Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Statistics & Probability Letters 66 (2004) 335–345
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Efficient estimators for functionals of Markov chains with parametric marginals

Spiridon Penev[a,1], Hanxiang Peng[b], Anton Schick[c,*,2], Wolfgang Wefelmeyer[d]

[a] Department of Statistics, University of New South Wales, Sydney NSW 2052, Australia
[b] Department of Mathematics, University of Mississippi, University, MS 38677, USA
[c] Department of Mathematical Science, Binghamton University, Box 6000, Binghamton, NY 13902-6000, USA
[d] Mathematisches Institut, Universität zu Köln, Weyertal 86-90, 509 31 Köln, Germany

Received April 2003

Abstract

Suppose we observe a geometrically ergodic Markov chain with a parametric model for the marginal, but no (further) information about the transition distribution. Then the empirical estimator for a linear functional of the joint law of two successive observations is no longer efficient. We construct an improved estimator and show that it is efficient. The construction is similar to a recent one for bivariate models with parametric marginals. The result applies to discretely observed parametric continuous-time processes.

MSC: primary 62G20; 62G30; 62M05

Keywords: Least-squares estimator; Series estimator; Orthonormal basis; Efficient influence function; Reversible Markov chain; Discretely observed diffusion; Constrained model

1. Introduction

Let $X_0,\dots,X_n$ be observations from a geometrically ergodic Markov chain with arbitrary state space. We want to estimate a linear functional $E[h(X_0,X_1)]$ of the joint stationary law of two successive observations. If nothing is known about the distribution of the chain, then the empirical estimator $\hat{H}=1/n\sum_{k=1}^{n}h(X_{k-1},X_k)$ is efficient; see Penev (1990, 1991), Bickel (1993), and Greenwood and Wefelmeyer (1995). Suppose now that we have a finite-dimensional parametric model F_ϑ, $\vartheta\in\Theta$,

* Corresponding author. Tel.: +1-607-777-2983; fax: +1-607-777-2450.
E-mail address: anton@math.binghamton.edu (A. Schick).
[1] Supported by a university research support grant, UNSW.
[2] Supported by NSF Grant DMS 0072174.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2003.10.022

for the marginal stationary law of the chain, but that we cannot or do not want to specify anything (else) about the transition distribution. Then we can construct better estimators for $E[h(X_0,X_1)]$. This includes the case where the transition distribution follows a parametric model involving the parameter ϑ and perhaps further parameters, but that we do not know this model or are not sure that it is correct. Our model is nonparametric, with a constraint that involves the unknown parameter ϑ.

Our results apply in particular to parametric continuous-time Markov processes that are discretely observed at fixed time intervals. Under such an observation scheme, estimators for the parameter ϑ were constructed in parametric diffusion processes by Pedersen (1995a,b), Bibby and Sørensen (1995, 1996, 1997, 2001), Kessler and Sørensen (1999) and Kessler (2000), and in general parametric continuous-time processes by Kessler and Sørensen (2002), see also Sørensen (1997). These estimators could be used to estimate the coefficients of the diffusion and then linear functionals $E[h(X_0,X_1)]$ as considered here. If the diffusion model is correctly specified, and if the estimators for ϑ are efficient (or nearly so), this would lead to better estimators for $E[h(X_0,X_1)]$ than ours. However, the marginals of a discretely observed process can be modeled much better than the dynamics. Estimators of $E[h(X_0,X_1)]$ based on a misspecified continuous-time model will usually be inconsistent. In contrast, our estimator uses only the information in the parametric model for the marginal law and is always $n^{1/2}$-consistent and asymptotically normal unless the marginals are misspecified.

Our results are closely related to results for bivariate models, which we recall first. Let $(Y_1,Z_1),\dots,(Y_n,Z_n)$ be i.i.d. bivariate random variables with joint law Q. We want to estimate a linear functional $E[h(Y,Z)] = \int h\,dQ$ for a fixed function $h \in L_2(Q)$. A natural estimator is the empirical estimator $\hat{H}_{\text{biv}} = 1/n\sum_{k=1}^n h(Y_k,Z_k)$. If additional structural assumptions on the joint law hold, this estimator can be improved.

Assume first that the marginals F and G of Q are *known*. In this case there is a large class of unbiased estimators. Indeed,

$$\hat{H}_{\text{biv}}(a,b) = \frac{1}{n}\sum_{k=1}^{n}(h(Y_k,Z_k) - a(Y_k) - b(Z_k))$$

is unbiased for each $a \in L_{2,0}(F)$ and $b \in L_{2,0}(G)$. Here, for any measure μ,

$$L_{2,0}(\mu) = \left\{ h \in L_2(\mu) : \int h\,d\mu = 0 \right\}.$$

The smallest variance is achieved by $\hat{H}_{\text{biv}}(a_Q,b_Q)$, where a_Q and b_Q minimize $E[(h(Y,Z) - a(Y) - b(Z))^2]$ over $a \in L_{2,0}(F)$ and $b \in L_{2,0}(G)$. Bickel, Ritov and Wellner (1991) have shown that any estimator equivalent to $\hat{H}_{\text{biv}}(a_Q,b_Q)$ is efficient, and have obtained such an estimator using the modified minimum chi-square estimator of Deming and Stephan (1940). Peng and Schick (2002) give a more direct construction, estimating a_Q and b_Q by a series estimator in terms of orthonormal bases $v_1,v_2,\dots$ of $L_{2,0}(F)$ and $w_1,w_2,\dots$ of $L_{2,0}(G)$. Their estimator is of the form

$$\frac{1}{n}\sum_{k=1}^{n}\left(h(Y_k,Z_k) - \sum_{i=1}^{M_n}\hat{\alpha}_{ni}v_i(Y_k) - \sum_{j=1}^{N_n}\hat{\beta}_{nj}w_j(Z_k) \right),$$

where M_n and N_n are integers that tend slowly to infinity with the sample size n, and $\hat{\alpha}_{n1},\dots,\hat{\alpha}_{nM_n}$, $\hat{\beta}_{n1},\dots,\hat{\beta}_{nN_n}$ are chosen to minimize

$$\sum_{k=1}^{n}\left(h(Y_k,Z_k)-\sum_{i=1}^{M_n}\alpha_i v_i(Y_k)-\sum_{j=1}^{N_n}\beta_j w_j(Z_k)\right)^2.$$

Of course, $\hat{\alpha}_{n1},\dots,\hat{\alpha}_{nM_n},\hat{\beta}_{n1},\dots,\hat{\beta}_{nN_n}$ are simply least-squares estimators for the response vector $H=(h(Y_1,Z_1),\dots,h(Y_n,Z_n))^\top$ and the design matrix with kth row formed by

$$(v_1(Y_k),\dots,v_{M_n}(Y_k),w_1(Z_k),\dots,w_{N_n}(Z_k)).$$

The assumption of known marginals is not always justifiable. A more realistic assumption is that the marginals depend on some unknown parameter ϑ, i.e., $F=F_\vartheta$ and $G=G_\vartheta$. This model is considered by Peng and Schick (2003). They replace, in the above construction, v_i by $v_i(\cdot,\hat{\vartheta})$ and w_i by $w_i(\cdot,\hat{\vartheta})$, where $v_1(\cdot,\vartheta),v_2(\cdot,\vartheta),\dots$ is a basis for $L_{2,0}(F_\vartheta)$; $w_1(\cdot,\vartheta),w_2(\cdot,\vartheta),\dots$ is a basis for $L_{2,0}(G_\vartheta)$; and $\hat{\vartheta}$ is a $n^{1/2}$-consistent estimator of ϑ. They show under mild assumptions on the bases that the resulting estimator $\hat{H}^*_{\mathrm{biv}}$ has an expansion

$$\hat{H}^*_{\mathrm{biv}}=\frac{1}{n}\sum_{k=1}^{n}(h(Y_k,Z_k)-a_Q(Y_k)-b_Q(Z_k))+D_{\mathrm{biv}}^\top(\hat{\vartheta}-\vartheta)+\mathrm{o_p}(n^{-1/2}) \tag{1.1}$$

if the parametric models for the marginals are Hellinger differentiable at ϑ with derivatives ϕ_ϑ and γ_ϑ, say. Here

$$D_{\mathrm{biv}}=E[a_Q(Y)\phi_\vartheta(Y)]+E[b_Q(Z)\gamma_\vartheta(Z)].$$

Bickel and Kwon (2001) have suggested that results on efficient estimation for bivariate models carry over to geometrically ergodic Markov chains. They point out that the calculation of efficient influence functions is identical if one parametrizes the Markov chain by the joint law of two successive observations, which corresponds to the description of the bivariate model by the joint law of (Y,Z). See also the discussion of Greenwood et al. (2001). Bickel and Kwon also suggest that the construction of efficient estimators for bivariate models should carry over to corresponding Markov chain models. In this paper we carry out this program for Markov chains with a parametric model F_ϑ, $\vartheta\in\Theta$, for the marginal stationary law. For the corresponding bivariate model we have $G_\vartheta=F_\vartheta$. Recall that the observations for the Markov chain are $X_0,\dots,X_n$. The Markov chain analog $\hat{H}^*$ of $\hat{H}^*_{\mathrm{biv}}$ is obtained by replacing the pairs (Y_k,Z_k) by pairs (X_{k-1},X_k) of successive observations. We show in Section 2 that the analog of (1.1) is

$$\hat{H}^*=\frac{1}{n}\sum_{k=1}^{n}(h(X_{k-1},X_k)-a_Q(X_{k-1})-b_Q(X_k))+D^\top(\hat{\vartheta}-\vartheta)+\mathrm{o_p}(n^{-1/2}) \tag{1.2}$$

under the assumption that the parametric model for the marginal stationary law is Hellinger differentiable at ϑ with derivative ϕ_ϑ. Now a_Q and b_Q are minimizers of $E[(h(X_0,X_1)-a(X_0)-b(X_1))^2]$ over a and b in $L_{2,0}(F_\vartheta)$, and

$$D=E[(a_Q(X_0)+b_Q(X_0))\phi_\vartheta(X_0)].$$

Kessler et al. (2001) have constructed an efficient estimator $\hat{\vartheta}$ of ϑ. If such an estimator is used, $\hat{H}^*$ is also efficient, as shown in Section 3.

We note that the results of this paper can be adapted to the case of a reversible chain. If the chain is known to be reversible, then Q is symmetric, $Q(\mathrm{d}x,\mathrm{d}y)=Q(\mathrm{d}y,\mathrm{d}x)$, and we can improve the empirical estimator $\hat{H}=1/n\sum_{k=1}^{n}h(X_{k-1},X_k)$ by symmetrization,

$$\hat{H}_{\mathrm{sym}}=\frac{1}{2n}\sum_{k=1}^{n}(h(X_{k-1},X_k)+h(X_k,X_{k-1})).$$

If Q is completely unknown, $\hat{H}_{\mathrm{sym}}$ is efficient; see Greenwood and Wefelmeyer (1999) and, for a simpler argument, Greenwood et al. (2001). If we have a parametric model F_ϑ for the marginal, it is natural to consider the symmetric improvement

$$\hat{H}^*_{\mathrm{sym}}=\hat{H}_{\mathrm{sym}}-\frac{1}{2n}\sum_{k=1}^{n}\sum_{i=1}^{M_n}\hat{\alpha}_{ni}(v_i(X_{k-1},\hat{\vartheta})+v_i(X_k,\hat{\vartheta})),$$

where $\hat{\alpha}_{n1},\ldots,\hat{\alpha}_{nM_n}$ are chosen to minimize

$$\sum_{k=1}^{n}\left(h(X_{k-1},X_k)+h(X_k,X_{k-1})-\sum_{i=1}^{M_n}\alpha_i(v_i(X_{k-1},\hat{\vartheta})+v_i(X_k,\hat{\vartheta}))\right)^2.$$

If $\hat{\vartheta}$ is efficient, so is $\hat{H}^*_{\mathrm{sym}}$. Efficient estimators for ϑ in reversible Markov chain models with parametric marginals are constructed in Kessler et al. (2001). We note that the diffusion models referred to above are reversible.

2. Stochastic expansion of the estimator

Let $X_0,\ldots,X_n$ be observations from a stationary Markov chain on an arbitrary state space S with countably generated σ-field, transition distribution $K(x,\mathrm{d}y)$, and marginal law $F_\vartheta(\mathrm{d}x)$, with ϑ in an open subset of $\mathbf{R}^r$. Let $Q(\mathrm{d}x,\mathrm{d}y)$ denote the law of two successive observations. We want to estimate an expectation $E[h(X_0,X_1)]=\int h\,\mathrm{d}Q$ for a fixed Q-square-integrable function h.

Let $v_1(\cdot,\vartheta),v_2(\cdot,\vartheta),\ldots$ be an orthonormal basis for $L_{2,0}(F_\vartheta)$, and let $\hat{\vartheta}$ be a $n^{1/2}$-consistent estimator of ϑ. Our estimator for $\int h\,\mathrm{d}Q$ is

$$\hat{H}^*=\frac{1}{n}\sum_{k=1}^{n}\left(h(X_{k-1},X_k)-\sum_{i=1}^{M_n}\hat{\alpha}_{ni}v_i(X_{k-1},\hat{\vartheta})-\sum_{j=1}^{N_n}\hat{\beta}_{nj}v_j(X_k,\hat{\vartheta})\right),$$

where M_n and N_n are integers, and $\hat{\alpha}_{n1},\ldots,\hat{\alpha}_{nM_n},\hat{\beta}_{n1},\ldots,\hat{\beta}_{nN_n}$ are chosen to minimize

$$\sum_{k=1}^{n}\left(h(X_{k-1},X_k)-\sum_{i=1}^{M_n}\alpha_i v_i(X_{k-1},\hat{\vartheta})-\sum_{j=1}^{N_n}\beta_j v_j(X_k,\hat{\vartheta})\right)^2.$$

We prove a stochastic expansion for $\hat{H}^*$ for a fixed parameter ϑ_0 under the following assumptions on the Markov chain, the parametric model, and the basis.

Assumption 1. The chain is geometrically ergodic in the L_2 sense: There is a $\lambda < 1$ such that for all $f \in L_{2,0}(F_{\vartheta_0})$,

$$\int \left(\int K(x, \mathrm{d}y) f(y) \right)^2 F_{\vartheta_0}(\mathrm{d}x) \leqslant \lambda \int f^2 \, \mathrm{d}F_{\vartheta_0}.$$

Assumption 2. The chain fulfills the following minorization criterion: There is an $\eta > 0$ such that for F_{ϑ_0}-a.a. x and all measurable B,

$$K(x, B) \geqslant \eta F_{\vartheta_0}(B).$$

Assumption 3. The parametric model is Hellinger differentiable at ϑ_0: There is a function $\varphi \in L_{2,0}(F_{\vartheta_0})^r$ such that

$$\int \left(\sqrt{\mathrm{d}F_{\vartheta_0+t}} - \sqrt{\mathrm{d}F_{\vartheta_0}} - \frac{1}{2} t^\top \varphi \sqrt{\mathrm{d}F_{\vartheta_0}} \right)^2 = \mathrm{o}(\|t\|^2).$$

Moreover, the Fisher information matrix $\int \varphi \varphi^\top \mathrm{d}F_{\vartheta_0}$ is positive definite.

Assumption 4. The basis elements are bounded: For each $i = 1, 2, \ldots$ and each $\vartheta \in \Theta$,

$$\sup_{x \in S} |v_i(x, \vartheta)| < \infty.$$

Assumption 5. The basis elements are locally Lipschitz: There are a neighborhood of ϑ_0 and constants $L_1, L_2, \ldots$ such that, for all s and t in the neighborhood,

$$\sup_{x \in S} |v_i(x, t) - v_i(x, s)| \leqslant L_i \|t - s\|.$$

Assumptions 1 and 3 were used in Kessler et al. (2001). Assumption 2 is equivalent to

$$Q(A \times B) \geqslant \eta F_{\vartheta_0}(A) F_{\vartheta_0}(B)$$

for all measurable A and B. This version was used for corresponding bivariate models in Bickel et al. (1991) and Peng and Schick (2002, 2003). Assumption 2 is used by Glynn and Ormoneit (2002) to prove a Hoeffding inequality for Markov chains that will be applied in the proof of our result. Assumptions 4 and 5 are as in Peng and Schick (2003).

To state our result, set $m_n = M_n \vee N_n$, and let

$$\Delta_n = \sum_{i=1}^{m_n} \sup_{x \in S} |v_i(x, \vartheta_0)|^2 \quad \text{and} \quad \Gamma_n = \sum_{i=1}^{m_n} L_i^2.$$

Theorem 1. *Let Assumptions* 1–5 *hold, and let* $\hat{\vartheta}$ *be a* $n^{1/2}$*-consistent estimator for* ϑ_0*. Assume that* M_n *and* N_n *tend to infinity, and*

$$\frac{m_n^2(\Delta_n + \Gamma_n)}{n} \to 0 \quad \textit{and} \quad \frac{\Gamma_n \log(1 + \Gamma_n)}{n} \to 0. \tag{2.1}$$

Then $\hat{H}^*$ has the stochastic expansion

$$\hat{H}^* = \frac{1}{n}\sum_{k=1}^{n}(h(X_{k-1},X_k) - a_Q(X_{k-1}) - b_Q(X_k)) + D^\top(\hat{\vartheta} - \vartheta_0) + o_p(n^{-1/2}),$$

where a_Q and b_Q minimize

$$\int (h(x,y) - a(x) - b(y))^2 Q(\mathrm{d}x,\mathrm{d}y)$$

over $a, b \in L_2(F_{\vartheta_0})$, and

$$D = \int (a_Q + b_Q)\varphi \,\mathrm{d}F_{\vartheta_0}.$$

A specific basis with these properties in the case of real state space and continuous distribution functions F_ϑ is given in Peng and Schick (2003). It is of the form $v_i(x,\vartheta) = \sqrt{2}\cos(i\pi F_\vartheta(x))$. For this basis, Assumption 4 holds, and Assumption 5, with $L_i = ci$, follows from Assumption 3. In this case the rate conditions (2.1) are equivalent to $m_n^5/n \to 0$.

Suppose now that $\hat{\vartheta}$ is asymptotically linear, i.e.,

$$n^{1/2}(\hat{\vartheta} - \vartheta_0) = n^{-1/2}\sum_{k=1}^{n} J(X_{k-1},X_k) + o_p(1)$$

for some $J \in L_2^r(Q)$ with $E(J(X_0,X_1)|X_0) = 0$. Then $\hat{H}^*$ is asymptotically normal. We show in Section 3 that $\hat{H}^*$ is also efficient if $\hat{\vartheta}$ is efficient.

Our proof is similar to that for the bivariate model in Peng and Schick (2003). Their exponential inequality, Lemma 2, must be replaced by an appropriate version for Markov chains, which we state first.

Lemma 1. *Let $B = \{t \in \mathbf{R}^q : \|t\| \leqslant C\}$ be the closed ball of radius C in $\mathbf{R}^q$. Let u_t, $t \in B$, be a family of functions on S such that $u_0 = 0$ and, for some $L > 0$,*

$$|u_t(x) - u_s(x)| \leqslant L\|t - s\|, \quad x \in S, \ s,t \in B.$$

Suppose Assumption 2 holds. Then

$$U_n(t) = \frac{1}{n}\sum_{k=1}^{n}\left(u_t(X_k) - \int u_t \,\mathrm{d}F_{\vartheta_0}\right), \quad t \in B,$$

fulfills, for each $\varepsilon > 0$ and $n\varepsilon > 2LC/\eta$,

$$P\left(\sup_{t\in B}|U_n(t)| > 3\varepsilon\right) \leqslant 2\left(1 + \frac{2q^{1/2}LC}{\varepsilon}\right)^q \exp\left(-\frac{\eta^2(n\varepsilon - 2LC/\eta)^2}{2nL^2C^2}\right).$$

The proof of this result is identical to that of Lemma 2 in Peng and Schick (2003) for the case of independent observations. Instead of the classical Hoeffding inequality we now use the Markovian extension given by Glynn and Ormoneit (2002).

Proof of Theorem 1. It suffices to show

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{M_n}(\hat{\alpha}_{ni}v_i(X_{k-1},\hat{\vartheta})-a_Q(X_{k-1}))+\left(\int a_Q\varphi^\top\,\mathrm{d}F_{\vartheta_0}\right)(\hat{\vartheta}-\vartheta_0)=\mathrm{o_p}(n^{-1/2}), \tag{2.2}$$

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{N_n}(\hat{\beta}_{nj}v_j(X_k,\hat{\vartheta})-b_Q(X_k))+\left(\int b_Q\varphi^\top\,\mathrm{d}F_{\vartheta_0}\right)(\hat{\vartheta}-\vartheta_0)=\mathrm{o_p}(n^{-1/2}). \tag{2.3}$$

We only show (2.2); (2.3) is similar. Let V_m denote the linear span of $v_1(\cdot,\vartheta_0),\dots,v_m(\cdot,\vartheta_0)$. Let $a_n=\sum_{i=1}^{M_n}\alpha_{ni}v_i(\cdot,\vartheta_0)$ and $b_n=\sum_{j=1}^{N_n}\beta_{nj}v_j(\cdot,\vartheta_0)$ be chosen to minimize $\int(h(x,y)-a(x)-b(x))^2Q(\mathrm{d}x,\mathrm{d}y)$ over $a\in V_{M_n}$ and $b\in V_{N_n}$. As shown in Peng and Schick (2002), a_n and b_n are uniquely determined, and $a_n\to a_Q$ and $b_n\to b_Q$ in $L_2(F_{\vartheta_0})$. Assumption 1 and the Cauchy–Schwarz inequality imply that for $k=3,4,\dots$ and $f\in L_2(Q)$

$$|E[f(X_0,X_1)f(X_{k-1},X_k)]|=|E[f(X_0,X_1)K(Kf)(X_{k-2})]|\leqslant\lambda^{(k-2)/2}E[f^2(X_0,X_1)].$$

Thus we obtain for $C=1+2/(1-\lambda^{1/2})$ that

$$E\left[\left(\frac{1}{n}\sum_{k=1}^{n}f(X_{k-1},X_k)\right)^2\right]\leqslant\frac{C}{n}E[f^2(X_0,X_1)]\quad\text{for } f\in L_{2,0}(Q). \tag{2.4}$$

This immediately gives

$$\frac{1}{n}\sum_{k=1}^{n}a_n(X_{k-1})=\frac{1}{n}\sum_{k=1}^{n}a_Q(X_{k-1})+\mathrm{o_p}(1).$$

As in Peng and Schick (2003) we have

$$\sum_{i=1}^{M_n}\alpha_{ni}\int v_i(x,\hat{\vartheta})F_{\vartheta_0}(\mathrm{d}x)+\left(\int a_Q\varphi^\top\,\mathrm{d}F_{\vartheta_0}\right)(\hat{\vartheta}-\vartheta_0)=\mathrm{o_p}(1).$$

Thus it suffices to show

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{M_n}(\hat{\alpha}_{ni}-\alpha_{ni})v_i(X_{k-1},\hat{\vartheta})=\mathrm{o_p}(n^{-1/2}), \tag{2.5}$$

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{M_n}\alpha_{ni}\left(v_i(X_{k-1},\hat{\vartheta})-v_i(X_{k-1},\vartheta_0)-\int v_i(x,\hat{\vartheta})\,F_{\vartheta_0}(\mathrm{d}x)\right)=\mathrm{o_p}(n^{-1/2}). \tag{2.6}$$

As in Peng and Schick (2003) one can show

$$\sum_{i=1}^{M_n}(\hat{\alpha}_{ni}-\alpha_{ni})^2=\mathrm{O_p}\left(\frac{M_n(\Gamma_n+\Delta_n)}{n}\right). \tag{2.7}$$

The proof is essentially the same, but now using (2.4) to deal with the appropriate averages. It is shown in Peng and Schick (2003) that

$$n\sum_{i=1}^{M_n}\left(\int v_i(x,\hat\vartheta)\,F_{\vartheta_0}(\mathrm{d}x)\right)^2 = \mathrm{O_p}(M_n). \tag{2.8}$$

It follows from (2.4) that

$$n\sum_{i=1}^{M_n}\left(\frac{1}{n}\sum_{k=1}^{n} v_i(X_{k-1},\vartheta_0)\right)^2 = \mathrm{O_p}(M_n). \tag{2.9}$$

In view of (2.7)–(2.9), statement (2.5) is equivalent to

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{M_n}(\hat\alpha_{ni}-\alpha_{ni})\left(v_i(X_{k-1},\hat\vartheta)-v_i(X_{k-1},\vartheta_0)-\int v_i(x,\vartheta_0)F_{\vartheta_0}(\mathrm{d}x)\right) = \mathrm{o_p}(n^{-1/2}). \tag{2.10}$$

Relations (2.5) and (2.6) are verified as in Peng and Schick (2003), but now using the above Lemma 1 instead of their Lemma 2. □

3. Efficiency of the estimator

Let us now prove that $\hat H^*$ is efficient if an efficient estimator $\hat\vartheta$ for ϑ_0 is used. We need the following notation. Let $\bar K(y,\mathrm{d}x)$ denote the transition distribution of the reversed chain, defined by $F_{\vartheta_0}(\mathrm{d}x)K(x,\mathrm{d}y)=\bar K(y,\mathrm{d}x)F_{\vartheta_0}(\mathrm{d}y)$. For a function $g\in L_{2,0}(Q)$ write $Kg(x)=\int K(x,\mathrm{d}y)g(x,y)$ and $\bar Kg(y)=\int \bar K(y,\mathrm{d}x)g(x,y)$. Let K^j and $\bar K^j$ be the operators on $L_{2,0}(F_{\vartheta_0})$ defined by $K^jf(X_0)=E(f(X_j)\,|\,X_0)$ and $\bar K^jf(X_j)=E(f(X_0)\,|\,X_j)$, $j=1,2,\ldots$. Let $U=\sum_{j=0}^{\infty}K^j$ and $\bar U=\sum_{j=0}^{\infty}\bar K^j$ be the corresponding potentials. Let now

$$\mathscr{T}=\{t\in L_2(Q): Kt=0\},$$

and let A be the operator from $L_{2,0}(F_{\vartheta_0})$ into $\mathscr{T}$ defined by $Af(x,y)=Uf(y)-KUf(x)$.

We can now recall the characterization of efficient estimators in Kessler et al. (2001). Consider (Hellinger differentiable) perturbations $K_{nt}(x,\mathrm{d}y)\doteq K(x,\mathrm{d}y)(1+n^{-1/2}t(x,y))$ consistent with the parametric model for the stationary law. The space $\mathscr{T}_*$ of all such functions t is called the *tangent space* for our model. It is a subset of $\mathscr{T}$. An r-dimensional functional χ of K is called *differentiable* with *gradient* g if $g\in\mathscr{T}^r$ and

$$n^{1/2}(\chi(K_{nt})-\chi(K))\to\int gt\,\mathrm{d}Q \quad \text{for all } t\in\mathscr{T}_*.$$

The *canonical gradient* is the (componentwise) projection g_* of g onto $\mathscr{T}_*^r$. An estimator $\hat\chi$ for χ is called *regular* if there is a random vector L such that

$$n^{1/2}(\hat\chi-\chi(K_{nt}))\Rightarrow L \quad \text{under } K_{nt} \text{ for all } t\in\mathscr{T}_*.$$

An estimator $\hat{\chi}$ for χ is called *asymptotically linear* with *influence function h* if $h \in \mathscr{T}^r$ and

$$n^{1/2}(\hat{\chi} - \chi(K)) = n^{-1/2}\sum_{k=1}^{n} h(X_{k-1}, X_k) + o_p(1).$$

An estimator is regular and efficient if and only if it is asymptotically linear with influence function equal to the canonical gradient. Moreover, an asymptotically linear estimator is regular if and only if its influence function is a gradient. In particular, the canonical gradient can be obtained as the projection onto $\mathscr{T}_*^r$ of the influence function of an arbitrary regular and asymptotically linear estimator.

As shown in Kessler et al. (2001), the tangent space for our model is

$$\mathscr{T}_* = \{t \in \mathscr{T} : \bar{U}\bar{K}t \in [\varphi]\},$$

where $[\varphi]$ is the linear span of the Hellinger derivative φ. Moreover, the influence function of an efficient estimator $\hat{\vartheta}$ of ϑ_0 is

$$g_*(x, y) = \left(\int e_* \varphi^\top \, dF_{\vartheta_0}\right)^{-1} A e_* \quad \text{with } e_* = (\bar{U}\bar{K}A)^{-1}\varphi.$$

Note that $\bar{U}\bar{K}$ corresponds to $\bar{V}$ in Kessler et al. (2001). If an efficient estimator $\hat{\vartheta}$ is used, then by Theorem 1 the influence function of our estimator $\hat{H}^*$ is

$$h_*(x, y) = h_0(x, y) - a_Q(x) - b_Q(y) + D^\top g_*(x, y),$$

where $h_0 = h - \int h \, dQ$. Efficiency of $\hat{H}^*$ follows if we show that h_* is in $\mathscr{T}_*$ and equals the projection of the influence function of the empirical estimator $\hat{H}$, which is

$$\tilde{h}(x, y) = h_0(x, y) - Kh_0(x) + AKh_0(x, y)$$

by Greenwood and Wefelmeyer (1995). Showing these two properties amounts to showing that $Kh_* = 0$ and $\int \tilde{h}t \, dQ = \int h_* t \, dQ$ for all $t \in \mathscr{T}_*$.

By definition of a_Q and b_Q we have that $h_Q(X_0, X_1) = h_0(X_0, X_1) - a_Q(X_0) - b_Q(X_1)$ is orthogonal to $a(X_0) + b(X_1)$ for all $a, b \in L_{2,0}(F_{\vartheta_0})$. Thus $E(h_Q(X_0, X_1) \,|\, X_0) = 0$ and $E(h_Q(X_0, X_1)|X_1) = 0$. The former shows that $Kh_* = Kh_Q + D^\top K g_* = 0$. It also gives $Kh_0 - a_Q - Kb_Q = 0$, which implies

$$a_Q + b_Q = (I - K)b_Q + Kh_0. \tag{3.1}$$

Now fix $t \in \mathscr{T}_*$. Then $\bar{U}\bar{K}t = \varphi^\top u$ for some $u \in \mathbf{R}$. We have

$$\int \tilde{h}t \, dQ = \int h_0 t \, dQ + \int AKh_0 \cdot t \, dQ = \int h_Q t \, dQ + \int (A(I - K)b_Q + AKh_0)t \, dQ.$$

Here we have used that $b = U(I - K)b$ and that $Kt = 0$. It was shown in Kessler et al. (2001) that $\int g_* t \, dQ = u$ and

$$\int tAf \, dQ = \int \bar{U}\bar{K}t \cdot f \, dF_{\vartheta_0} = \int f\varphi^\top u \, dF_{\vartheta_0}.$$

In particular, if $f = a_Q + b_Q$, we get from (3.1) that

$$\int (A(I-K)b_Q + AKh_0)t \,\mathrm{d}Q = D^\top u = D^\top \int g_* t \,\mathrm{d}Q.$$

Hence, we get $\int \tilde{h} t \,\mathrm{d}Q = \int h_* t \,\mathrm{d}Q$.

For the case $r = 1$, Kessler et al. (2001) construct an efficient estimator $\hat{\vartheta}$ of ϑ_0 under the additional assumption of *continuous* Hellinger differentiability of F_ϑ. The construction carries over to r-dimensional ϑ.

References

Bibby, B.M., Sørensen, M., 1995. Martingale estimating functions for discretely observed diffusion processes. Bernoulli 1, 17–39.

Bibby, B.M., Sørensen, M., 1996. On estimation for discretely observed diffusions: a review. Theory Stoch. Process. 2 (18), 49–56.

Bibby, B.M., Sørensen, M., 1997. A hyperbolic diffusion model for stock prices. Finance Stoch. 1, 25–41.

Bibby, B.M., Sørensen, M., 2001. Simplified estimating functions for diffusion models with a high-dimensional parameter. Scand. J. Statist. 28, 99–112.

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

Bickel, P.J., Kwon, J., 2001. Inference for semiparametric models: Some questions and an answer (with discussion). Statist. Sinica 11, 863–960.

Bickel, P.J., Ritov, Y., Wellner, J.A., 1991. Efficient estimation of linear functionals of a probability measure *P* with known marginal distributions. Ann. Statist. 19, 1316–1346.

Deming, W.E., Stephan, F.F., 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Statist. 11, 427–444.

Glynn, P.W., Ormoneit, D., 2002. Hoeffding's inequality for uniformly ergodic Markov chains. Statist. Probab. Lett. 56, 143–146.

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

Greenwood, P.E., Wefelmeyer, W., 1999. Reversible Markov chains and optimality of symmetrized empirical estimators. Bernoulli 5, 109–123.

Greenwood, P.E., Schick, A., Wefelmeyer, W., 2001. Comment [on Bickel and Kwon, 2001]. Statist. Sinica 11, 892–906.

Kessler, M., 2000. Simple and explicit estimating functions for a discretely observed diffusion process. Scand. J. Statist. 27, 65–82.

Kessler, M., Sørensen, M., 1999. Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli 5, 299–314.

Kessler, M., Sørensen, M., 2002. On time-reversibility and estimating functions for Markov processes. MaPhySto Research Report 2002-25, Department of Mathematical Sciences, University of Aarhus. http://www.maphysto.dk/.

Kessler, M., Schick, A., Wefelmeyer, W., 2001. The information in the marginal law of a Markov chain. Bernoulli 7, 243–266.

Pedersen, A.R., 1995a. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Statist. 22, 55–71.

Pedersen, A.R., 1995b. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Bernoulli 1, 257–279.

Penev, S., 1990. Convolution theorem for estimating the stationary distribution of Markov chains. C. R. Acad. Bulgare Sci. 43, 29–32.

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

Peng, H., Schick, A., 2002. On efficient estimation of linear functionals of a bivariate distribution with known marginals. Statist. Probab. Lett. 59, 83–91.

Peng, H., Schick, A., 2003. Estimation of linear functionals of bivariate distributions with parametric marginals. Technical Report, Department of Mathematical Sciences, Binghamton University. http://math.binghamton.edu/anton/preprint.html.

Sørensen, M., 1997. Estimating functions for discretely observed diffusions: a review. In: Basawa, I.V., Godambe, V.P., Taylor, R.L. (Eds.), Selected Proceedings of the Symposium on Estimating Functions. Lecture Notes—Monograph Series 32, Institute of Mathematical Statistics, Hayward, California, pp. 305–325.