2} \pi((\bar{Q}_n^i u - \bar{Q}^i u)(Q_n - Q)Q_n^{m-2-i} w).$$

Thus a sufficient condition for (5.6) is that

$$n \sum_{i=1}^{m-2} \pi(|\bar{Q}_n^i u - \bar{Q}^i u|^2)\, \pi(|(Q_n - Q)Q_n^{m-2-i} w|^2) = o_{P_n}(1).$$

Remark 12. Suppose $m = 3$. Then the left-hand side of (5.3) becomes

$$\iint \pi(dx)Q(x,dy) \left| \int (Q_n(y,dz) - Q(y,dz)) f(x,y,z) \right| \\ + \iint \pi(dy)Q(y,dz) \left| \int (\bar{Q}_n(y,dx) - \bar{Q}(y,dx)) f(x,y,z) \right|,$$

while the left-hand side of (5.4) becomes

$$\int \pi(dy) \left| \iint (\bar{Q}_n(y,dx) - \bar{Q}(y,dx))(Q_n(y,dz) - Q(y,dz)) f(x,y,z) \right|.$$

The latter is also a bound for the left-hand side of (5.5) which equals the above without the absolute values. These expressions simplify further if $f(x,y,z) = u(x,y)w(y,z)$. In this case they reduce to

$$\pi(\bar{u}_* |W_n| + w_* |U_n|) \quad \text{and} \quad \pi(|U_n W_n|)$$

with $\bar{u}_*(y) = \int \bar{Q}(y,dx)|u(x,y)|$ and $w_*(y) = \int Q(y,dz)|w(y,z)|$, and with

$$U_n(y) = \int \bar{Q}_n(y,dx)u(x,y) - \int \bar{Q}(y,dx)u(x,y),$$

$$W_n(y) = \int Q_n(y,dz)w(y,z) - \int Q(y,dz)w(y,z).$$

Sufficient conditions for (5.3)–(5.5) are then

$$\pi(U_n^2) + \pi(W_n^2) = o_{P_n}(1) \tag{5.8}$$

and

$$n\pi(U_n^2)\pi(W_n^2) = o_{P_n}(1). \tag{5.9}$$

These conditions in turn follow if for some $\alpha \in (0,1)$,

$$n^\alpha \pi(U_n^2) + n^{1-\alpha}\pi(W_n^2) = o_{P_n}(1). \tag{5.10}$$

If $f(x_1, x_2, x_3) = u_1(x_1)u_2(x_2)u_3(x_3)$, then we can take

$$u(x,y) = u_1(x)|u_2(y)|^{1/2} \quad \text{and} \quad w(y,z) = \operatorname{sign}(u_2(y))|u_2(y)|^{1/2}u_3(z),$$

and (5.10) becomes

$$n^\alpha \pi((\bar{Q}_n u_1 - \bar{Q}u_1)^2|u_2|) + n^{1-\alpha}\pi((Q_n u_3 - Qu_3)^2|u_2|) = o_{P_n}(1). \tag{5.11}$$

6. A special case

Assume that the state space is a compact interval $S = [a,b]$ endowed with its Borel σ-field $\mathcal{S}$, that both Q and $\bar{Q}$ are V-uniformly ergodic for $V = 1$, and that $\pi \otimes Q$ has a density g which is bounded away from zero on $[a,b] \times [a,b]$ and satisfies the Lipschitz condition

$$|g(t,y) - g(s,y)| + |g(x,t) - g(x,s)| \le L|t-s|, \quad s,t,x,y \in [a,b], \tag{6.1}$$

for some finite constant L. These assumptions on g imply that the stationary distribution π has a density which is Lipschitz continuous, bounded and bounded away from zero on S. We take f to be of the form

$$f(x_1, \dots, x_m) = u_1(x_1) \cdot \ldots \cdot u_m(x_m), \quad x_1, \dots, x_m \in S,$$

for bounded measurable functions $u_1, \dots, u_m$ on S. Thus estimating $E[f(X_1, \dots, X_m)]$ includes as special cases estimating $E[X_1 X_m]$ and $E[I[X_1 \le s_1] \cdot \ldots \cdot I[X_m \le s_m]]$ for $s_1, \dots, s_m$ in S.

We shall use kernel estimators of Q and $\bar{Q}$. For this let k be a Lipschitz continuous symmetric density with support $[-1,1]$, and a_n a bandwidth such that $cn^{-1/3} \le a_n \le Cn^{-1/3}$ for constants $0 < c < C < \infty$. Set $k_n(x) = k(x/a_n)/a_n$. We estimate Q by the random kernel Q_n defined by

$$Q_n(x,A) = \frac{\sum_{j=1}^n k_n(x - X_{j-1})I[X_j \in A]}{\sum_{j=1}^n k_n(x - X_{j-1})}, \quad x \in [a,b], A \in \mathcal{S}.$$

This estimator has the property that

$$\sup_{x\in S}|Q_n u(x) - Qu(x)| = o_{P_n}(n^{-1/3}\log n) \tag{6.2}$$

for every bounded measurable function u on $[a,b]$. Indeed, it follows from the V-uniform ergodicity with $V=1$ that the chain is ϕ-mixing with geometrically decaying coefficients. Thus we obtain the desired result from Remark 3.3.5 or Theorem 3.4.9 in Györfi et al. (1989). (There is a misprint in their condition (H.14): Convergence should be to 0.) We estimate $\bar{Q}$ by the random kernel $\bar{Q}_n$ defined by

$$\bar{Q}_n(x,A) = \frac{\sum_{j=1}^n k_n(x-X_j)I[X_{j-1}\in A]}{\sum_{j=1}^n k_n(x-X_j)}, \quad x\in[a,b], A\in\mathcal{S}.$$

This estimator also satisfies

$$\sup_{x\in S}|\bar{Q}_n u(x) - \bar{Q}u(x)| = o_{P_n}(n^{-1/3}\log n) \tag{6.3}$$

for every bounded measurable function u on $[a,b]$. It is obvious that (5.2) holds. With the aid of (6.2) and (6.3) it is now easy to check that (5.3)–(5.5) hold for our choice of f. For (5.5) use also the representation (5.7).

Thus the construction (5.1) for our present f yields an estimator which has influence function Tf and is hence efficient for estimating $E[u_1(X_1)\cdot\ldots\cdot u_m(X_m)]$. Denote this estimator by $\tau_n(f,Q_n,\bar{Q}_n)$ to stress its dependence on f and the choice of estimators Q_n and $\bar{Q}_n$ of Q and $\bar{Q}_n$. If the chain is known to be reversible, we can replace Q_n and $\bar{Q}_n$ by $Q_n^{sym} = \frac{1}{2}(Q_n+\bar{Q}_n)$ and use the estimator $\tau_n^{sym} = \tau_n(f^{sym}, Q_n^{sym}, Q_n^{sym})$. This estimator has influence function Tf^{sym} and hence is an efficient estimator of $t_f(Q)$ in the model of all reversible Markov chains.

7. Proof of Theorem 2

1. We show first that $t_f(Q) = \pi\otimes Q^{\otimes(m-1)}(f)$ has gradient

$$g = AQ^{\otimes(m-1)}f + \sum_{i=2}^m \bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f.$$

In part 2 of the proof, the canonical gradient g_0 is obtained by projecting g onto H.

Fix $h\in H$. We need to show that for each f satisfying Assumption 2 we have

$$n^{1/2}\Big(\pi_{n,h}\otimes Q_{n,h}^{\otimes(m-1)}(f) - \pi\otimes Q^{\otimes(m-1)}(f)\Big) \to \pi\otimes Q(gh).$$

Here we interpret $\pi_{n,h} \otimes Q_{n,h}^{\otimes 0}$ as $\pi_{n,h}$ and $\pi \otimes Q^{\otimes 0}$ as π. We shall prove this by induction on m. By Lemma 1, the result is true for the case $m = 1$. Now assume the result holds for $m - 1$. The induction hypothesis, with f replaced by Qf, yields

$$n^{1/2}\Big(\pi_{n,h} \otimes Q_{n,h}^{\otimes(m-2)}(Qf) - \pi \otimes Q^{\otimes(m-2)}(Qf)\Big) \to \pi \otimes Q((g - h\bar{Q}^{\odot(m-2)}f))$$

for all $h \in H$. Since we can write

$$\pi_{n,h} \otimes Q_{n,h}^{\otimes(m-1)}(f) = \pi_{n,h} \otimes Q_{n,h}^{\otimes(m-2)} \otimes (Q_{n,h} - Q)(f) + \pi_{n,h} \otimes Q_{n,h}^{\otimes(m-2)}(Qf),$$

we need to show that

$$n^{1/2}\Big(\pi_{n,h} \otimes Q_{n,h}^{\otimes(m-2)} \otimes (Q_{n,h} - Q)(f)\Big) \to \pi \otimes Q(h\bar{Q}^{\odot(m-2)}f). \tag{7.1}$$

With

$$k_n(x_1, \dots, x_{m-1}) = \int Q(x_{m-1}, dy)h_n(x_{m-1}, y)f(x_1, \dots, x_{m-1}, y), \quad x_1, \dots, x_{m-1} \in S,$$

we can express the left-hand side of (7.1) as

$$\pi \otimes Q(h_n\bar{Q}^{\odot(m-2)}f) + \pi_{n,h} \otimes Q_{n,h}^{\otimes(m-2)}(k_n) - \pi \otimes Q^{\otimes(m-2)}(k_n).$$

We have used the fact that $\pi \otimes Q(h_n\bar{Q}^{\odot(m-2)}f) = \pi \otimes Q^{\otimes(m-2)}(k_n)$. Since $\bar{Q}^{\odot(m-2)}f \in L_2(\pi \otimes Q)$ and $h_n \to h$ in $L_2(\pi \otimes Q)$, we obtain that

$$\pi \otimes Q(h_n\bar{Q}^{\odot(m-2)}f) \to \pi \otimes Q(h\bar{Q}^{\odot(m-2)}f).$$

It follows from (3.9), Assumption 2 and $|h_n| \le 2n^{1/8}$ that

$$\pi_{n,h} \otimes Q_{n,h}^{\otimes(m-2)}(k_n) - \pi \otimes Q^{\otimes(m-2)}(k_n) \to 0.$$

Combining these results, we obtain the desired (7.1).

2. It remains to project g onto H. The function $AQ^{\otimes(m-1)}f$ is in H. The projections of $\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f$ onto H are obtained by subtracting the conditional expectations. The projections can be written with the operator Δ introduced in (3.7), i.e., as

$$\Delta\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f(x,y) = \bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f(x,y) - \bar{Q}^{\odot(i-2)}Q^{\otimes(m-i+1)}f(x).$$

Hence the canonical gradient of $t_f(Q) = \pi \otimes Q^{\otimes(m-1)}(f)$ in the full nonparametric model is

$$g_0 = Tf = AQ^{\otimes(m-1)}f + \sum_{i=2}^{m} \Delta\bar{Q}^{\odot(i-2)}Q^{\otimes(m-i)}f.$$

8. Proof of Theorem 3

We have already seen that τ_n has influence function g_0. Thus it is enough to verify (5.6). We shall use the results of Schick (1998) to do so. Set

$$h(x,y) = \sum_{i=2}^{m} \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i)} f(x,y) - \sum_{i=3}^{m} \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i+1)} f(x)$$

and

$$h_n(x,y) = \sum_{i=2}^{m} \bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i)} f(x,y) - \sum_{i=3}^{m} \bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i+1)} f(x).$$

Note that $\tau_n = n^{-1} \sum_{j=1}^{n} h(X_{j-1}, X_j)$. By the properties of Q, $\bar{Q}$, Q_n and $\bar{Q}_n$,

$$\sup_{x,y\in S} \frac{|\bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i)} f(x,y) - \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i)} f(x,y)|}{V^{\alpha_1+\cdots+\alpha_{i-1}}(x) V^{\alpha_i+\cdots+\alpha_m}(y)} = O_{P_n}(1), \quad i = 2, \dots, m,$$

and

$$\sup_{x\in S} \frac{|\bar{Q}_n^{\odot(i-2)} Q_n^{\otimes(m-i+1)} f(x) - \bar{Q}^{\odot(i-2)} Q^{\otimes(m-i+1)} f(x)|}{V(x)} = O_{P_n}(1), \quad i = 3, \dots, m.$$

Thus, by Theorem 3.7 in Schick (1998), the desired (5.6) follows from

(8.1) $$\pi \otimes Q(h_n - h) = o_{P_n}(n^{-1/2})$$

and

(8.2) $$\pi \otimes Q(|h_n - h|) = o_{P_n}(1).$$

The second condition is a consequence of (5.3) and (5.4). To obtain (8.1), verify that $\pi \otimes Q(h_n - h) = \Delta_n(f)$, where

$$\Delta_n = \pi \otimes Q \otimes Q_n^{\otimes(m-2)} - \pi \otimes Q^{\otimes(m-1)} - \sum_{i=3}^{m} \bar{Q}_n^{\odot(i-2)} \odot \pi \otimes (Q_n - Q) \otimes Q^{\otimes(m-i)}$$

Now use $Q_n^{\otimes(m-2)} - Q^{\otimes(m-2)} = \sum_{i=3}^{m} Q^{\otimes(i-3)} \otimes (Q_n - Q) \otimes Q_n^{\otimes(m-i)}$ and $\bar{Q}^{\odot a} \odot \pi = \pi \otimes \bar{Q}^{\otimes a}$ for $a = 1, \dots, m$, to obtain

$$\Delta_n = \sum_{i=3}^{m} (\bar{Q}^{\odot(i-2)} - \bar{Q}_n^{\odot(i-2)}) \odot \pi \otimes (Q_n - Q) \otimes Q_n^{\otimes(m-i)}.$$

It is now easy to see that (5.5) is equivalent to (8.1).

9. References

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

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

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

Greenwood, P.E. and Wefelmeyer, W. (1997). Maximum likelihood estimator and Kullback-Leibler information in misspecified Markov chain models. *Teor. Veroyatnost. i Primenen*, **42**, 169-178.

Greenwood, P.E. and Wefelmeyer, W. (1999). Reversible Markov chains and optimality of symmetrized empirical estimators. *Bernoulli*, **9**, 109-123.

Greenwood, P.E., McKeague, I.W. and Wefelmeyer, W. (1999). Splitting Markov fields and combining empirical estimators. In preparation.

Györfi, L., Härdle W., Sarda, P. and Vieu, P. (1989). Nonparametric Curve Estimation from Time Series. *Lecture Notes in Statistics*, **60**, Springer-Verlag, Berlin.

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

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

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.

Kartashov, N. V. (1996). *Strong Stable Markov Chains.* VSP, Utrecht.

Koshevnik, Y. A. and Levit, B. Y. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl.*, **21**, 738-753.

Levit, B. Y. (1974). On optimality of some statistical estimates. In: *Proceedings of the Prague Symposium on Asymptotic Statistics*, (J. Hájek, eds.), **2**, 215-238, Charles University, Prague.

Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

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

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, **2**, 13-25.

Roussas, G. G. (1965). Asymptotic inference in Markov processes. *Ann. Math. Statist.*, **36**, 987–992.

Roussas, G. G. (1972). *Contiguous Probability Measures: Some Applications in Statistics.* Cambridge University Press, London.

Schick, A. (1998). Sample splitting with Markov chains. Preprint. http://math.binghamton.edu/anton/preprint.html

Anton Schick
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000
USA
anton@math.binghamton.edu
http://math.binghamton.edu/anton/index.html

Wolfgang Wefelmeyer
Universität-GH Siegen
Fachbereich 6 Mathematik
Walter-Flex-Str. 3
57068 Siegen
Germany
wefelmeyer@mathematik.uni-siegen.de
http://www.math.uni-siegen.de/statistik/wefelmeyer.html