On parameter estimation for ergodic Markov chains with unbounded loss functions

A.Yu.Veretennikov
Institute for Information Transmission Problems
19 Bolshoy Karetnii, 101447, Moscow, Russia
email: ayu@sci.lpi.ac.ru *

November 13, 1998

1991 Mathematics Subject Classification. 62-02, 62F12

Keywords and phrases. Asymptotic normality, efficiency, polynomial loss function, maximum likelihood estimation, Hajek - Le Cam efficiency.

*The work was supported by grants DFG 436 RUS 17/133/97 and FRBR 98-01-00431. The preprint was written during the visit of the author to Weierstrass Institute for Applied Analysis and Stochastics, Berlin

Abstract

We establish the Hajek - Le Cam asymptotic efficiency of maximum likelihood estimators for "polynomially ergodic" Markov regular experiments in the class of loss functions with a polynomial growth.

The contents

1 Introduction

Parameter estimation of random processes is a classical problem which remains to attract an attention. Various settings of this problem were investigated in details by many authors for independent experiments, Markov processes, semimartingales, diffusion processes. In particular, the local asymptotical normality (LAN) property (see Le Cam (1986)) which is crucial for the modern notion of asymptotic efficiency was studied for Markov processes in Roussas (1972), Milhaud, Oppenheim, Viano (1983), Hoepfner, Jacod, Ladelli (1990), and, in particular, for certain diffusion processes in Ibragimov, Khas'minskij (1981), Kutoyants (1984), (1994) et al. In Ibragimov, Khas'minskij (1981) the asymptotic efficiency with loss functions of polynomial growth was established under general assumptions. There are examples how to check those conditions (see Ibragimov, Khas'minskij (1984)) but in Markov case usual conditions are rather restrictive, they imply, as a rule, the exponential bounds and hence give much more than needed for polynomial loss functions.

The aim of this paper is to give weak sufficient condition for the asymptotic efficiency of the MLE for loss functions with a polynomial growth for a wide class of Markov processes which we call "polynomially ergodic". The "technical" definition of this class (see section 3 below) can be checked, however, for rather natural examples (see Proposition 1 below) using the inequalities for mixing and convergence rate for Markov processes from Veretennikov (1997), (1998). The advantage of the use of this class is that it provides the efficiency indeed for polynomial loss functions and not more, e.g. not exponential. For the latter one can, in fact, consider "exponentially ergodic" classes analogously.

Here we formulate one example for which our general theorem 2 below works: $\theta \in (-1, +1)$,

$$X_{n+1} = X_n - (2+\theta)\,\mathrm{sign}\,X_n(1+|X_n|)^{-1}\log(1+|X_n|) + W_{n+1},$$

$$\{W_n\} \sim \mathcal{N}(0,1), \ \text{ i.i.d.}$$

2 Setting of the problem

One observes a Markov process $X_n \in R^1$, $n \geq 0$ which depends on a parameter $\theta \in (a,b) \subset R^1$ via its transition density $f_\theta(x,y)$ w.r.t. the Lebesgue measure (for simplicity).

We assume that the process X_n is ergodic under any θ with certain mixing and convergence bounds uniformly in θ. Those bounds will be fixed a bit later.

Assume that f is continuously differentiable in θ in some neighbourhood of (a,b) and so that the standard differentiation is possible under the integrals while deriving the Cramér-Rao inequality.

The goal is to establish the asymptotic efficiency of the maximum likelihood estimator (MLE) for loss functions with a polynomial growth.

The class of ergodic Markov processes with "polynomial" ergodicity is exposed in section 3. A remark 2 in section 4 is devoted to the Cramèr-Rao inequality for our experiments. Remark 4 in section 5 concerns the uniform asymptotic normality property. In section 6 we prove the consistency and asymptotic efficiency of the MLE in the Hájek - Le Cam sense. The method of Ibragimov and Khas'minskij is used. Also we use some elementary facts from the martingale theory.

To formulate our main result about the efficiency of the MLE we need to introduce our polynomial ergodic classes, make some useful remarks and remind some facts and definitions from Ibragimov, Khas'minskij (1981). Because of this the rigorous statement of this result is in the last section. A non-rigorous formulation is as follows:
an MLE is asymptotically efficient in the sense of Hajek - LeCam for loss functions with a polynomial growth in the class of polynomially ergodic Markov process (see section 3 below) under some additional regularity and growth assumptions on transition densities and identification type conditions.

3 Polynomial ergodicity

We assume that the initial data $X_0 = x_0$ is non-random. The changes in the general case will be evident. Denote by $S_{m,m',k}$ $(m, m', k > 0)$ the class of ergodic Markov processes which satisfy the bounds

(i)
$$var(\mu_n^x - \mu) + \beta_n^x \leq C(1+|x|^m)(1+n)^{-(k+1)},$$

(ii)

$$\sup_t E_x|X_t|^{m'} \le C(1+|x|^m),$$

(iii)

$$\int |x|^m \, \mu(dx) < \infty.$$

(iv) For any interval $Q = (-N, N)$ with N large enough there exist such constants $\kappa = \kappa(Q,k) > 0$ and $C_k = C_{Q,k} < 1$ for which

$$P_x(\tau_n^Q \ge \kappa^{-1} n) \le C_k(1+|x|^m)(1+n)^{-k}, \quad n = 1, 2, \ldots,$$

where $\tau_0^Q = 0$, $\tau_{n+1}^Q = \inf(t \ge \tau_n^Q + 1 : \, X_t \in Q)$.

Here $X_0 = x$, $var(\mu^x - \mu)$ is a distance in variance, μ is a (unique) invariant measure of the process X, β^x is a complete regularity coefficient

$$\beta_t^x = \sup_{s\ge 0} E \; sup_{B\in F_{\ge t+s}}(P(B|F_{\le s}) - P(B)),$$

and $F_D = \sigma\{X_s : \; s \in D \subset R^1\}$.

Proposition 1 (Veretennikov (1998)) *Let process X satisfy the recurrent equation*

$$X_{n+1} = f(X_n) + W_{n+1}, \quad (W_n) \; i.i.d.,$$

under conditions

$$EW_0 = 0, \;\; E|W_0|^{m_0} < \infty, \;\; m_0 > 4,$$

f is locally bounded,

$$(|f(x)|/|x| - 1)|x|^2 \to -\infty, \quad |x| \to \infty,$$

and the "process on Q" (i.e. the process in the successive times of hitting the set Q) satisfies the Doeblin type condition for any N large enough, namely, there exists such $n_0 > 0$ that

$$(D_\ell) \qquad \inf_{x,x'\in Q} \int \min\left\{\frac{P^Q(n_0, x, dy)}{P^Q(n_0, x', dy)}, 1\right\} P^Q(n_0, x', dy) > 0,$$

where $P(dy)/P'(dy)$ means the derivative of the absolute continuous part of one measure w.r.t. another and P^Q denotes the transition probability of the "process on Q" for n_0 steps. Then $X \in S_{m,m',k}$ for any

$$2 < m' < m - 2 \le m_0 - 2, \quad 0 < k < (m'-2)/2. \tag{1}$$

Moreover, one can choose κ which does not depend on k and C which does not depend on Q!!?

If additionally $m_0 = \infty$ then $X \in S := \bigcap\{S_{m,m',k} : \; (m,m',k)$ satisfy condition (1) with $m_0 = \infty\}$.

In particular, the example in the introduction satisfies all conditions of proposition 1 uniformly in $\theta \in (-1,1)$. It is sufficient (but not necessary) for condition (D_ℓ) that the density of W_n is positive everywhere.

In the sequel we will use the assumption $X \in S$ rather than $X \in S_{m',m,k}$ with some m', m, k for simplicity.

4 Remarks

Denote

$$L_{\theta,n} = \sum_{i=1}^{n} \log \frac{f_\theta(X_{i-1}, X_i)}{f_{\theta_0}(X_{i-1}, X_i)} \equiv \sum_{i=1}^{n} h_\theta(X_{i-1}, X_i).$$

Let $g_n(\theta) = E_\theta \theta_n$ for any estimator θ_n.

Remark 1 (Cramèr-Rao inequality)

$$D_\theta \theta_n \geq \frac{(g_n'(\theta))^2}{E_\theta (L'_{\theta,n})^2}.$$

The proof is standard and we omit it. The denominator in the r.h.s. is a Fisher information. It may be important to know its asymptotics.

Remark 2 (Fisher information asymptotics) *Let $(X_n) \in S$ and let there exist such $m \geq 0$ that*

$$|h_\theta'(x,x')| \leq C(1 + |x|^{m/2} + |x'|^{m/2}). \tag{2}$$

Then

$$n^{-1} E_\theta (L'_{\theta,n})^2 \to \sigma_\theta^2 := E_\theta^{inv} (h_\theta'(X_0, X_1))^2 \tag{3}$$

uniformly in $\theta \in \Theta$.

Indeed, we have

$$n^{-1} E_\theta (L'_{\theta,n})^2 = (1/n) \sum_{i=1}^{n} E_\theta (h_\theta'(X_{i-1}, X_i))^2$$

$$+(2/n) \sum_{1 \leq i < j \leq n} E_\theta h_\theta'(X_{i-1}, X_i) h_\theta'(X_{j-1}, X_j).$$

Let us omit the parameter θ for the moment. All bounds will be uniform. We estimate

$$|E(h'(X_{i-1}, X_i))^2 - E^{inv}(h'(X_{i-1}, X_i))^2|$$

$$= |\int (h_\theta'(x,x'))^2 (\mu_{i-1}^{x_0} - \mu)(dx) f(x,x') dx|$$

$$\leq C \int (1 + |x|^m + |x'|^m) |\mu_{i-1}^{x_0} - \mu|(dx) f(x,x') dx$$

$$\leq C \int (1 + |x|^m) |\mu_{i-1}^{x_0} - \mu|(dx) \leq C(1 + |x_0|^m)(1+i)^{-(k+1)}.$$

So,

$$|(1/n) \sum_{i=1}^{n} E_\theta (h_\theta'(X_{i-1}, X_i))^2 - E_\theta^{inv} (h_\theta'(X_0, X_1))^2|$$

$$\leq C(1 + |x_0|^{m/2}) n^{-1} \sum_{i \geq 0} (1+i)^{-(k+1)} \to 0.$$

¿From our assumptions the equalities follow for any $i \neq 0$:

$$E_\theta h'_\theta(X_0, X_1) h'_\theta(X_i, X_{i+1}) = 0. \tag{4}$$

This shows remark 2.

Of course, $\sigma_\theta^2 \geq 0$. But *we can and will assume that*

$$0 < C^{-1} < \sigma_\theta^2 < C < \infty, \quad \textit{for some } C > 0 \tag{5}$$

for any θ. It is reasonable because of the formula (3) and because σ_θ^2 is continuous in θ. Indeed, it is easy to show that the invariant density is continuous in θ and so is σ_θ^2.

Further, following Borovkov (1988), §§16, 20, one can consider the class of asymptotically unbiased estimators $\tilde{K}_0$, i.e. such estimators θ_n that for any θ

$$g_n(\theta) - \theta = o(1/\sqrt{n}) \text{ and } g'_n(\theta) - 1 = o(1), \quad n \to \infty.$$

Remarks 1 and 2 imply the following fact.

Remark 3 (limiting Cramèr-Rao inequality) *In the class* $\tilde{K}_0$ *under conditions of remark 2*

$$\liminf_{n \to \infty} n D_\theta \theta_n \geq \sigma_\theta^2. \tag{6}$$

In the other words, σ_θ^2 is a limiting normalized Fisher information for our markovian experiment with a normalizing coefficient $1/n$. In particular, it follows from theorem 2 below that the MLE belongs to this class and is asymptotically effective there. However, our main interest is, as we told, the efficiency in the sense of Hajek - LeCam.

5 Asymptotic normality

LAN was established in Roussas (1972), Ogata, Inagaki (1977), Milhaud, Oppenheim, Viano (1983), Hoepfner, Jacod, Ladelli (1990) et al. under various ergodocity and stationarity assumptions. Notice that a uniform positive recurrence condition is satisfied for our class S. We will show a uniform asymptotic normality (see Ibragimov, Khas'minskij (1981)) which is needed in the next section.

Consider the likelihood function

$$Z_{\theta,n}(u) = \prod_{k=1}^{n} \frac{f_{\theta+u/\sqrt{n}}}{f_\theta}(X_{k-1}, X_k).$$

Let us also define

$$Z_{\theta,\ell,n}(u) = \prod_{k=\ell}^{n} \frac{f_{\theta+u/\sqrt{n}}}{f_\theta}(X_{k-1}, X_k).$$

For the sake of simplicity we suppose that all denominators in this expression are positive. It is well-known how one can relax this assumption (cf. Ibragimov, Khas'minskij (1981)).

The experiment satisfies the uniform asymptotic normality property if (cf. Ibragimov, Khas'minskij (1981), definition 2.2.2, we gives an equivalent form using the continuity of σ_θ^2 in θ) the log *of the likelihood function can be represented in the form*

$$\log Z_{\theta,n}(u) = u\Delta_{\theta,n} - u^2\sigma_\theta^2/2 + \psi_n(u,\theta),$$

where

$$\Delta_{\theta,n} \stackrel{P_\theta}{\Longrightarrow} N(0,\sigma_\theta^2), \quad \psi_n(u,\theta) \stackrel{P_\theta}{\to} 0, \quad n \to \infty,$$

and, moreover, for any compact $K \subset \Theta$, any u, any $\theta \in K$ and any sequences $\theta_m \in K$, $u_m \to u$, $\theta_m \to \theta$, $n_m \to \infty$, $\theta_m + u_m/\sqrt{n_m} \in K$ also

$$\log Z_{\theta_m,n_m}(u_m) = u\Delta_{n_m,\theta_m} - u^2\sigma_\theta^2/2 + \psi_{n_m}(u_m,\theta_m),$$

where

$$\Delta_{\theta_m,n_m} \stackrel{P_{\theta_m}}{\Longrightarrow} N(0,\sigma_\theta^2), \quad \psi_{n_m}(u_m,\theta_m) \stackrel{P_{\theta_m}}{\to} 0, \quad m \to \infty, \tag{7}$$

Remark 4 *Let assumptions of remark 2 be satisfied. Then*

$$n^{-1/2}L'_{\theta,n} \Longrightarrow N(0,\sigma_\theta^2)$$

uniformly in $\theta \in \Theta$, i.e.

$$\Delta_{\theta_m,n_m} \stackrel{P_{\theta_m}}{\Longrightarrow} N(0,\sigma_\theta^2), \quad m \to \infty$$

with notations as in the definition above.

Indeed, first of all, because of the convergence in variation it suffices to establish the desired property in the stationary regime of X. For this one can use theorem 18.5.3 from [5]. Remind that theorem (an equivalent form):
Let a stationary sequence ξ_n satisfy the strong mixing property with a coefficient $\alpha(n)$, there exists $\delta > 0$

$$E|\xi_n|^{2+\delta} < \infty, \quad \textit{and} \quad \sum_n (\alpha(n))^{\delta/(2+\delta)} < \infty. \tag{8}$$

Then $\sigma_\theta^2 = E\xi_0^2 + 2\sum_{j\ge1} cov(\xi_0,\xi_j) < \infty$ and

$$n^{1/2}\sum_{j=1}^n \xi_j \Longrightarrow N(0,\sigma_\theta^2).$$

It is a straightforward consequence of the proof of this theorem (see Ibragimov, Linnik (1971), ch. 18) that the uniform convergence (cf. (7)) holds if one assumes the uniform convergence in both parts of (8).

Let us check these assumptions. We use the inequality (cf. [5])

$$\alpha(t) \leq \beta(t),$$

where $\beta(t) = E^{inv}\beta^{X_0}(t)$. Notice that conditions (i), (iii) in the description of the ergodic classes imply the bound

$$\beta(t) \leq C_k(1+t)^{-(k+1)} \quad (\forall k).$$

Then the existence of $\delta > 0$ follows from the condition $m < m_0$ $(= \infty)$ while $\alpha(n)$ decreases faster than any polynomial. So all assumptions of theorem 18.5.3 are satisfied. Therefore, we get the desired uniform weak convergence since all bounds are uniform w.r.t. θ. The remark follows.

Proposition 2 (uniform asymptotic normality) *Let assumptions of remark 2 be satisfied. Then the experiment $\{X, P_\theta\}$ satisfies a uniform asymptotic normality property.*

Proof follows from *uniform* bounds analogous to those in the proof of theorem 3.4.1 from Ibragimov, Khas'minskij (1981) or in other papers on the subject without large changes. So we prefer to propose slightly different way, which use, essentially, also very close approximation idea.

It follows from Harris' representation formula for invariant measures (see below, for the reference cf. Meyn and Tweedie (1993)) and the description of the class S that we can smooth all distributions of our processes in the following way. At each moment n we add to the value X_n an independent normal value η_n with a zero mean and a small variance, say, ϵ. The observation is that such a perturbed process will be still in the call S (in fact, it even does not depend on the value of the variance of the perturbation). Hence, our new perturbed process will have infinitely smooth distributions. Now, if we prove any estimate which only concerns first derivatives, it is very likely that then we can pass to a limit when the variance of our perturbation tends to zero. Indeed, denote a perturbed process by X^ϵ and perturbed likelihood function by Z^ϵ. Suppose we have an assertion

$$\log Z^\epsilon_{\theta_m,n_m}(u_m) = u\Delta_{n_m,\theta_m,\epsilon} - u^2\sigma^2_{\theta,\epsilon}/2 + \psi_{n_m}(u_m,\theta_m,\epsilon),$$

with

$$\Delta_{\theta_m,n_m,\epsilon} \stackrel{P_{\theta_m,\epsilon}}{\Longrightarrow} N(0,\sigma^2_{\theta,\epsilon}), \quad \psi_{n_m}(u_m,\theta_m,\epsilon) \stackrel{P_{\theta_m,\epsilon}}{\to} 0, \quad m \to \infty.$$

Then the desired result will follow if we show that $\sigma^2_{\theta,\epsilon} \to \sigma^2_\theta$, $\epsilon \to 0$. But both values are expectations w.r.t. invariant measures, so we should pass to the limit under the integral. And this is exactly what Harris' representation allows to do. Indeed, due to the properties of the class S we have (using evident new notations)

$$\sigma^2_{\theta,\epsilon} = \int (h'_\theta(x,x'))^2\, \mu^{\theta,\epsilon}_Q(dx) f_\theta(x,x')\, dx'$$

$$= \left(E^{inv,\epsilon}_Q \tau\right)^{-1} E^{inv,\epsilon}_Q \textstyle\sum_{i=1}^{\tau} (h'_\theta(X_i,X_{i+1}))^2$$

(Harris' representation where Q is some compact, $\tau = \inf k \geq 1 : \ X_k \in Q$ and $E_Q^{inv,\epsilon}$ means the expectation w.r.t. the invariant measure $\mu_Q^{\theta,\epsilon}$ of the *process on* Q). For Q fixed, the invariant measures on Q depend weakly continuously on $\epsilon \geq 0$ due to the (geometrical) convergence.

Now, choosing Q a bit larger or a bit smaller, we get the convergence of both terms in the above formula to their limits as $\epsilon \to 0$ by virtue of the conditions $(H1) - (H4)$. Namely, if $\nu > 0$, $Q = [-N, N]$ and $Q_{\pm}^{\nu} = [-N \mp \nu, N \pm \nu]$ then for ϵ small enough one obtains

$$\limsup_{\epsilon \to 0} E_{Q_+^{\nu}}^{inv,\epsilon} \tau \leq \liminf_{\epsilon \to 0} E_Q^{inv,\epsilon} \tau \leq \limsup_{\epsilon \to 0} E_Q^{inv,\epsilon} \tau \leq \limsup_{\epsilon \to 0} E_{Q_-^{\nu}}^{inv,\epsilon} \tau$$

and analogous inequalities hold also for the numerators. The assertion $\sigma_{\theta,\epsilon}^2 \to \sigma_{\theta}^2, \ \epsilon \to 0$ follows from these inequalities after passing to the limit as $\nu \to 0$.

We demonstrate the use of this remark now. So, we assume for a while that h has three derivatives in θ which are polynomially bounded. We have,

$$\log Z_{\theta,n}(u) = \log \textstyle\prod_{k=1}^n [1 + h'_\theta(X_{k-1}, X_k) u/\sqrt{n}$$

$$+ h''_\theta(X_{k-1}, X_k) u^2/(2n) + h'''_{\tilde{\theta}_{n,k}}(X_{k-1}, X_k) u^3/(6n^{3/2})]$$

$$\equiv \textstyle\sum_{1 \leq k \leq n} \log[1 + \eta_{k,n}^{(1)} + \eta_{k.n}^{(2)} + \eta_{k,n}^{(3)}]$$

with some $\tilde{\theta}_{n,k} = \theta + a_n u/\sqrt{n}$, $|a_n| \leq 1$.

Denote $A_n = \{\omega \in \Omega : \ \max_{1 \leq k \leq n} |\eta_{k,n}^{(i)}| \leq 1/4 \ \ i = 1, 2, 3\}$. Then $P_\theta(A_n) \to 1, \ \ n \to \infty$. Indeed (cf. Ibragimov, Khas'minskij (1981), proof of theorem 2.1.1),

$$P_\theta(\max_{k \leq n} n^{-1/2} |h'_\theta(X_{k-1}, X_k)| > 1/4) \leq \sum_{k=0}^n P_\theta(n^{-1/2} |h'_\theta(X_{k-1}, X_k)| > 1/4)$$

$$\leq \sum_{k=0}^n n^{-(1+\delta/2)} 4^{(2+\delta)} E_\theta |h'_\theta(X_{k-1}, X_k)|^{2+\delta} \leq C(X_0) n^{-\delta/2} \to 0, \quad n \to \infty.$$

Other parts with $\eta^{(2)}$ and $\eta^{(3)}$ are estimated even simpler.

One has on A_n,

$$\log Z_{\theta,n}(u) = \sum_{k=0}^n \Bigg\{ [h'_\theta(X_{k-1}, X_k) u/\sqrt{n}$$

$$+ h''_\theta(X_{k-1}, X_k) u^2/(2n) + h'''_{\tilde{\theta}_{n,k}}(X_{k-1}, X_k) u^3/(6n^{3/2})]$$

$$-(1/2) \Big[h'_\theta(X_{k-1}, X_k) u/\sqrt{n}$$

$$+h''_\theta(X_{k-1},X_k)u^2/(2n)+h'''_{\tilde\theta_{n,k}}(X_{k-1},X_k)u^3/(6n^{3/2})\Big]^2$$

$$+(1/3)a\Big[h'_\theta(X_{k-1},X_k)u/\sqrt{n}$$

$$+h''_\theta(X_{k-1},X_k)u^2/(2n)+h'''_{\tilde\theta_{n,k}}(X_{k-1},X_k)u^3/(6n^{3/2})\Big]^3\Big\}$$

Now, we get

$$n^{-1/2}u\sum_{k=0}^n h'_\theta(X_{k-1},X_k)\Longrightarrow uN(0,\sigma^2_\theta), \tag{9}$$

$$n^{-1}\sum_{k=0}^n h''_\theta(X_{k-1},X_k)\stackrel{P_\theta}{\to}E^{inv}h''_\theta(X_0,x_1)=0,$$

$$E_\theta n^{-3/2}\sum_{k=0}^n|h'''_{\tilde\theta_n}(X_{k-1,X_k})|\le n^{-1/2}C(1+|X_0|^m)\to 0,\quad n\to\infty,$$

$$-(1/2)n^{-1}\sum_{k=0}^n|h'_\theta(X_{k-1},X_k)|^2\stackrel{P_\theta}{\to}-\sigma^2_\theta/2.$$

Other terms tend to zero at least with the rate $n^{-1/2}$. Moreover, we already know (see remark 4) that the weak convergence in (9) is uniform and it is easy to see that all other limits are also uniform in θ. Thus, the uniform asymptotic normality holds under additional assumption about three derivatives. Due to the remark above, this assumption is not a restriction, hence, proposition 2 is proved.

Fix some $\theta_0\in\Theta$. We call a one step Fisher information the function

$$I_1(\theta,x)=E_{\theta_0}\left(\frac{f'_\theta(x,X_1)}{f'_{\theta_0}(x,X_1)}\right)^2.$$

We *assume* that $I_1(\theta,x)$ is continuous in θ.

Theorem 1 (on polynomial bounds) *Let conditions of Proposition 2 be satisfied as well as additional identificability assumptions*

(Δ_1) $\forall Q=(-N,N),\ K\subset\Theta,\ K$ *compact,*

$$0<\inf_{\theta\in K,x\in Q}I_1(\theta,x)\le\sup_{\theta\in K,x\in Q}I_1(\theta,x)<\infty;$$

(Δ_2) $\forall Q=(-N,N),\ \forall K\subset\Theta,\ K$ *compact and for any* $\delta>0$,

$$\inf_{x_1\in Q}\inf_{\theta\in K}\inf_{h:\,\theta+h\in\Theta,\,|h|\ge\delta}\int(f^{1/2}_{\theta+h}(x_1,x_2)-f^{1/2}_\theta(x_1,x_2))^2\,dx_2>0.$$

Then for any $k=1,2,\ldots$ *and* $n=1,2,\ldots$

$$E_\theta Z^{1/2}_{\theta,n}(u)\le\exp(-cu^2)+C_k(1+n)^{-k}.$$

Comment. Because of our example in the introduction, it is not natural, in fact, to require uniform inequalities in conditions $(\Delta_{1,2})$ w.r.t. $x \in R^1$. This is the reason why there is no exponential inequalities similar to those in the i.i.d. case (cf. Ibragimov, Khas'minskij (1981), proof of theorem 3.3.2). Roughly speaking, the idea is the following. Due to polynomial ergodicity, our process X visits any compact Q often enough (occupation time $\geq \kappa n$ as $n \to \infty$) with a probability close to 1 (namely, this probability $\geq 1 - C_k(1+n)^{-k}$ for any k). When X belongs to Q, we get exponential inequalities. But the probability $(X \notin Q)$ has no exponential bound, in general, only polynomial ones.

Proof. First of all notice that assumptions (Δ_1) and (Δ_2) imply the inequality

$$\inf_{\theta\in K}\inf_{x\in Q}\inf_{h:\,\theta+h\in\Theta,\,|h|\geq\delta}\int |f_{\theta+h}^{1/2}(x,x') - f_\theta^{1/2}(x,x')|^2\,dx' \geq c\delta^2(1+\delta^2)^{-1}$$

(see Ibragimov, Khas'minskij (1981), proof of theorem 3.3.2) which in turn implies

$$\sup_\theta E_\theta Z_{\theta,1}^{1/2}(u/\sqrt{n}) \leq \exp(-cu^2n^{-1}) \tag{10}$$

on the set $\{x \in Q\}$ (see Ibragimov, Khas'minskij (1981), lemma 1.5.3).

Consider the stopping times $\{\tau_t,\ t = 1,2,\ldots\}$:

$$\tau_1 = \inf(s = 0,1,\ldots:\ X_s \in Q), \quad \tau_{t+1} = \inf(s = \tau_t+1, \tau_t+2,\ldots:\ X_s \in Q).$$

Choose k and $Q = Q(k)$ s.t. $\kappa E^{inv}\tau_1 < 1$. It follows from assumption (iv) that

$$P(\tau_{[\kappa n]} \geq n) \leq \frac{C_k}{(1+n)^k}, \quad \forall n$$

because we can include the (fixed) initial data X_0 in C_k. Now let us estimate

$$E_\theta Z_{\theta,n}^{1/2}(u) = E_\theta Z_{\theta,n}^{1/2}(u)1(\tau_{[\kappa n]} \geq n) + E_\theta Z_{\theta,n}^{1/2}(u)1(\tau_{[\kappa n]} < n).$$

We have,

$$E_\theta Z_{\theta,n}^{1/2}(u)1(\tau_{[\kappa n]} \geq n) \leq (EZ_{\theta,n}(u))^{1/2}(P(\tau_{[\kappa n]} \geq n))^{1/2} \leq \left(\frac{C_k}{(1+n)^k}\right)^{1/2}.$$

Further, since $Z_{\theta,n}^{1/2}$ is a P_θ-supermartingale with $Z_{\theta,0} = 1$ then

$$1(\tau_{[\kappa n]} < n)E_\theta\left(Z_{\theta,\tau_{[\kappa n]}+1,n}^{1/2} \mid F_{\tau_{[\kappa n]}}\right) \leq 1(\tau_{[\kappa n]} < n),$$

due to the optional theorem for supermartingales. Hence, the second term can be estimated as

$$E_\theta Z_{\theta,n}^{1/2}(u)1(\tau_{[\kappa n]} < n) \leq E_\theta Z_{\theta,\tau_{[\kappa n]}}^{1/2}(u).$$

Let us show that

$$E_\theta\left(Z_{\theta,\tau_{k-1}+1,\tau_k}^{1/2}(u) \mid F_{\tau_{k-1}}\right) \leq \exp(-cu^2n^{-1}),\ \ 1 \leq k \leq n.$$

We get,

$$E_\theta \left(Z^{1/2}_{\theta,\tau_{k-1}+1,\tau_{k-1}+2}(u) \mid F_{\tau_{k-1}} \right) \le \exp(-cu^2 n^{-1})$$

because of (10). Also (assume $Z_{\theta,k,m} = 1$ if $k > m$),

$$E_\theta \left(Z^{1/2}_{\theta,\tau_{k-1}+2,\tau_k}(u) \right) (1(\tau_{k-1}+1 = \tau_k) + 1(\tau_{k-1}+1 < \tau_k))$$

$$= 1(\tau_{k-1}+1 = \tau_k) + (1(\tau_{k-1}+1 < \tau_k) E_\theta \left(Z^{1/2}_{\theta,\tau_{k-1}+2,\tau_k} \mid F_{\tau_{k-1}+1} \right) \le 1.$$

So,

$$E \left(Z^{1/2}_{\theta,\tau_{k-1}+1,\tau_k} \mid F_{\tau_{k-1}} \right) \le \exp(-cu^2 n^{-1}).$$

By induction we obtain

$$E \left(Z^{1/2}_{\theta,0,\tau_{[\kappa n]}} \mid F_{\tau_{k-1}} \right) \le \exp(-cu^2 [\kappa n]/n) \le \exp(-cu^2), \quad n > 1$$

(with another c). This proves the theorem.

6 MLE efficiency

The MLE $\hat\theta_n$ is defined by the formula

$$L_{\theta,n} \to \max_{[a,b]}.$$

Notice that if there is more than one point in the set $\arg\max_{[a,b]}$ then still it is possible to choose $\hat\theta_n$ as a random value, due to the measurable choice theorem. The statements below concern any such a choice.

Theorem 2 (MLE asymptotic efficiency) *Under conditions of theorem 1 we have,*

(1) *the MLE is consistent a.s. uniformly in $\theta \in K$;*

(2) *the MLE is asymptotically normal:*

$$\sqrt{n}(\hat\theta_n - \theta) \stackrel{P_\theta}{\Longrightarrow} N(0, \sigma^2_\theta);$$

(3) *all moments of $n^{1/2}(\hat\theta_n - \theta)$ tend to the ones of $N(0, \sigma^2_\theta)$;*

(4) *the MLE asymptotically efficient in the Hajek - Le Cam sense, i.e.*

$$\lim_{n\to\infty} [\inf_{\theta_n} \sup_{u \in U} E_u w(\sqrt{n}(\theta_n - \theta)) - \sup_u E_u w(\sqrt{n}(\hat\theta_n - \theta))] = 0.$$

for any loss function from the class W_p (see Ibragimov, Khas'minskij (1981), section 1.2).

Remark 5 (MLE consistency) *$X \in S$ and (Δ). Then*

$$\hat{\theta}_n \to \theta \quad P_\theta - a.s.$$

This is a statement from the previous theorem. However, its proof may be derived also by standard scheme (cf. Ibragimov, Khas'minskij (1981), theorem 1.4.3 and remark 1.4.1) with the series of polynomially decreasing members instead of exponents, due to theorem 1.
Comment. Under the assumptions of the theorem 2 we have P_θ-a.s. for large n

$$n^{-1} L'_{\theta,n} = 0 \tag{11}$$

(see Ibragimov, Khas'minskij (1981)). Indeed, the MLE is consistent a.s.

Remark 6 (MLE asymptotic efficiency "in Cramèr-Rao sense") *Under assumptions of theorem 2, $\hat{\theta}_n \in \tilde{K}_0$ and the asymptotic covariance of the MLE is equivalent to $n^{-1}\sigma_\theta^2$.*

Indeed, the standard Dugues scheme works well (cf. [2]).

Proof of theorem 2. All assertions follow from theorems 3.1.1. and 3.1.3 from Ibragimov, Khas'minskij (1981). To show this, we should check basic assumptions of those theorems which consist of four conditions, (H1) - (H4). We remind them for the reader's convenience, in a slightly simplified form adjusted to our case. Namely, we omit (H2) which is trivial for our normalizing coefficient $n^{1/2}$ which does not depend on θ.

(H1) For any compact $K \subset \Theta$, the experiment satisfies the uniform asymptotic normality property.

(H3) For any compact $K \subset \Theta$ there exist such $\beta > 0, m > 0, B > 0, a > 0$ that

$$\sup_{\theta \in K} \sup_{u,v \in (\Theta - \theta)\sqrt{n}} |u - v|^{-\beta} E_\theta |Z_{\theta,n}^{1/m}(u) - Z_{\theta,n}^{1/m}(v)|^m < B(1 + R^a).$$

(H4) For any compact $K \subset \Theta$ and any $N > 0$ there exists such n_0 that

$$\sup_{\theta \in K} \sup_{n > n_0} \sup_{u \in (\Theta - \theta)\sqrt{n}} |u|^N E_\theta Z_{\theta,n}^{1/2}(u) < \infty.$$

Now, proposition 2 gives us (H1). Condition (H4) it follows from theorem 1. Indeed, $|u| \leq C\sqrt{n}$ with some $C > 0$. Hence, we get from the assertion of this theorem that

$$E_\theta Z_{\theta,n}^{1/2}(u) \leq \exp(-cu^2) + C'_k (1 + u^2)^{-n}, \quad \forall k = 1, 2, \ldots$$

So it remains to check (H3). For this aim we will use lemma 3.3.1 from Ibragimov, Khas'minskij (1981) which says that a sufficient condition for (H3) with $\beta = m = 2$ is (in one-dimensional case)

$$\sup_{\theta \in K} \sup_{|u| < R,\, \theta + u \in \Theta} I_n(\theta + u) I_n^{-1}(\theta) \leq B(1 + R^a).$$

We get due to inequality (3) and condition (5) that

$$\lim_{n\to\infty} I_n(\theta+u)/I_n(\theta) \leq C$$

($I_n(\theta) := E_\theta(L'_{\theta,n})^2$). Hence, assumption of lemma 3.1.1 from Ibragimov, Khas'minskij (1981) is satisfied which gives one (H3) with $\beta = m = 2$. Theorem 2 now follows from theorems 3.1.1 and 3.1.3 from Ibragimov, Khas'minskij (1981).

Acknowledgements.

The author is grateful to V.E.Bening, D.M.Chibisov, A.A.Gushchin, Yu.Kutoyants for very useful discussions.

References

[1] A.A.Borovkov (1988) Mathematical statistics. Parameter Estimation. Supplementary chapters. Moscow, Mir.

[2] D.Duguè (1958) Statistique théorique et appliquée, Paris, Masson et C-ie.

[3] R.Hoepfner, J.Jacod, L.Ladelli (1990) Local asymptotic normality and mixed normality for Markov statistical models, PTRF, 86, 105-129.

[4] I.A.Ibragimov, R.Z.Khas'minskij (1981) Statistical estimation, asymptotic theory. Berlin et al., Springer.

[5] I.A.Ibragimov, Yu.V.Linnik (1971) Independent and stationary sequences of random variables. Groningen, The Netherlands, Wolters-Noordhoff.

[6] Yu.A.Kutoyants (1984) Parameter estimation for stochastic processes. Berlin, Heldermenn.

[7] Yu.Kutoyants (1994) Identification of dynamical systems with small noise, Dordrecht, Kluwer.

[8] L.LeCam (1986) Asymptotic methods in statistical decision theory. Berlin et al., Springer.

[9] S.P.Meyn, R.Tweedie (1993) Markov chains and stochastic stability, Springer, Berlin et al.

[10] X.Milhaud, G.Oppenheim, M.C.Viano (1983) Sur la convergence du processus de vraisemblance en variables markoviennes, Z.Wahrscheinlichkeitstheorie verw. Gebiete, 64, 49-65.

[11] Y.Ogata, N.Inagaki (1977) The weak-convergence of the likelihood ratio random fields for Markov observations, Ann. Inst. Statist. Math. 29 A, 165-187.

[12] G.Roussas (1972) Contiguity of probability measures, London, Cambridge Univ. Press.

[13] A.Yu.Veretennikov (1997) On polynomial mixing bounds for stochastic differential equations, Stoch. Processes Appl. 70, 115-127.

[14] A.Yu.Veretennikov (1998) On polynomial mixing and convergence rate for stochastic difference and differential equations, preprint No. 393, WIAS Berlin.