Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R.J. Chitashvili, N.L. Lazrieva, T.A. Toronjadze

Asymptotic theory of M-estimators in general statistical models
Part II: On asymptotic behaviour of estimators in the presence of nuisance parameters

Department of Operations Research, Statistics, and System Theory Report BS-R9020 June

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Asymptotic Theory of M-Estimators in General Statistical Models

Part II: On asymptotic behaviour of estimators in the presence of nuisance parameters

R.J. Chitashvili, N.L. Lazrieva, T.A. Toronjadze
Department of Probability theory and Mathematical statistics
Tbilisi Mathematical Institute, Academy of Sciences of the Georgian SSR
150 a, Agmashenebeli Avenue, Tbilisi 380012, USSR

The asymptotic behaviour of the solutions to the equation for M-estimators is considered under model disturbance in a general scheme of statistical models.

1980 Mathematics Subject Classification: 60G40.
Keywords & Phrases: semimartingales, statistical estimation.
Note: To appear in Transactions of Tbilisi Mathematical Institute, 1989. This work was completed while the first author was visiting the Centre for Mathematics and Computer Science, Amsterdam.

1. Introduction

In this part (see CWI Report BS-9019 for Part I) the methods proposed in [8] by the present authors are further developed. The range of statistical problems covered by these methods is extended. In particular, the following questions are studied: regularity of M-estimators, one-step approximation technique, stability of estimators with respect to model variations. Problems of parameter estimation in presence of a finite-dimensional nuisance parameters are treated in detail, in particular, the asymptotic properties of so called pseudo M-estimators, the method of skew projection, the characterization of the limiting distributions of regular estimators. Besides, a scheme is proposed in which structural parameter estimators are constructed by the partial likelihood technique and the asymptotic variances are compared for estimators obtained by projection technique, partial likelihood technique and MLE (under known nuisance parameters).

Unlike [8] this paper deals with the multidimensional parameter case. To make the discription complete, detailed proofs of results are given - brief schemes of these proofs are presented in [8]. Several parts of this work were published earlier in short communications of N.L. Lazrieva and T.A. Toronjadze. Partly it was completed while the first author was visiting the Centre for Mathematics and Computer Science, and was discussed at the local seminar on martingale methods in statistics, organized by K. Dzhaparidze. The authors are thankful to all participants of these useful discussions.

2. Specification of the model

2.1. Let

$$\mathcal{E} = (\Omega, \mathcal{F}, F, P_\theta, P), \quad \theta \in \Theta \subset R^d, \quad d \geq 1,$$

be a general statistical experiment with a filtration, where $(\Omega, \mathcal{F}, F, P)$ is a probability space with filtration $F = (\mathcal{F}_t)$, $t \geq 0$, satisfying usual conditions, P_θ a probability measure depending on the parameter $\theta \in \Theta$, Θ an open subset of the Euclidean space R^d. It is assumed for all $\theta \in \Theta$ that $P_\theta \overset{\text{loc}}{\sim} P$.

Let $P(t) = P|\mathcal{F}_t$, $P_\theta(t) = P_\theta|\mathcal{F}_t$ be the restrictions of the measures P and P_θ to the σ-algebra $\mathcal{F}_t$, and let $\rho_\theta = (\rho_\theta(t))$, $t \geq 0$ be the likelihood ratio process (we consider a right-continuous modification with left-hand limits):

$$\rho_\theta(t) = \frac{dP_\theta(t)}{dP(t)}.$$

Report BS-R9020
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

For simplicity we assume $\rho_\theta(0)=1$.

It is well known (see, e.g. [2]) that there exists a local P-martingale $M_\theta=(M_\theta(t))_{t\geq 0}$ such that the process ρ_θ can be expressed as Dolean's exponential

$$\rho_\theta = \mathscr{E}(M_\theta):=\exp\{M_\theta - \tfrac{1}{2}\langle M_\theta^c\rangle\}\prod(1+\Delta M_\theta)e^{-\Delta M_\theta},$$

where M^c is a continuous part, $\langle M\rangle$ the square characteristic and ΔM the jump of a martingale M.

2.2. Let Q be some other probability measure on $(\Omega,\mathscr{F},F)$ such that $Q\overset{\text{loc}}{\ll}P$ and $dQ/dP=\mathscr{E}(M)$, where M is a P-martingale. If m is some local P-martingale, then by the general Girsanov theorem [2] the process L (L-transformation)

$$L(m,M):=m-\langle m^c,M^c\rangle-\sum\frac{\Delta m\Delta M}{1+\Delta M}$$

is a local Q-martingale.

2.3. An experiment $\mathscr{E}$ is called regular if the following conditions are satisfied:

(1) For each $t\geq 0$ $P-a.s.$ the function $\theta\to M_\theta(t,\omega)$ is continuously differentiable and the derivative $\dot{M}_\theta:=(\frac{\partial}{\partial\theta_1}M_\theta,\ldots,\frac{\partial}{\partial\theta_d}M_\theta)'$ (here and elsewhere below "prime" is the transposition sign) for all $\theta\in\Theta$ is a local P-martingale.

(2) For all $t\geq 0$ $P-a.s$:

a) $\frac{\partial}{\partial\theta}\ln\rho_\theta:=(\frac{\partial}{\partial\theta_1}\ln\rho_\theta,\ldots,\frac{\partial}{\partial\theta_d}\ln\rho_\theta)'$ exists;

b) $\frac{\partial}{\partial\theta}\ln\rho_\theta=L(\dot{M}_\theta,M_\theta)$, where $L(\dot{M}_\theta,M_\theta)=(L(\frac{\partial}{\partial\theta}M_\theta,M_\theta),\ldots,L(\frac{\partial}{\partial\theta_d}M_\theta,M_\theta))'$;

c) $L(\dot{M}_\theta,M_\theta)\in M^2(P_\theta)$ ($M^2(P_\theta)$ is the class of square integrable P_θ-martingales).

Denote by $\hat{I}(\theta)=(\hat{I}_t(\theta))_{t\geq 0}$ the Fisher information process i.e.

$$\hat{I}(\theta)=\langle L(\dot{M}_\theta,M_\theta)\rangle:=(\langle L(\frac{\partial}{\partial\theta_i}M_\theta,M_\theta),L(\frac{\partial}{\partial\theta_j}M_\theta,M_\theta)\rangle)_{i,j=\overline{1,d}}$$

(3) the Fisher information matrix

$$I(\theta):=E_\theta\hat{I}(\theta)$$

is finite and positively defined.

The likelihood equation takes here the form

$$L(\dot{M}_\theta,M_\theta)=0.$$

Of course, this is the special member of the following family of equations

$$L(m_\theta,M_\theta)=0$$

with certain d-dimentional P-martingales m_θ depending on $\theta\in\Theta$. These are estimational equations in the sense that their solutions are viewed as estimators of the unknown parameter θ, the so called M-estimators [8]. To preserve the classical terminology we shall say that the martingale m_θ defines the M-estimator and P_θ-martingale $L(m_\theta,M_\theta)$ is the influence martingale.

2.4. It will be convenient to consider a scheme of series i.e. a sequence of the regular statistical experiments (models)

$$\mathbf{E}=(\mathscr{E}_n)_{n\geq 1}=(\Omega^n,\mathscr{F}^n F^n=(\mathscr{F}_t^n)_{0\leq t\leq T},P_\theta^n,P^n),\ \theta\in\Theta\subset R^d,\ n\geq 1,$$

$T>0$ is a number.

A sequence $\mathbf{E}=(\mathcal{E}_n), n\geq 1$ will be called ergodic if the following conditions are satisfied:

(i) there exists a numerical sequence $c_n, n\geq 1$ (a sequence of numbers) such that $c_n>0$, $\lim_{n\to\infty} c_n=0$, and for each $\theta\in\Theta$

$$c_n^2 I_T^n(\theta) \underset{n\to\infty}{\to} I(\theta),$$

$$c_n^2 \hat{I}_T^n(\theta) \xrightarrow{P_\theta^n} I(\theta),$$

where $I(\theta)$ is finite positive definite matrix for all $\theta\in\Theta$, sign "$\xrightarrow{P_\theta^n}$" denotes the convergence in probability P_θ^n (i.e. $\xi_n \xrightarrow{P_\theta^n} \xi$ means that $P_\theta^n\{|\xi_n-\xi|>\rho\}\to 0$, $n\to\infty$, $\forall\rho>0$); or

(j) denote $c_n(\theta)=(I_T^n(\theta))^{-\frac{1}{2}}$. Then

$$\lim_{n\to\infty} \|c_n(\theta)\|=0,$$

$$c_n^2(\theta)\,\hat{I}_T^n(\theta) \xrightarrow{P_\theta^n} I,$$

where $\|\cdot\|$ is the usual matrix norm and I the unit matrix.

For convenience we will use below both of these definitions.

2.5. Example 1. *I.I.D.*

Let μ and μ_θ, $\theta\in\Theta\subset R^1$, be probability measures, defined on some measurable space $(X,\mathcal{B})$; let $\mu_\theta\sim\mu$ and $d\mu_\theta/d\mu\,(x)=f(x,\theta)$. Put T=1.

Corresponding scheme of experiments has the form $\mathbf{E}=(\mathcal{E}_n)_{n\geq 1}=(\Omega^n,\mathcal{F}^n,F^n=(F_t^n), 0\leq t\leq 1, P_\theta^n, P^n)$, $n\geq 1$, where $\Omega^n=X^n$, $\mathcal{F}^n=\mathcal{B}^n$, $\mathcal{F}_t^n=\mathcal{B}^{[nt]}$ ([·]-denotes an integer part), $P_\theta^n=\mu_\theta\times\cdots\times\mu_\theta$, $P^n=\mu\times\cdots\times\mu$. It can be easily seen that for $\omega=(x_1,\ldots,x_n)\in X^n$

$$\rho_\theta^n(t,\omega)=\prod_{i=1}^{[nt]} f(x_i,\theta)=\mathcal{E}_t(M_\theta^n),$$

where

$$M_\theta^n(t)=\sum_{i=1}^{[nt]}(f(x_i,\theta)-1).$$

Now, if for μ- almost all x the function $f(x,\theta)$ is differentiable w.r.t. θ and

$$\frac{\partial}{\partial\theta}\int f(x,\theta)\,d\mu(x)=\int \dot{f}(x,\theta)d\mu(x)=0,$$

then

$$\frac{\partial}{\partial\theta}\ln\rho_\theta^n(t,\omega)=\sum_{i=1}^{[nt]}\frac{\dot{f}(x_i,\theta)}{f(x_i,\theta)}=\sum_{s\leq t}\frac{\Delta\dot{M}_s^n}{1+\Delta M_s^n}=L_t(\dot{M}_\theta^n,M_\theta^n),$$

and if $0<I(\theta):=\int(\frac{\dot{f}(x,\theta)}{f(x,\theta)})^2 f(x,\theta)d\mu(x)<\infty$, then the regularity conditions are satisfied. Note that $I_1^n(\theta)=nI(\theta)$. By the law of large numbers it follows that the ergodicity conditions are satisfied with $c_n^2=n$ (or $c_n^2(\theta)=nI(\theta)$)

The MLE equation has the form

$$L_1(\dot{M}_\theta^n,M_\theta^n)=\sum_{i=1}^{n}\frac{\dot{f}(x_i,\theta)}{f(x_i,\theta)}=0.$$

Now, if $m_\theta^n(t)=\sum_{i=1}^{[nt]}\phi(x_i,\theta)$ where $\int\phi(x,\theta)\mu(dx)=0\ \forall\theta$, then m_θ^n will be a P^n-martingale and the equation for M-estimator takes the form

$$L_1(m_\theta^n,M_\theta^n)=\sum_{i=1}^{n}\frac{\phi(x_i,\theta)}{f(x_i,\theta)}=0.$$

Example 2. *Diffusion.*
For every $n\geq 1$ let ξ_n be a diffusion type process with the differential

$$d\xi_n(t)= a_n(t,\xi_n,\theta)\,dt + dw_n(t),\ 0\leq t\leq T,\ \xi_n(0)=0, \tag{2.1}$$

where $\theta\in\Theta\subset R^1$ is an unknown parameter.

This case is covered by a general scheme of experiments $\mathbf{E}=(\mathcal{E}_n)_{n\geq 1}$ in the following manner:

$$\Omega^n = C_{[0,T]},\ \mathcal{F}^n = \mathcal{B}_T=\sigma\{x:x_s,s\leq T\}.$$

$$F^n=\{\mathcal{F}^n_t = \sigma(x:x_s,s\leq t),\ 0\leq t\leq T\},$$

P^n_θ is a distribution of the process ξ_n (with given θ) and P^n the Wiener measure. For all $n\geq 1$ assume

$$P^n\{\int_0^T a_n^2(s,x,\theta)\,ds<\infty\}=P^n_\theta\{\int_0^T a_n^2(s,x,\theta)ds<\infty\}=1$$

Under these conditions there exists the unique weak solution to the equation (2.1), $P^n_\theta\sim P$ and the local density has the form

$$\rho^n_\theta(t)=\exp(\int_0^t a_n(s,x,\theta)dx_s - \tfrac{1}{2}\int_0^t a_n^2(s,x,\theta)ds)=\mathcal{E}_t(M^n_\theta)\ P^n-a.s.$$

where $M^n_\theta(t)=\int_0^t a_n(s,x,\theta)dx_s$ is a local P^n-martingale.

Further assume that for all $n\geq 1$, $x\in C_{[0,T]}$ and $t\in[0,T]$ the function $\theta\to a_n(t,x,\theta)$ is continuously differentiable and

$$\frac{\partial}{\partial\theta}\int_0^t a_n(s,x,\theta)dx_s=\int_0^t\frac{\partial}{\partial\theta}a_n(s,x,\theta)dx_s,\ P^n-a.s.$$

$$\frac{\partial}{\partial\theta}\int_0^t a_n^2(s,x,\theta)ds=2\int_0^t(\frac{\partial}{\partial\theta}a_n(s,x,\theta))\,a_n(s,x,\theta)ds,\ P^n-a.s.$$

$$0<I^n_T(\theta):=E^n_\theta\int_0^T(\frac{\partial}{\partial\theta}a_n(s,x,\theta))^2ds<\infty.$$

Then the regularity conditions are satisfied and

$$\frac{\partial}{\partial\theta}\ln\rho^n_\theta(t)=\int_0^t\dot a_n(s,x,\theta)(dx_s - a_n(s,x,\theta)ds)$$

(here $\dot a_n(s,x,\theta):=\frac{\partial}{\partial\theta}a_n(s,x,\theta)$), so the MLE equation is

$$L_T(\dot M^n_\theta,M^n_\theta)=\int_0^T\dot a_n(s,x,\theta)(dx_s-a_n(s,x,\theta)ds)=0,$$

and the natural generalization of this is the equation for M-estimator

$$L_T(m^n_\theta,M^n_\theta)=\int_0^T\phi_n(s,x,\theta)(dx_s-a_n(s,x,\theta)ds)=0$$

where $\phi_n(\cdot,\cdot,\theta)$ is some nonanticipative functional with $E^n_\theta\int_0^T\phi_n^2(s,x,\theta)ds<\infty$ and $m^n_\theta(t)=\int_0^t\phi_n(s,x,\theta)dx_s$.

The Fisher information process is $\hat I^n_t(\theta)=\int_0^t(\dot a_n(s,x,\theta))^2ds$ and the ergodicity conditions turn into:

(i) there exists a sequence $c_n, n\geq 1$, $c_n>0$, $c_n\to 0$ such that

$$c_n^2 I^n_T(\theta)\to I(\theta),\ c_n^2\int_0^T(\dot a_n(s,x,\theta))^2ds\xrightarrow{P^n_\theta}I(\theta),$$

or

(j) $$c_n^2(\theta)=(I^n_T(\theta))^{-1}=(E^n_\theta\int_0^T(\dot a_n(s,x,\theta))^2ds)^{-1}$$

and

$$\int_0^T (\dot{a}_n(s,x,\theta))^2 ds \ (E_\theta^n \int_0^T (\dot{a}_n(s,x,\theta))^2 ds)^{-1} \xrightarrow{P_\theta^n} 1.$$

Now we give an example of diffusion, when the ergodicity condition is automatically fulfilled.

The ergodic diffusion process. Let $\xi=(\xi_t)$, $t\geqslant 0$, be the diffusion process with

$$d\xi_t = a(\xi_t,\theta)dt + dw_t,\ \xi_0 = 0.$$

The equation for MLE is

$$\int_0^t \dot{a}(x_s,\theta)(dx_s - a(x_s,\theta)ds) = 0, \tag{2.2}$$

while the equation for M-estimator is

$$\int_0^t \phi(x_s,\theta)(dx_s - a(x_s,\theta)ds) = 0. \tag{2.3}$$

We assume that $\hat{\theta}_t$ is constructed from (2.2) or (2.3), and study its asymptotic properties.

This problem is equivalent to studying asymptotic properties of the estimator $\hat{\theta}_n$ as $n\to\infty$ in the following scheme of series:

$$d\xi_t^n = a_n(\xi_t^n,\theta)dt + dw_t^n,\ \xi_0^n = 0,\ 0\leqslant t\leqslant T$$

where $a^n(x,\theta) = \sqrt{n}\ a(\sqrt{n}x,\theta)$, $\xi_t^n = \frac{1}{\sqrt{n}}\xi_{nt}$, $w_t^n = \sqrt{n}w_{nt}$, and $\hat{\theta}_n$ is constructed from the equation

$$\int_0^T \phi_n(\xi_t^n,\theta)(d\xi_t^n - a_n(\xi_t^n,\theta)dt) = 0 \quad (\text{or } \int_0^T \dot{a}_n(\xi_t^n,\theta)(d\xi_t^n - a_n(\xi_t^n,\theta)dt) = 0),$$

where

$$\phi_n(x,\theta) = \sqrt{n}\phi(\sqrt{n}x,\theta).$$

Suppose that the process ξ (for each θ) is ergodic, i.e. $G(\theta)<\infty$ where

$$G(\theta) = \int_{-\infty}^{+\infty} e^{2\int_0^x a(y,\theta)dy} dx.$$

Then, by well known theorem of Maruyama and Tanaka it follows that :

1) ξ has the limit distribution $\mu_\theta(x) = G^{-1}(\theta) \int_{-\infty}^{x} e^{\int_0^y a(u,\theta)du} dy$ and

2) for any measurable μ_θ-integrable function $\psi(x)$ the following relations hold

$$P_\theta\{\lim_{t\to\infty} \frac{1}{t}\int_0^t \psi(\xi_s)ds = \int_{-\infty}^{+\infty} \psi(x)\mu_\theta(dx)\} = 1,$$

$$\lim_{t\to\infty} \frac{1}{t} E_\theta \int_0^t \psi(\xi_s)ds = \int_{-\infty}^{+\infty} \psi(x)\mu_\theta(dx). \tag{2.4}$$

In the scheme of series

$$I_T^n(\theta) = E_\theta^n \int_0^T (\dot{a}_n(x_s,\theta))^2 ds = E_\theta^n \int_0^{nT} (\dot{a}(x_s,\theta))^2 ds, \hat{I}_T^n(\theta) = \int_0^{nT} (\dot{a}(x_s,\theta))^2 ds$$

and by virtue of (2.4) the ergodicity conditions are satisfied with

$$c_n = \frac{1}{\sqrt{n}} \quad \text{or} \quad c_n(\theta) = (I_T^n(\theta))^{-1/2}.$$

EXAMPLE 3. *Point processes.*

Let $\mathcal{E} = (\Omega,\mathcal{F},F,P_\theta,P)$, $\theta\in\Theta\subset R^1$ where Ω is a space of piecewise continuous functions $X = (X_t)_{t\geqslant 0}$ such that $X_0 = 0$, $X_t - X_{t-} = 0$ or 1, $\mathcal{F} = \sigma\{X : X_s, s\geqslant 0\}$, $F = \{\mathcal{F}_t\}_{t\geqslant 0}$, $\mathcal{F}_t = \sigma\{X : X_s, s\leqslant t\}$, P_θ is a measure on $(\Omega,\mathcal{F})$ such that the coordinate process $X = (X_t)_{t\geqslant 0}$ is a point process with a compensator

$$A(t,\theta)=\int_0^t a(s,\theta)d\alpha_s$$

and P is a measure such that X is a point process with compensator $\alpha=(\alpha_t)_{t\geq 0}$. Assume that α is continuous and $\int_0^1(1-\sqrt{a(s,\alpha)})^2 d\alpha_s<\infty$. In this case $P_\theta \overset{\text{loc}}{\sim} P$ and

$$\rho_\theta(t)=\mathcal{E}_t(M_\theta)$$

where $M_\theta(t)=\int_0^t(a(s,\theta)-1)\,d(X_s-\alpha_s)$. Assuming that $a(s,\theta)$ is differentiable, we can write

$$\dot{M}_\theta(t)=\int_0^t \dot{a}(s,\theta)\,d(X_s-\alpha_s)$$

and therefore

$$\frac{\partial}{\partial\theta}\ln\rho_\theta(t)=\int_0^t\frac{\dot{a}(s,\theta)}{a(s,\theta)}(dX_s-a(s,\theta)d\alpha_s)=L_t(\dot{M}_\theta,M_\theta),$$

Hence the MLE equation has the form

$$\int_0^t\frac{\dot{a}(s,\theta)}{a(s,\theta)}(dX_s-a(s,\theta)d\alpha_s)=0$$

while the equation for M-estimator is

$$\int_0^t\frac{\psi(s,\theta)}{a(s,\theta)}(dX_s-a(s,\theta)d\alpha_s)=0.$$

The Fisher information $I_t(\theta)$ in this case has the form

$$I_t(\theta)=E_\theta\int_0^t\left[\frac{\dot{a}(s,\theta)}{a(s,\theta)}\right]^2 a(s,\theta)d\alpha_s.$$

Denote $c_t^{-2}(\Theta)=I_t(\theta)$. Ergodicity of the model means

$$c_t^2(\theta)\int_0^t\left[\frac{\dot{a}(s,\theta)}{a(s,\theta)}\right]^2 a(s,\theta)d\alpha_s\xrightarrow{P_\theta}1$$

(or: there exists a function c_t such that $c_t>0$, $c_t\to 0$, $t\to\infty$

$$\lim_{t\to\infty}c_t^{-2}\int_0^t\left[\frac{\dot{a}(s,\theta)}{a(s,\theta)}\right]^2 a(s,\theta)d\alpha_s\xrightarrow{P_\theta}I_\theta,\ \lim_{t\to\infty}c_t^{-2}I_t(\theta)=I(\theta)).$$

3. Local asymptotic normality (LAN).

3.1. Let a sequence of statistical experiments be given:

$$\mathsf{E}=(\mathcal{E}_n)_{n\geq 1}=(\Omega^n,\mathcal{F}^n,F^n=(\mathcal{F}_t^n)_{0\leq t\leq T},P_\theta^n,P^n)_{n\geq 1},\theta\in\Theta\subset R^d.$$

Define the process of normalized likelihood ratio

$$Z_\theta^n(u):=\frac{dP^n_{\theta+c_nu}}{dP^n_\theta}$$

where $u\in U_n(\theta)=\{u:\theta+c_nu\in\Theta\}\subset R^d$.

We will need some additional conditions of regularity:

(i) for all $z,\theta\in\Theta$
$L(\dot{M}_z^n,M_\theta^n)\in M^2(P_\theta^n)$; $L(\dot{M}_z^n,M_\theta^n)$ and $I^n(z,\theta):=E_\theta^n L^2(\dot{M}_z^n,M_\theta^n)$ are z-continuous;

(ii) $\frac{\partial}{\partial z}L(M_z^n,M_\theta^n)=L(\dot{M}_z^n,M_\theta^n)$;

(iii) for all $z,\theta\in\Theta$

$$L\left(\int_0^1 (\dot{M}^n_{\theta+sz},z)ds, M^n_\theta\right) = \int_0^1 L((\dot{M}^n_{\theta+sz},z),M^n_\theta)ds$$

THEOREM 3.1. *Let the following conditions hold:*

(a) *The sequence of experiments* $\mathbb{E}=(\mathcal{E}_n)_{n\geq 1}$ *is such that the conditions (i), (iii) (or (i), (ii)) are satisfied.*

(b) $c_n^2|X|^2 I_{\{|X|>c_n\epsilon\}} * \nu_T^n \xrightarrow{P^n_\theta} 0$ *for* $\forall\epsilon>0$, *where* ν^n *is the compensator of the jump measure of the process* $L(\dot{M}^n_\theta, M^n_\theta)$ *w.r.t. the measure* P^n_θ;

(c) $\lim_{r\to 0}\overline{\lim_{n\to\infty}}\sup_{y:\|y-\theta\|<r} c_n^2\|E^n_\theta\langle L(\dot{M}^n_y-\dot{M}^n_\theta,M^n_\theta)\rangle\|=0$

(or for every $\rho>0$

(c¹) $\lim_{r\to 0}\overline{\lim_{n\to\infty}} P^n_\theta\{\sup_{y:\|y-\theta\|<r} c_n^2\|\langle L(\dot{M}^n_y-\dot{M}^n_\theta,M^n_\theta)\rangle_T\|>\rho\}=0)$.

Then the family of measures $\{P^n_\theta,\theta\in\Theta\}_{n\geq 1}$ *possesses the LAN property at each point* $\theta\in\Theta$, *i.e.*

$$\ln Z^n_\theta(u) = (\Delta^n_\theta,u) - \frac{1}{2}(I(\theta)u,u) + \psi^n_\theta(u), \tag{3.1}$$

where

$$\mathfrak{L}(\Delta^n_\theta|P^n_\theta)\Rightarrow N(0,I(\theta)),\ \Delta^n_\theta := c_n L(\dot{M}^n_\theta,M^n_\theta),\ \psi^n_\theta \xrightarrow{P^n_\theta} 0$$

and $\mathfrak{L}(\xi|P)$ *is the distribution of a random variable* ξ *w.r.t. a measure P. (Here and below "⇒" denotes the weak convergence of distributions, and* $(a,b)=a'b$ *the scalar product of two vectors* $a,b\in R^d$*)*

PROOF. Applying the formula for the product of two Dolean exponentials we easily arrive at

$$Z^n_\theta(u) = \mathcal{E}(L(M^n_{\theta+c_n u}-M^n_\theta,M^n_\theta)) := \mathcal{E}(L^n),$$

which implies

$$\ln Z^n_\theta(u) = L^n - \tfrac{1}{2}\langle L^{n,c}\rangle + \sum(\ln(1+\Delta L^n)-\Delta L^n).$$

Denoting

$$J^n_\theta(u) = L(M^n_{\theta+c_n u}-M^n_\theta,M^n_\theta) - c_n(L(\dot{M}^n_\theta,M^n_\theta),u),$$

$$\psi^n_\theta(u) = J^n_\theta(u) - \tfrac{1}{2}\langle L^{n,c}\rangle + \sum(\ln(1+\Delta L^n) - \tfrac{1}{2}(I(\theta)u,u),$$

we will show that the representation (3.1) holds with Δ^n_θ and $\psi^n_\theta(u)$ introduced above.

The functional central limit theorem for square integrable martingales [3] with conditions (a) and (b) ensure the convergence

$$\mathfrak{L}(\Delta^n_\theta|P^n_\theta)\Rightarrow N(0,I(\theta)),\quad n\to\infty.$$

We shall prove below that for all $n\in R^d$

$$\psi^n_\theta(u)\xrightarrow{P^n_\theta}0,\ n\to\infty.$$

It will be proved first that for any $u\in R^d$

$$J^n_\theta(u)\xrightarrow{P^n_\theta}0,\ n\to\infty. \tag{3.2}$$

Using the generalized Newton-Leibnitz formula we obtain

$$M^n_{\theta+c_n u}-M^n_\theta = \int_0^1 (\dot{M}^n_{\theta+sc_n u},c_n u)ds.$$

Consequently, in virtue of (i) and (iii) we have

$$J^n_\theta(u) = L(M^n_{\theta+c_n u}-M^n_\theta - (\dot{M}^n_\theta,c_n u),M^n_\theta) = \int_0^1 (L(\dot{M}^n_{\theta+sc_n u}-\dot{M}^n_\theta,M^n_\theta),c_n u)ds.$$

This gives

$$E_\theta^n(J_\theta^n(u))^2 \leq \int_0^1 (E_\theta^n \langle L(\dot{M}_{\theta+sc_nu}^n - \dot{M}_\theta^n, M_\theta^n\rangle_T c_n u, c_n u) ds$$
$$\leq c_n^2 \sup_{y:\|y-\theta\| \leq c_n u} \|E_\theta^n \langle L(\dot{M}_y - \dot{M}_\theta, M_\theta)\rangle_T\| \cdot \|u\|^2 \to 0.$$

It remains to prove that

$$\psi_\theta^n \xrightarrow{P_\theta^n} 0.$$

Note first that

$$\sup_{s \leq T} |\Delta_s L(M_{\theta+sc_nu}^n - M_\theta^n, M_\theta^n)| \xrightarrow{P_\theta^n} 0. \tag{3.3}$$

Indeed, from the definition of $J_\theta^n(u)$ we have

$$|\Delta_s L(M_{\theta+sc_nu}^n - M_\theta^n, M_\theta^n)| \leq |\Delta_s J_\theta^n(u)| + |\Delta_s(L(\dot{M}_\theta^n, M_\theta^n), c_n u)|$$

so that the relation (3.3) follows from the relation

$$E_\theta^n \sup_{s \leq T} |\Delta_s J_\theta^n(u)|^2 \leq E_\theta^n [J_\theta^n(u)]_T \to 0$$

and Lindeberg's condition.

Further, by (3.3) and Taylor's formula we have

$$-\frac{1}{2} < L^{n,c} > + \sum(\ln(1+\Delta L^n) - \Delta L^n) = -\frac{1}{2}\langle L^{n,c}\rangle - \frac{1}{2}\sum(\Delta L^n)^2 + \sum o((\Delta L^n)^2).$$

Therefore, by virtue of (3.2) again we obtain (here and elsewhere below $P_\theta^n - \lim_{n\to\infty} \xi_n$ means $\{\xi_n \xrightarrow{P_\theta^n}\}$)

$$P_\theta^n - \lim_{n\to\infty} \psi_\theta^n(u) = P_\theta^n - \lim_{n\to\infty} (-\frac{1}{2}\langle L^{n,c}\rangle - \frac{1}{2}\sum(\Delta L^n)^2 + \frac{1}{2}(I(\theta)u,u))$$
$$= P_\theta^n - \lim_{n\to\infty} \{(I(\theta)u,u) - [L^n]\}.$$

We have

$$[L^n] = [L(M_{\theta+c_nu}^n - M_\theta^n, M_\theta^n)] = [L(M_{\theta+c_nu}^n - M_\theta^n - (\dot{M}_\theta^n, c_n u), M_\theta^n) + L((\dot{M}_\theta^n, c_n u), M_\theta^n)]$$
$$= [L(M_{\theta+c_nu}^n - M_\theta^n - (\dot{M}_\theta^n, c_n u), M_\theta^n] + 2[L(M_{\theta+c_nu}^n - M_\theta^n - (\dot{M}_\theta^n, c_n u), M_\theta^n), L((\dot{M}_\theta^n, c_n u), M_\theta^n)]$$
$$+ [L((\dot{M}_\theta^n, c_n u), M_\theta^n)].$$

The first two terms of the last relation tend to zero in probability P_θ^n by virtue of conditions (a) and (b) and the relation (3.2). The conditions (a) and (b) imply that the third term converges to $(I(\theta)u,u)$. □

Note that Theorem 3.1 has been asserted without proving in [9] and [10].

REMARK 3.1. Theorem 3.1 is proved similarly under conditions (i), (ii) and (c^1) by using Lenglart's inequality.

4. ASYMPTOTIC PROPERTIES OF SOLUTIONS TO AN ESTIMATIONAL EQUATION

4.1. For every $\theta \in \Theta \subset R^d$ let a sequence of probability measures $\{Q_\theta^n\}_{n \geq 1}$, ($Q_\theta^n \sim P^n$) and d-dimensional random vectors $L_n(\theta)$, $n \geq 1$, be given on a measurable space $(\Omega^n, \mathcal{F}_T^n)$ as well as a sequence of positive numbers $(c_n)_{n \geq 1}$.

LEMMA 4.1. *Let the following conditions hold:*

a) $\lim_{n\to\infty} c_n = 0.$

b) *For each $n \geqslant 1$ the mapping $\theta \rightsquigarrow L_n(\theta)$ is continuously θ-differentiable Q_θ^n a.s.*
c) *For each $\theta \in \Theta$ there exists a function $\Delta_Q(\theta,y)$, $\theta, y \in \Theta$, such that*

$$Q_\theta^n - \lim_{n\to\infty} c_n^2 L(y) = \Delta_Q(\theta,y),$$

and the equation (w.r.t. y)

$$\Delta_Q(\theta,y)=0$$

has the unique solution $\theta^* = b(\theta)$.
d) $Q_\theta^n - \lim_{n\to\infty} c_n^2 \dot{L}_n(\theta^*) = -\gamma_Q(\theta)$ *where $\gamma_Q(\theta)$ is a positive definite matrix for every $\theta \in \Theta$.*
e) $\lim_{r\to 0} \overline{\lim_{n\to\infty}} Q_\theta^n \{ \sup_{y:\|y-\theta^*\|<r} c_n^2 \|\dot{L}_n(y) - \dot{L}_n(\theta^*)\| > \rho \} = 0$ *for each $\rho > 0$.*

Then for each $\theta \in \Theta$ there exists a sequence of random vectors $\hat{\theta} = (\hat{\theta}_n)_{n\geqslant 1}$ taking on values in Θ, such that

I. $\lim_{n\to\infty} Q_\theta^n \{ L_n(\hat{\theta}_n) = 0 \} = 1;$

II. $Q_\theta^n - \lim_{n\to\infty} \hat{\theta}_n = b(\theta);$

III. *if $\tilde{\theta} = (\theta_n)_{n\geqslant 1}$ is another sequence with properties* I *and* II, *then*

$$\lim_{n\to\infty} Q_\theta^n \{ \hat{\theta}_n = \tilde{\theta}_n \} = 1;$$

IV. *if the sequence of distributions $\mathfrak{L}(c_n L_n(\theta^*) | Q_\theta^n)$, $n \geqslant 1$, converges to a certain distribution Φ, then*

$$\mathfrak{L}(\gamma_Q(\theta) c_n^{-1} (\hat{\theta}_n - \theta^*) | Q_\theta^n) \Rightarrow \Phi.$$

PROOF 1°. We shall show first that the decomposition

$$c_n^2 L_n(y) = c_n^2 L_n(\theta^*) - \gamma_Q(\theta)(y-\theta^*) + \epsilon_n(y,\theta^*)(y-\theta^*) \tag{4.1}$$

takes place with $\epsilon_n(y,\theta^*) \in \mathscr{F}_T^n$ and

$$\lim_{r\to 0} \overline{\lim_{n\to\infty}} Q_\theta^n \{ \sup_{y:\|y-\theta^*\|<r} \|\epsilon_n(y,\theta^*)\| > \rho \} = 0 \tag{4.2}$$

for each $\rho > 0$.

By Taylor's formula we have

$$c_n^2 L_n(y) = c_n^2 L_n(\theta^*) + c_n^2 \dot{L}_n(\theta^*)(y-\theta^*) + c_n^2 [\dot{L}_n(\bar{\theta}) - \dot{L}_n(\theta^*)](y-\theta^*)$$

where $\bar{\theta}(\theta) = \theta^* + \alpha(\theta^*)(y-\theta^*)$, $\alpha(\theta^*) \in [0,1]$ and the point $\bar{\theta}$ is chosen so that $\bar{\theta} \in \mathscr{F}_T^n$.

Further, in view of condition d) we have

$$c_n^2 L_n(y) = c_n^2 L_n(\theta^*) - \gamma_Q(\theta)(y-\theta^*) + \epsilon_n(\bar{\theta}(\theta),\theta^*)(y-\theta^*)$$

where

$$\epsilon_n(y,\theta^*) = c_n^2 [\dot{L}_n(y) - \dot{L}_n(\theta^*)] + [c_n^2 \dot{L}_n(\theta^*) + \gamma_Q(\theta)], y \in \Theta.$$

Evidently, conditions d) and e) ensure the property (4.2).

2°. We shall show now that there exists a family $\{\Omega_\theta(n,r): n \geqslant 1, r > 0, \theta \in \Theta\}$ such that

$$1)\ \Omega_\theta(n,r) \in \mathscr{F}_T^n,$$

$$2)\ \lim_{r\to 0} \underline{\lim_{n\to\infty}} Q_\theta^n \{ \Omega_\theta(n,r) \} = 1,$$

and for any $r > 0$, $n \geqslant 1$ and $\omega \in \Omega_\theta(n,r)$ the equation

$$L_n(y) = 0$$

has the unique solution $\hat{\theta}_n(\theta,\omega)$ in the ball $\|y-\theta^*\|\leq r$.

The decomposition (4.1) gives

$$(c_n^2 L_n(\theta^*+u),u)=(c_n^2 L_n(\theta^*),u)-u'\gamma_Q(\theta)u+u'\epsilon_n(\theta^*+u,\theta^*)u. \tag{4.3}$$

For any $\theta\in\Theta$ and $n\geq 1$ define

$$\Omega_\theta(n,r)=\{\omega\in\Omega^n : \|c_n^2 L_n(\theta^*)\|<\frac{\tilde{\gamma}_Q(\theta)r}{2}, \sup_{y:\|y-\theta^*\|\leq r}\|\epsilon_n(y,\theta^*)\|<\frac{\tilde{\gamma}_Q(\theta)}{2}\},$$

where $\tilde{\gamma}_Q(\theta)=\inf_{u:\|u\|=1}\{u'\gamma_Q(\theta)u\}>0$, since $\gamma_Q(\theta)$ is positive definite.

Obviously, $\Omega_\theta(n,r)\in\mathcal{F}_T^n$. Hence if $\omega\in\Omega_\theta(n,r)$, then in view of $u'\gamma_Q(\theta)u\geq\tilde{\gamma}_Q(\theta)\|u\|^2$, from (4.3) we get (for $\|u\|=r$)

$$(c_n^2 L_n(\theta^*+u),u)<0.$$

Since the mapping $u\rightsquigarrow c_n^2 L_n(\theta^*+u)$ is continuous w.r.t. u, then for $\|u\|\leq r$ the equation

$$c_n^2 L_n(\theta^*+u)=0$$

has at least one solution $u_n(\theta^*)$ with $\|u_n(\theta^*)\|\leq r$ (this is a wellknown fact in the classical analysis).

It can be easily seen that if $\omega\in\Omega_\theta(n,r)$ and $\|u\|\leq r$, then the matrix $c_n^2\dot{L}_n(\theta^*+u)$ is negative definite.

On the other hand, for $\omega\in\Omega_\theta(n,r)$ and $\|u\|\leq r$

$$L_n(\theta^*+u,\omega)-L_n(\theta^*+u_n(\theta^*),\omega)=\int_0^1\frac{\partial}{\partial\alpha}[L_n((\theta^*+u_n(\theta^*))+\alpha(u-u_n(\theta^*)),\omega)d\alpha.$$

Consequently,

$$L_n(\theta^*+u,\omega)=\int_0^1\dot{L}_n(\theta^*+u_n(\theta^*)+\alpha(u-u_n(\theta^*)),\omega)(u-u_n(\theta^*))d\alpha$$

and

$$(L_n(\theta^*+u,\omega),u-u_n(\theta^*))=\int_0^1(u-u_n(\theta^*))'\dot{L}_n(\theta^*+u_n(\theta^*)+\alpha(u-u_n(\theta^*)),\omega)(u-u_n(\theta^*))d\alpha<0,$$

provided $u\neq u_n(\theta^*)$. Hence $L_n(\theta^*+u,\omega)\neq 0$ for $\|u\|\leq r$ and $u\neq u_n(\theta^*)$. By the construction of the set $\Omega_\theta(n,r)$ and conditions c), d) and e) of the lemma it is easily seen that

$$\lim_{r\to 0}\varliminf_{n\to\infty} Q_\theta^n\{\Omega_\theta(n,r)\}=1.$$

3°. We shall construct the sequence $\hat{\theta}=(\hat{\theta}_n),n\geq 1$ with properties I, II and III. Define

$$\Omega_n^\theta=\bigcup_{k>0}\Omega_\theta(u,\frac{1}{k}).$$

Evidently, $\Omega_n^\theta\in\mathcal{F}_T^n$. Let $\omega\in\Omega_n^\theta$. Then from the previous statement it follows that there exists a number $k(\omega)>0$ such that the equation $L_n(y)=0$ has the unique solution $\check{\theta}^n(\omega,\theta)$ in the ball $\|y-\theta^*\|\leq\frac{1}{k(\omega)}$ with the mapping $\omega\rightsquigarrow\check{\theta}^n(\omega,\theta)$ defined on $(\Omega_n^\theta,\Omega_n^\theta\cap\mathcal{F}_T^n)$, which is $\Omega_n^\theta\cap\mathcal{F}_T^n$-measurable (see, e.g. [4])

Set

$$\check{\theta}^n(\omega)=\begin{cases}\check{\theta}^n(\omega,\theta), & \omega\in\Omega_n^\theta,\\ \theta_0, & \omega\notin\Omega_n^\theta,\end{cases}$$

where θ_0 is a certain point in Θ.

It is easily seen that by construction $\hat{\theta}_n(\omega,\theta)$ possesses properties I , II and III.

4°. Finally, we shall prove assertion IV. By the decomposition (4.1) we have

$$\|(c_nL_n(\hat\theta_n)-c_nL_n(\theta^*))-\gamma_Q(\theta)c_n^{-1}(\hat\theta_n-\theta^*)\|\leqslant\|\epsilon_n(\hat\theta_n,\theta^*).\ \gamma_Q^{-1}(\theta)\|\cdot\|\gamma_Q(\theta)c_n^{-1}(\hat\theta_n-\theta^*)\| \tag{4.4}$$

and

$$\lim_{n\to\infty} Q_\theta^n\{\|\epsilon_n(\hat\theta_n,\theta^*)\|\geqslant\rho\}=0 \text{ for } \forall\rho>0,$$

which follows directly from the relation

$$\{\|\hat\theta_n-\theta^*\|\leqslant r\}\cap\{\sup_{y:\|y-\theta^*\|\leqslant r}\|\epsilon_n(y,\theta^*)\|<\rho\}\subset\{\|\epsilon_n(\hat\theta_n,\theta^*)\|<\rho\}$$

The assertion of problem 2 in [6], section 1.4, can be generalized as follows: let $\xi_n\xrightarrow{\mathfrak{D}}\xi$ (that is ξ_n converges weakly to ξ) and let the inequality

$$\|\xi_n-\eta_n\|\leqslant\xi_n\|\eta_n\|+\alpha_n$$

holds with $\mathcal{G}\xrightarrow{P}0$ and $\alpha_n\xrightarrow{P}0$, then $\eta_n\xrightarrow{\mathfrak{D}}\xi$. Applying now this assertion with $\alpha_n\equiv0$ to the case in question we get

$$\lim_{n\to\infty}\mathcal{L}\{\gamma_Q(\theta)c_n^{-1}(\hat\theta_n-\theta^*)|Q_\theta^n\}=\lim_{n\to\infty}\mathcal{L}\{c_nL_n(\theta^*)|Q_\theta^n\}.$$

This completes the proof of the lemma. □

Note that the proof of the lemma is essentially based on Cramer-Dugue's method generalized by A. Le Breton in [4].

Remark 4.1. It follows from the proof of Lemma 4.1 that all assertions remain true if the condition b) and c) are changed by the following condition b'):

(i) for each $n\geqslant1$ the random vector $L_n(\theta)$ is θ-continuous Q_θ^n- a.s.,

(ii) the sequence $L_n(\theta),n\geqslant1$ is asymptotically differentiable in the following sence: there exists the sequence of random matrices $\dot L_n(\theta),n\geqslant1$ such that

$$\lim_{r\to0}\overline{\lim_{n\to\infty}}\ P_{\theta^*}^n\{\sup_{\|y-\theta^*\|<r}c_n^2\frac{\|L_n(y)-L_n(\theta^*)-\dot L_n(\theta^*)(y-\theta^*)\|}{\|y-\theta^*\|}>\rho\}=0$$

for all $\rho>0$.

Remark 4.2. It is easy to see that for $\hat\theta=(\hat\theta_n)_{n\geqslant1}$ we have asymptotic expression ($\Theta\subset R^1$).

$$\hat\theta_n=\theta^*-\frac{L_n(\theta^*)}{\dot L_n(\theta^*)}+r_n(\theta^*,\hat\theta_n),$$

where

$$r_n(\theta^*,\hat\theta_n)=\frac{\epsilon_n(\theta^*,\hat\theta_n)(\hat\theta_n-\theta^*)}{c_n^2\dot L_n(\theta^*)}+\frac{c_n^2L_n(\hat\theta_n)}{c_n^2\dot L_n(\theta^*)}.$$

Under the conditions of Lemma 4.1 (see also Remark 4.1) we get using assertions I and II that $r_n(\theta^*,\hat\theta_n)\xrightarrow{P_{\theta^*}^n}0$.

This expression allows us (see remark 5.3 below) to conclude that M-estimator $\hat\theta=(\hat\theta_n)_{n\geqslant1}$ is a CLAN-estimator (CLAN: consistent linear asymptotically normal).

4.2. Global asymptotic behaviour of the solution of the equation $L_n(y,\omega)=0$.

We assume $\Theta=(a,b)^d$, and for convenience we set $a=-\infty$ and $b=+\infty$.

For every $\theta\in\theta$ consider the set

$$S_\theta=\{\hat\theta=(\hat\theta_n)_{n\geqslant1}:\text{for every } n\geqslant1\ \hat\theta_n\in\mathcal{F}_T^n \text{ and } Q_\theta^n-\lim_{n\to\infty}c_n^2L_n(\hat\theta_n)=0\}.$$

THEOREM 4.1. *Let the following condition (sup c) holds:*
(sup c_1) *the function $\Delta_Q(\theta,y)$ is y-continuous for every θ,*
(sup c_2) *for any $K, 0<K<\infty$, $\rho>0$*

$$\lim_{n\to\infty} Q_\theta^n\{\sup_{\|y\|\leq K}\|c_n^2L_n(y)-\Delta_Q(\theta,y)\|>\rho\}=0.$$

Then

I. *The following alternative holds: if $\hat{\theta}\in S_\theta$, then either*

$$Q_\theta^n-\lim_{n\to\infty}\hat{\theta}_n=\theta^*=b(\theta) \tag{4.5}$$

or

$$\overline{\lim_{n\to\infty}}\ Q_\theta^n\{\|\hat{\theta}_n\|>K\}>0 \tag{4.6}$$

for any K, $0<K<\infty$.

II. *If, in addition, the condition*

$$(c^+)\qquad \lim_{\|y\|\to\infty}\|\Delta_Q(\theta,y)\|=c(\theta)>0$$

holds and for all $\rho>0$

$$\lim_{n\to\infty} Q_\theta^n\{\sup_{y\in\Theta}\|c_n^2L_n(y)-\Delta_Q(\theta,y)\|>\rho\}=0,$$

then (4.5) *is valid.*

PROOF. Let $\hat{\theta}=(\hat{\theta}_n)_{n\geq1}\in S_\theta$ and suppose that (4.6) is not satisfied. Then there is a number $K_0>0$ such that

$$\lim_{n\to\infty} Q_\theta^n\{\|\hat{\theta}_n\|>K_0\}=0.$$

Therefore

$$Q_\theta^n\{\|c_n^2L_n(\hat{\theta}_n)-\Delta_Q(\theta,\hat{\theta}_n)\|>\rho\}\leq Q_\theta^n\{\|\hat{\theta}_n\|>K_0\}+Q_\theta^n\{\|c_n^2L_n(\hat{\theta}_n)-\Delta_Q(\theta,\hat{\theta}_n)\|>\rho,\ \|\hat{\theta}_n\|\leq K_0\}$$
$$\leq Q_\theta^n\{\|\hat{\theta}_n\|>K_0\}+Q_\theta^n\{\sup_{\|y\|\leq K_0}\|c_n^2L_n(y)-\Delta_Q(\theta,y)\|>\rho\}\to0.$$

On the other hand

$$Q_\theta^n-\lim_{n\to\infty} c_n^2L_n(\hat{\theta}_n)=0$$

and, hence,

$$Q_\theta^n-\lim_{n\to\infty}\Delta_Q(\theta,\hat{\theta}_n)=0. \tag{4.7}$$

Now assume that (4.5) fails too. Then one can choose $\epsilon>0$ such that

$$\overline{\lim_{n\to\infty}}\ Q_\theta^n\{\|\hat{\theta}_n-b(\theta)\|>\epsilon\}>0.$$

By condition (sup c_1) for any $\epsilon>0$

$$\Delta(\epsilon)=\inf_{y:\|y-b(\theta)\|>\epsilon,\|y\|\leq K_0}\|\Delta_Q(\theta,y)\|>0.$$

This gives

$$\overline{\lim_{n\to\infty}}\ Q_\theta^n\{\|\Delta_Q(\theta,\hat{\theta}_n)\|>\Delta(\epsilon)\}\geq\overline{\lim_{n\to\infty}}\ Q_\theta^n\{\|\Delta_Q(\theta,\hat{\theta}_n)\|>\Delta(\epsilon),\|\hat{\theta}_n\|\leq K_0\}$$
$$\geq\overline{\lim_{n\to\infty}}\ Q_\theta^n\{\|\hat{\theta}_n-b(\theta)\|>\epsilon,\ \|\hat{\theta}_n\|\leq K_0\}>0,$$

which contradicts (4.7).

In order to prove the second assertion of the theorem it suffices to note that under condition (c^+)

$$\inf_{y:\|y-b(\theta)\|\geq\epsilon} \|\Delta_Q(\theta,y)\|>0$$

and to repeat the previous arguments. □

For every $n\geq 1$ consider the set $A_n=\{\omega$: the equation $L_n(y,\omega)=0$ has at least one (measurable) solution$\}$

Evidently, for any $\theta\in\Theta$ we have $\Omega_n^\theta\subset A_n$, where the set Ω_n^θ is defined in 3° (see the proof of Lemma 4.1). Since under the conditions of Lemma 4.1 $Q_\theta^n\{\Omega_n^\theta\}\to 1$, for any $\theta\in\Theta$ we have

$$\lim_{n\to\infty} Q_\theta^n\{A_n\}=1.$$

Define the set S_{sol} of sequences $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ by introducing first the set

$$S_n=\{\tilde{\theta}_n:\tilde{\theta}_n\in\mathcal{F}_T^n, L_n(\tilde{\theta}_n)=0,\ \omega\in A_n;\tilde{\theta}_n=\theta_0,\omega\in\overline{A}^n\}$$

for each $n\geq 1$, and then setting

$$S_{sol}=\{\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}:\forall n\geq 1,\tilde{\theta}_n\in S_n\}$$

Evidently, if $B_n\in\mathcal{F}_T^n$ and $\tilde{\theta}_n^1,\ \tilde{\theta}_n^2\in S_n$, then $\tilde{\theta}_n^3=\tilde{\theta}_n^1 I_{B_n}+\tilde{\theta}_n^2 I_{\overline{B}_n}\in S_n$.

COROLLARY 4.1. *If in addition to the conditions of Lemma 4.1 condition (sup c) is satisfied for any* $\theta\in\Theta$, *then there exists an estimator* $\theta^*=(\theta_n^*)_{n\geq 1}\in S_{sol}$ *such that*

$$Q_\theta^n-\lim_{n\to\infty}\theta_n^*=b(\theta) \tag{4.8}$$

for any $\theta\in\Theta$.

If, besides, for any $\theta\in\Theta$ *condition* (c^+) *is satisfied, then any estimator* $\hat{\theta}=(\hat{\theta}_n)_{n\geq 1}\in S_{sol}$ *has property* (4.8).

PROOF. It is sufficient to construct an estimator $\theta^*=(\theta_n^*)_{n\geq 1}$ for which (4.6) fails for each $\theta\in\Theta$.

For any $n\geq 1$ and $\epsilon>0$ there is $\theta_n^*\in S_n$ such that

$$\|\theta_n^*\|\leq \operatorname*{essinf}_{\tilde{\theta}_n\in S_n}\|\tilde{\theta}_n\|+\epsilon \quad P^n-a.s.$$

By virtue of Lemma 4.1, for any $\theta\in\Theta$ there exists an estimator $\hat{\theta}(\theta)=\{\hat{\theta}_n(\theta)\}_{n\geq 1}$ such that

$$\lim_{n\to\infty} Q_\theta^n\{L_n(\hat{\theta}_n(\theta))=0\}=1 \tag{4.9}$$

and

$$Q_\theta^n-\lim_{n\to\infty}\hat{\theta}_n(\theta)=b(\theta). \tag{4.10}$$

For every c, $0<c<\infty$, we have

$$\overline{\lim_{n\to\infty}}\, Q_\theta^n\{\|\theta_n^*\|>c\}\leq \overline{\lim_{n\to\infty}}\, Q_\theta^n\{\|\theta_n^*\|>c, L_n(\hat{\theta}_n(\theta))\neq 0\}+\overline{\lim_{n\to\infty}}\, Q_\theta^n\{\|\theta_n^*\|>c,\ L_n(\hat{\theta}_n(\theta))=0\}$$
$$\leq \overline{\lim_{n\to\infty}}\, Q_\theta^n\{L_n(\hat{\theta}_n(\theta))\neq 0\}+\overline{\lim_{n\to\infty}}\, Q_\theta^n\{\|\hat{\theta}_n(\theta)\|+\epsilon>c\}$$

The first term on the right-hand side of this inequality converges to zero by virtue of (4.9), and the second one by virtue of (4.10) for sufficiently large c. □

4.3. Stable asymptotic solutions
Consider a set of sequences of experiments

$$\{\mathbf{E}=(\mathscr{E}_n)_{n\geq 1}=(\Omega^n,\mathscr{F}^n,F^n,Q_\theta^n,P^n)_{n\geq 1},\theta\in\Theta\subset R^d\},$$

each of which is regular and ergodic. Suppose that for any sequence of experiments $(\mathscr{E}_{n\geq 1})$ condition c) of Lemma 4.1 is satisfied with $b(\theta)=\theta$.

DEFINITION 4.1. The sequence $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}\in S_{sol}$ is called stable if for any family of measures $\{Q_\theta^n, n\geq 1, \theta\in\Theta\}$ defining the sequence of experiments $(\mathscr{E}_n)_{n\geq 1}$ we have

$$\tilde{\theta}_n \xrightarrow{Q_\theta^n} \theta,$$

for all $\theta\in\Theta$.

THEOREM 4.2. *Let for any $(\mathscr{E}_n)_{n\geq 1}$ the conditions of Lemma 4.1 be satisfied as well as (supc) of theorem 4.1. Then there exists stable sequence $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}\in S_{sol}$. If, in addition, condition (c^+) is satisfied, then any sequence $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}\in S_{sol}$ is stable.*

PROOF. Obviously, the desired sequence $\theta^*=(\theta_n^*)_{n\geq 1}$ is constructed in the course of proving corollary 4.1. □

Theorem 4.2 has a formal character. An example of a nontrivial situation in which the conditions of the theorem are satisfied, will be given below.

5. ASYMPTOTIC PROPERTIES OF *M*-ESTIMATORS.
5.1. Suppose that the martingale m_θ^n (defining the *M*-estimator) is regularly related to the experiment $\mathbf{E}$ in the following sence: components of the vector $m_\theta^n=(m_\theta^{n,1},\ldots,m_\theta^{n,d})'$ satisfy condition (1) of section 2.3 with $L(m_\theta^n,M_\theta^n)$, $L(\dot{m}_\theta^n,M_\theta^n):=L(\frac{\partial}{\partial\theta_i}m_\theta^{n,j},M_\theta^n))_{i,j=\overline{1,d}}\in M^2(P_\theta^n)$ and the matrix

$$\frac{\partial}{\partial\theta}L(m_\theta^n,M_\theta^n):=\dot{L}(m_\theta^n,M_\theta^n):=(\frac{\partial}{\partial\theta_i}L(m_\theta^{n,j},M_\theta^n))_{i,j=\overline{1,d}}$$

which takes the form

$$\dot{L}(m_\theta^n,M_\theta^n)=L(\dot{m}_\theta^n,M_\theta^n)-[L(m_\theta^n,M_\theta^n),L(\dot{M}_\theta^n,M_\theta^n)],$$

where $[M,N]=[M^i,N^j],i,j=\overline{1,d}$ is the matrix of mutual square variations and, finally the elements of the matrix $\dot{L}(m_\theta^n,M_\theta^n)$ are θ-continuous for all $t\in[0,T]$ P^n-a.s.

The asymptotic behaviour of the *M*-estimator defined by the martingale m_θ^n described by the following theorem.

THEOREM 5.1. *Let for every $\theta\in\Theta$ the following conditions be satisfied:*

(a) $\quad P_\theta^n-\lim\limits_{n\to\infty} c_n^2\langle L(m_\theta^n,M_\theta^n)\rangle_T=\Gamma(\theta),$

where $\Gamma(\theta)=(\Gamma_{ij}(\theta)),i,j=\overline{1,d}$, is a positive definite symmetric matrix;

(b) $\quad P_\theta^n-\lim\limits_{n\to\infty} c_n^2[L(m_\theta^n,M_\theta^n),\ L(\dot{M}_\theta^n,M_\theta^n)]_T=\gamma(\theta),$

where $\gamma(\theta)=(\gamma_{ij}(\theta)),i,j=\overline{1,d}$, is a positive definite matrix;

(c) $\quad P_\theta^n-\lim\limits_{n\to\infty} c_n^2 L_T(\dot{m}_\theta^n,M_\theta^n)=0;$

(d) $\quad P_\theta^n-\lim\limits_{n\to\infty}\int_0^T\int_{\|x\|>\epsilon}\|x\|^2\,\nu_\theta^n(ds,dx)=0,$ *for* $\forall\epsilon\in(0,1],$

where ν_θ^n is the P_θ^n-compensator of the jump measure of the process $c_n L(m_\theta^n, M_\theta^n)$;

(e) $$\lim_{r\to 0} \overline{\lim_{n\to\infty}} P_\theta^n\{\sup_{y:\|y-\theta\|<r} c_n^2 \|\dot L(m_y^n, M_y^n) - \dot L(m_\theta^n, M_\theta^n)\| \geq \rho\} = 0,$$

for any $\rho>0$ ($\|A\|$ is a norm of a matrix A). Then for any $\theta\in\Theta$ there exists an M-estimator $\hat\theta = (\hat\theta_n)_{n\geq 1}$ with properties

(I) $\lim_{n\to\infty} P_\theta^n\{L(m_{\hat\theta_n}^n, M_{\hat\theta_n}^n) = 0\} = 1$;

(II) $P_\theta^n - \lim_{n\to\infty} \hat\theta_n = \theta$;

(III) *if $\tilde\theta = (\tilde\theta_n)_{n\geq 1}$ is another estimator with properties* (I) *and* (II), *then*

$$\lim_{n\to\infty} P_\theta^n\{\tilde\theta_n = \hat\theta_n\} = 1;$$

(IV) $\mathcal{L}\{c_n^{-1}(\hat\theta_n - \theta) | P_\theta^n\} \Rightarrow N(0, \gamma^{-1}(\theta)\Gamma(\theta)(\gamma^{-1}(\theta))')$.

PROOF. To prove the theorem apply first Lemma 4.1 with $L(m_\theta^n, M_\theta^n)$ instead of $L_n(\theta)$ and P_θ^n instead of Q_θ^n. (Observe that condition c) of Lemma 4.1 is satisfied with $b(\theta) = \theta$ by virtue of condition (a) of the theorem, and condition d) of Lemma 4.1 by virtue of conditions (b) and (c)). Apply then the functional central limit theorem for martingales [3] which, under conditions (a) and (d) of the theorem, gives

$$\mathcal{L}\{c_n L(m_\theta^n, M_\theta^n) | P_\theta^n\} \Rightarrow N(0, \Gamma(\theta)).$$ □

REMARK 5.1. Conditions (c), (d) and (e) may be, if necessary, expressed in terms of characteristics of envolved martingales (that are $m_\theta^n, M_\theta^n, \dot M_\theta^n$ and $\dot m_\theta^n$).

REMARK 5.2. It is not hard to see that in multidimensional case we also have a "drift theorem" (that is Theorem 3.1 [8]). Assume now that the conditions of the "global theorem" (that is Theorem 4.1) are satisfied for two sequences of measures $\{P_\theta^n\}$ and $\{\tilde P_\theta^n\}$ (hypothetical and alternative). Then the estimator $\{\theta_n^*\}$ constructed in Corollary 4.1 (i.e. the estimator with the minimal modulus) will get a drift.

REMARK 5.3. Rewrite the relation obtained in Remark 4.2 in the following form (here $\theta^* = \theta$)

$$\hat\theta_n = \theta + \frac{L(m_\theta^n, M_\theta^n)}{\langle L(m_\theta^n, M_\theta^n), L(\dot M_\theta^n M_\theta^n)\rangle} + R_n(\theta, \hat\theta_n),$$

where

$$R_n(\theta, \hat\theta_n) = r_n(\theta, \hat\theta_n) - \frac{L(m_\theta^n, M_\theta^n)}{\dot L(m_\theta^n, M_\theta^n)} - \frac{L(m_\theta^n, M_\theta^n)}{\langle L(m_\theta^n, M_\theta^n), L(\dot M_\theta^n, M_\theta^n)\rangle}.$$

From the conditions (b), (c) and (d) of Theorem 5.1 and formula

$$\dot L(m_\theta^n, M_\theta^n) = L(\dot m_\theta^n, M_\theta^n) - [L(m_\theta^n, M_\theta^n), L(\dot M_\theta^n, M_\theta^n)]$$

it is easy to see that

$$R_n(\theta, \hat\theta_n) \xrightarrow{P_\theta^n} 0.$$

Hence M-estimator $\hat\theta = (\hat\theta_n)_{n\geq 1}$ is a CLAN-estimator.

Similarily to Theorem 6.2 [8], we can establish conditions sufficient for the "global" Theorem 4.1, provided the involved martingale has an integral representation. However, we will not linger on this problem here.

6. Stability of estimators. Partial likelihood estimators

6.1. In this section we shall deal with stability problems for M-estimator defined by equation

$$L(m_\theta^n, M_\theta^n) = 0.$$

In particular, we rise the following question: how to choose the martingales $\{m_\theta^n, \theta \in \Theta\}_{n \geq 1}$ in order to guarantee the stability of the corresponding estimator? Lemma 4.1 (in particular the condition c)) suggests that the main requirement for the stability is the following: for any family of measures $\{Q_\theta^n, \theta \in \Theta\}_{n \geq 1}$, defining (in subsection 4.3) a set of sequences of experiments $\mathbf{E} = (\mathcal{E}_n)_{n \geq 1}$, the following relations hold

$$Q_\theta^n - \lim_{n \to \infty} c_n^2 L(m_\theta^n, M_\theta^n) = 0, \quad Q_\theta^n - \lim_{n \to \infty} c_n^2 \dot{L}(m_\theta^n, M_\theta^n) = \text{const.}$$

This suggests naturally that the processes $L(m_\theta^n, M_\theta^n)$ and $L(\dot{m}_\theta^n, M_\theta^n)$ behave as martingales w.r.t. any such measure Q_θ^n.

We present now a special scheme (which can be viewed as a natural generalization of the partial likelihood scheme) of such situation.

6.2. *Partial Likelihood Estimators*

Let $(\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{0 \leq t \leq T}, P)$ be a probability space with a filtration satisfying the usual conditions. Besides, let two families $\{N_\theta, \theta \in \Theta\}$ and H (of martingales from $H^1(F,P)$) be given on this space,[1] such that the following conditions are satisfied:

(I_1) for all $\theta \in \Theta$ and $X \in H$ we have $N_\theta \perp H$ and $N_\theta(0) = X(0) = 0$;

(I_2) for any $X \in H$, $\Delta X > -1$.

(I_3) for any $\theta \in \Theta$ and $X \in H$ Dolean's exponential has the property

$$\mathcal{E}_T(X)\mathcal{E}_T(N_\theta) = \mathcal{E}_T(M_\theta^X),$$

where $M_\theta^X = X + N_\theta + [X, N_\theta]$ defines the density of some measure Q_θ^X w.r.t. a measure P, i.e. $E\mathcal{E}_T(M_\theta^X) = 1$.

(I_4) for any $\theta \in \Theta$, $E\mathcal{E}_T(N_\theta) = 1$, $\Delta M_\theta > -1$, $N_\theta(0) = 0$.

Further, let $\mathcal{L}^1(H)$ be the stable subspace[2] of martingales from $H^1(F,P)$ generated by H. Denote by H_0 the largest subset from $\mathcal{L}^1(H)$ satisfying together with $\{N_\theta, \theta \in \Theta\}$ conditions $(I_1)-(I_4)$. Evidently $\mathcal{L}^1(H_0) = \mathcal{L}^1(H)$.

Consider the set $\mathbf{X}$ of various mappings $X: \Theta \to H_0$ and the set of experiments $\mathbf{E} = \{\mathbf{E}^X, X \in \mathbf{X}\}$ where

$$\mathbf{E}^X = (\Omega, \mathcal{F}, F, Q^X = \{Q_\theta^X, \theta \in \Theta\}, P),$$

with a family of measures $Q^X = \{Q_\theta^X, \theta \in \Theta\}$ such that for all $\theta \in \Theta$, $Q_\theta^X \sim P$, $Q_\theta^X|\mathcal{F}_0 = P|\mathcal{F}_0$ and

$$\frac{dQ_\theta^X}{dP} = \mathcal{E}(X_\theta)\mathcal{E}(N_\theta) \tag{6.1}$$

For convenience let us fix some basic experiment $\mathbf{E}^{X_0}$. All experiments from $\mathbf{E}$ are considered as alternatives. Therefore the family of martingales H_0 will be called the family defining alternatives.

Following *Cox* [11], we shall call the second multiplier in the right-hand side of (6.1) the partial likelihood and the estimator $\hat{\theta} = (\hat{\theta}_t), 0 \leq t \leq T$ defined by the relation

$$\hat{\theta}_t = \arg\max_\theta \mathcal{E}_t(N_\theta)$$

1) $H^1(F,P) = \{M \in \mathcal{L}_{loc}(F,P) : E \sup_{0 \leq t \leq T} |M_t| < \infty\}$.

2) Two local martingales M and N are called orthogonal ($M \perp N$) if their product MN is a local martingale and $M(0)N(0) = 0$. $\mathcal{L}^1(H)$ is a minimal subspace of $H^1(F,P)$ containing (together) with every $M \in H$ all stochastic integrals $h \cdot H$ with $h \in L^1(M)$, where $L^1(M) = \{h : E(\int_0^T h^2 d[M])^{1/2} < \infty\}$.

as the solution of the equation

$$L_t(\dot{N}_\theta, N_\theta) = 0,$$

the partial likelihood estimator (PLE).[3)].

Defining the PLE as the root of the above equation, we certainly assume that for all t there exist the derivatives $\frac{\partial}{\partial\theta} N_\theta := \dot{N}_\theta$ and $\frac{\partial}{\partial\theta} \ln \mathcal{E}_t(N_\theta)$ $P-a.s.$, besides $\dot{N}_\theta \in M_{loc}(F,P)$ and the equality

$$\frac{\partial}{\partial\theta} \ln \mathcal{E}_t(N_\theta) = L_t(\dot{N}_\theta, N_\theta)$$

holds.

Before studying the asymptotic properties of PLE we give two lemmas which play a key role in our further arguments.

Let martingales N and m from $M_{loc}(F,P)$ be such that $N = N_{\theta_0}$ for some $\theta_0 \in \Theta$ and $m \perp H$.

LEMMA 6.1. *The process $L(m,N)$ is a P^X-local martingale w.r.t. any measure P^X ($P^X \sim P$) with the density of the form*

$$\frac{dP^X}{dP} = \mathcal{E}(X)\,\mathcal{E}(N) = \mathcal{E}(M^X), \tag{6.2}$$

where $X \in H_0$, $M^X = X + N + [X,N]$.

PROOF. In view of the generalized Girsanov theorem it is sufficient to prove that for any $X \in H_0$ a P-martingale m^X can be found such that

$$L(m,N) = L(m^X, M^X).$$

Since $m \perp H$ and, hence, $[m,X] := \langle m^c, X^c\rangle + \Sigma \Delta m \Delta X$ is a P-local martingale, we also have $m^X = m + [m,X] \in M_{loc}(F,P)$

We will show that

$$L(m^X, M^X) = L(m,N)$$

By definition of L-transformation we have

$$L(m^X, M^X) = m^X - \langle m^{X,c}, M^{X,c}\rangle - \sum \frac{\Delta m^X \Delta M^X}{1+\Delta M^X}.$$

Besides, since $[m,X]$, $[X,N] \in M^d_{loc}(F,P)$ we have

$$m^{X,c} = m^c$$

$$M^{X,c} = X^c + N^c$$

$$\Delta m^X = \Delta m(1-\Delta X),\ \Delta M^X = \Delta X(1+\Delta N) + \Delta N.$$

Therefore

$$L(m^X, M^X) = m + \langle m^c, X^c\rangle + \sum \Delta m \Delta X - \langle m^c, X^c + N^c\rangle$$
$$- \sum \frac{\Delta m(1+\Delta X)(\Delta X(1+\Delta N)+\Delta N)}{(1+\Delta X)(1+\Delta N)} m - \langle m^c, N^c\rangle - \sum \frac{\Delta m \Delta N}{1+\Delta N} = L(m,N). \quad \square$$

It follows from Lemma 6.1 that each local martingale $m, m \perp H_0$ defines a process $L(m,N)$, which is a local martingale w.r.t. all measures $P^X, X \in H_0$ simultaneously, with densities of form (6.2).

3) At the end of this section examples will be given which will make it clear that the introduced PLE are generalazions to the case of a general statistical experiment of the PLE introduced by *Cox* for models in discrete time.

The following lemma proves that the processes of the form $L(m,N)$ exhaust the class of all F-adapted processes with regular trajectories, which are local martingales w.r.t. all measures $P^X, X\in H_0$ (with property (6.2)) simultaneously.

LEMMA 6.2. *Let $l=(l_t, 0\leq t\leq T)$ be an F-adapted process with regular trajectories and the for all measures $P^x, X\in H_0$, with property (6.2), $l\in M_{loc}(F,P^X)$. Then a local P-martingale m can be found such that $m\perp H_0$ and*

$$l=L(m,N).$$

PROOF. Since for any $X\in H_0$ we have $l\in M_{loc}(F,P^X)$ and

$$\frac{dP}{dP^X}=\mathscr{E}(L(-M^X,M^X))$$

where $M^X=X+N+[X,N]$, then by the generalized Girsanov theorem for any $X\in H_0$ we have

$$L(l,L(-M^X,M^X))\in M_{loc}(F,P).$$

It can be easily computed that

$$L(l,L(-M^X,M^X))=l+[l,M^X].$$

Hence for any $X\in H_0$ we have the process $m^X=l+[l,M^X]$ is a local P-martingale. Since $0\equiv X\in H_0$ we have $l+[l,N]:=m\in M_{loc}(F,P)$ It can be easily seen that

$$l=L(m,N).$$

Indeed, since $\langle m^c,N^c\rangle=\langle l^c,N^c\rangle$ and $\Delta m=\Delta l(1+\Delta N)$ we have

$$l+[l,N]-\langle l^c,N^c\rangle-\sum\frac{\Delta l\Delta N(1+\Delta N)}{1+\Delta N}=l.$$

It remains to prove that $m\perp H_0$. We have $m^X-m\in M_{loc}(F,P)$ for all $X\in H_0$. But

$$m^X-m=\langle l^c,X^c\rangle+\sum\Delta l(1+\Delta N)\Delta X=[m,X],$$

since $\langle l^c,X^c\rangle=\langle m^c,X^c\rangle$, $\Delta m=\Delta l(1+\Delta N)$.

Hence for any $X\in H_0$ we have $[m,X]\in M_{loc}(F,P)$. □

REMARK. It can be easily seen that for any $X\in H_0$ the process l can be represented as

$$l=L(m^X,M^X).$$

To study the asymptotic properties of PLE, we consider the usual scheme of series.

For every $n\geq 1$ define all objects introduced above, in particular, families $\{N_\theta^n,\theta\in\Theta\}$, $H^n,H_0^n\subset H^1(F^n,P^n)$ satisfying conditions $I_1)-I_4)$ and the set $\mathbf{X}^n$ of all possible mappings $X^n:\Theta\to H_0^n$.

Consider the set $\mathbf{X}$ of sequences $X=(X^n)_{n\geq 1}$ with $X^n\in\mathbf{X}^n$, and associate with every $X\in\mathbf{X}$ a family of measures

$$Q^X=\{Q_\theta^{X,n},\theta\in\Theta\}_{n\geq 1}$$

which is such that for every $n\geq 1$ and all $\theta\in\Theta$ we have $Q^{X,n}\sim P^n$ and

$$\frac{dQ_\theta^{X,n}}{dP^n}=\mathscr{E}(X_\theta^n)\mathscr{E}(N_\theta^n)=\mathscr{E}(M_\theta^{X,n})$$

with

$$M_\theta^{X,n}=X_\theta^n+N_\theta^n+[X_\theta^n,N_\theta^n].$$

Consider corresponding experiment

$$\mathbf{E}^X = (\mathcal{E}^X_n)_{n\geq 1} = (\Omega^n, \mathcal{F}^n, F^n, \{Q^{X,n}_\theta, \theta\in\Theta\},\ P^n)_{n\geq 1}.$$

According to the above convention we shall fix some basic experiment $\mathbf{E}^{X_0}$ in the whole family of experiments $\mathbf{E} = (\mathbf{E}^X, X\in\mathbf{X})$. The corresponding $Q^{X_0,n}_\theta$ will be denoted by P^n_θ and $M^{X_0,n}_\theta$ by M^n_θ.

We will search for the family of martingales $\{m^n_\theta, \theta\in\Theta\}_{n\geq 1}$ which guarantee the stability (in the sense of Definition 4.1) of M-estimators defined by the equation

$$L(m^n_\theta, M^n_\theta) = 0 \tag{6.3}$$

w.r.t. all alternatives $\{Q^X, X\in\mathbf{X}\}$ forming a regular ergodic experiment.

As it has been noted in the introduction to the present section, it is essential for the stability that the processes $L(m^n_\theta, M^n_\theta)$ and $L(\dot{m}^n_\theta, M^n_\theta)$ are martingales w.r.t. all measures $Q^{X,n}_\theta$ (θ and n are fixed) simultaneously. As we have seen already, Lemmas 6.1 and 6.2 imply that only processes of the form $L(m, N^n_\theta)$ with some P-martingale $m\perp H^n_0$ are martingales w.r.t. all measures $Q^{X,n}_\theta, X\in\mathbf{X}$ It is evident therefore that at least those estimators are stable which are constructed by the equation

$$L(\overline{m}^n_\theta, N^n_\theta) = 0 \tag{6.4}$$

with some family of P^n-martingales $\{\overline{m}^n_\theta, \theta\in\Theta\}, n\geq 1$, such that for all $\theta\in\Theta$ and $n\geq 1$ we have $\overline{m}^n_\theta \perp H^n_0$.

REMARK. Equation (6.4) can be written as

$$L(m^{n,X}_\theta, M^{n,X}_\theta) = 0$$

for any $X\in\mathbf{X}$, where $m^{n,X}_\theta = \overline{m}^n_\theta + [\overline{m}^n_\theta, X^n_\theta]$.

Hence the estimators defined by (6.4) are usual M-estimators. Therefore to study the asymptotic properties of these estimators with respect to all alternatives $\{Q^X, X\in\mathbf{X}\}$, Theorem 5.1 can be applied.

On the other hand, one can immediately use Lemma 4.1 with $L_n(\theta) = L(\overline{m}^n_\theta, M^n_\theta)$ and thus avoid the smoothness requirement for the martingales X^n_θ. In what follows we shall write m^n_θ instead of $\overline{m}^n_\theta$.

For simplicity we consider one-dimensional case ($d=1$). The assertions of Theorem 6.1 below are true in multidimensional case ($d\geq 2$) as well.

Suppose that families $\{N^n_\theta, \theta\in\Theta\}_{n\geq 1}$ and $\{m^n_\theta, \theta\in\Theta\}_{n\geq 1}$ are such that the following conditions are satisfied:

(1) for every $n\geq 1$ and $0\leq t\leq T$ the mappings $\theta\rightsquigarrow m^n_\theta(t,\omega)$ are continuously differentiable P^n-a.s. w.r.t. θ, besides $\dot{N}^n_\theta, \dot{m}^n_\theta \in M_{loc}(F^n, P^n)$ and $m^n_\theta, \dot{N}^n_\theta, \dot{m}^n_\theta \perp H^n$ for all $\theta\in\Theta$;

(2) $\dfrac{\partial}{\partial\theta} L(m^n_\theta, N^n_\theta) = L(\dot{m}^n_\theta, N^n_\theta) - [L(m^n_\theta, N^n_\theta), L(\dot{N}^n_\theta, N^n_\theta)]$;

(3) $L(\dot{N}^n_\theta, N^n_\theta), L(m^n_\theta, N^n_\theta), L(\dot{m}^n_\theta, N^n_\theta) \in M^2(F^n, Q^{X,n}_\theta)$, for all $X\in\mathbf{X}$[1].

THEOREM 6.1. *Let $\{c_n\}, n\geq 1$ be a sequence with $c_n>0$, $c_n\to 0$, such that*

1) For all $Q^X = \{Q^{X,n}_\theta, \theta\in\Theta\}_{n\geq 1}$ and any $\theta\in\Theta$ the following conditions are satisfied:

(a) $Q^{X,n}_\theta - \lim\limits_{n\to\infty} c_n^2 \langle L(m^n_\theta, N^n_\theta)\rangle_T = \Gamma_X(\theta),\ \Gamma_X(\theta)>0$;

(b) $Q^{X,n}_\theta - \lim\limits_{n\to\infty} c_n^2 [L(m^n_\theta, N^n_\theta), L(\dot{N}^n_\theta, N^n_\theta)]_T = \gamma_X(\theta),\ \gamma_X(\theta)>0$;

(c) $Q^{X,n}_\theta - \lim\limits_{n\to\infty} c_n^2 L_T(\dot{m}^n_\theta, N^n_\theta) = 0$;

(d) $Q^{X,n}_\theta - \lim\limits_{n\to n} \int_0^T \int_{|U|>\epsilon} u^2 \nu^{n,X}_\theta(ds,du) = 0$ *for all $\epsilon\in(0,1]$, where $\nu^{X,n}_\theta$ is a $Q^{X,n}_\theta$ compansator of the jump*

1) Lemma 6.1 implies that under condition (1) we have $L(\dot{N}^n_\theta, N^n_\theta)$, $L(m^n_\theta, N^n_\theta)$, $L(\dot{m}^n_\theta, N^n_\theta) \in M_{loc}(F^n, Q^{X,n}_\theta), X\in\mathbf{X}$.

measure of the process $c_n L(m^n_\theta, N^n_\theta)$;

(e) $\lim_{r\to 0}\overline{\lim_{n\to\infty}} Q^{X,n}_\theta\{\sup_{y:|y-\theta|<r} c_n^2|\dot{L}(m^n_y,N^n_y)-\dot{L}(m^n_\theta,N^n_\theta)|>\rho\}=0$ *for any* $\rho>0$.

Then, for any θ *and* Q^X *there exists an M-estimator* $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ *(depending generally speaking on both* θ *and* Q^X*)*[1] *such that*

I. $\lim_{n\to\infty} Q^{X,n}_\theta\{L_T(m^n_{\tilde{\theta}_n}, N^n_{\tilde{\theta}_n})=0\}=1$,

II. $Q^{X,n}_\theta - \lim_{n\to\infty}\tilde{\theta}_n=0$,

III. if $\tilde{\tilde{\theta}}=(\tilde{\tilde{\theta}}_n)_{n\geq 1}$ *is another estimator with properties I and II, then*

$$\lim_{n\to\infty} Q^{X,n}_\theta\{\tilde{\theta}_n=\tilde{\tilde{\theta}}_n\}=1,$$

IV. $\mathcal{L}\{c_n^{-1}(\tilde{\theta}_n-\theta)|Q^{X,n}_\theta\}\Rightarrow N(0,\gamma_X^{-2}(\theta)\Gamma_X(\theta))$.

2) *If* $\Theta=(a,b)$ *(without loss of generality we can assume that* $a=-\infty, b=+\infty$*) and for any* $Q^X, X\in\mathbf{X}$ *condition (sup c) of Theorem 4.1 holds instead of condition (c) as well as* c) *of Lemma 4.1 with* $b(\theta)=\theta$*, then there exists a stable estimator* $\hat{\theta}=(\hat{\theta}_n)_{n\geq 1}$ *with respect to a class of alternatives* $\{Q^X, X\in\mathbf{X}\}$. *That is*

$$Q^{X,n}_\theta - \lim_{n\to\infty}\hat{\theta}_n=\theta$$

for any $\theta\in\Theta$ *and* $Q^X, X\in\mathbf{X}$.

PROOF. Assertion 1) is proved similarly to Theorem 5.1, while assertion 2) is an immediate corollary of Theorem 4.1. Evidently, the estimator $(\theta^*_n)_{n\geq 1}$ constructed in Corollary 4.1 can be taken as a stable estimator. Note that this estimator is independent of θ and the choice of measures Q^X. □

REMARK. In section 4 we have introduced sets S_θ and S_{sol} associated with the equation $L_n(\theta)=0$. Observe that the set S_{sol} was constructed only by means of this equation, while the set S_θ was depending on the sequence $\{Q^n_\theta\}_{n\geq 1}$.

In the course of studying the asymptotic behaviour of the solutions to equations $L(m^n_\theta, N^n_\theta)=0$ with various $\{m^n_\theta, \theta\in\Theta\}_{n\geq 1}$ and $Q^X=\{Q^{X,n}_\theta, \theta\in\Theta\}_{n\geq 1}$, $X\in\mathbf{X}$, we denote these sets by $S^{m,X}_\theta$ and S^m_{sol}, respectively. Evidently, if conditions of 1) in Theorem 6.1 are satisfied for all Q^X, then

$$S^m_{sol}\subset S^{m,X}_\theta$$

for all $\theta\in\Theta$ and $X\in\mathbf{X}$. Furthermore, any estimator $\tilde{\theta}(X,\theta)\in S^m_{sol}$. Therefore if the set S^m_{sol} has the property that for any $\tilde{\theta}, \tilde{\tilde{\theta}}\in S^m_{sol}$

$$\lim_{n\to\infty} Q^{X,n}_\theta\{\tilde{\theta}_n=\tilde{\tilde{\theta}}_n\}=1$$

for all $\theta\in\Theta$ and Q^X, then any estimator $\hat{\theta}\in S^m_{sol}$ is stable.

The PLE $\hat{\theta}$ is an important particular case when $m^n_\theta=\dot{N}^n_\theta$. Along with the stability w.r.t. a class of alternatives $(Q^X, X\in\mathbf{X})$, it also has a certain optimality property: PLE $\hat{\theta}$ has the minimal asymptotic variance among all estimators $\tilde{\theta}$ defined by the equation (6.4).

Indeed, since $[M,N]\leq[M]^{1/2}[N]^{1/2}$, for all $M,N\in M_{loc}(F,P)$, we have

$$\frac{\sigma^2\hat{\theta}}{\sigma^2\tilde{\theta}}=Q^{X,n}_\theta - \lim_{n\to\infty}\frac{[L(m^n_\theta,N^n_\theta),L(\dot{N}^n_\theta,N^n_\theta)]^2_T}{[L(m^n_\theta,N^n_\theta)]_T[(\dot{N}^n_\theta,N^n_\theta)]_T}\leq 1.$$

2) This estimator is sometimes denoted by $\tilde{\theta}(\theta,X)$.

6.3. *Examples.*

First we shall describe a rather general scheme which will include, as one can see below, a number of cases studied by various authors (e.g. [11], [12], [13], [17])

For every $n\geq 1$ let $(\Omega^n, \mathcal{F}^n, F^n, P^n)$ be a probability space with filtration satisfying the usual conditions, E some Lusin space,

$$\tilde{\Omega}^n = \Omega^n \times R_+ \times E,\ \tilde{\mathcal{F}}^n = \mathcal{F}^n \otimes \mathcal{B}(R_+) \otimes \mathcal{B}(E),$$

μ^n an integer-valued random measure on $(R_+ \times E, \mathcal{B}(R_+)\otimes\mathcal{B}(E))$ and ν^n its P^n-compensator. Denote $D^n = \{(\omega,t): \mu^n\{\omega,\{t\},E\}\neq 0\}$.

Following [1] introduce the classes

$$K(\mu^n) = \{W: W \in \tilde{0}(F^n,P^n), W*\mu_n \in \tilde{A}(P^n)\},$$

$$G^2(\mu^n) = \{W: W\in\tilde{\mathcal{P}}(F^n,P^n),\ \mathcal{S}((\tilde{W}^n)^2)^{q/2} \in \mathcal{A}(P^n)\}, q=1,2$$

$$H^q(\mu^n) = \{V: V\in K(\mu^n),\ M_{\mu}^{P^n}(VI_{]0,\infty[}|\tilde{\mathcal{P}}(F^n,P^n)) = 0,\quad (V^2\mu^n)^{q/2} \in \mathcal{A}(P^n)\},$$

where

$$\tilde{0}(F^n,P^n) = 0(F^n,P^n)\otimes\mathcal{B}(E), \tilde{\mathcal{P}}(F^n,P^n) = \mathcal{P}(F^n,P^n)\otimes\mathcal{B}(E),$$

$0(F^n,P^n)$, and $\mathcal{P}(F^n,P^n)$ are optional and predictable σ-algebras respectively, $\mathcal{A}(P^n)$ a class of processes with an integrable (w.r.t. the measure P^n) variation, and $\tilde{\mathcal{A}}(P^n)$ is a set of equivalence classes of optional and integrable random measures. That is $E^n\{1*|\mu^n|_\infty\}<\infty$,

$$\tilde{W}_t^n = \int_E W(\omega,t,x)\mu^n(\{t\},dx) - \hat{W}_t^n,$$

$$\hat{W}_t^n = \int_E W(\omega,t,x)\nu^n(\{t\},dx),$$

and $M_{\mu_n}^{P_n}$ is Dolean's measure associated with μ^n, i.e.

$$M_{\mu}^{P^n}(d\omega,dt,dx) = P^n(d\omega)\mu^n(\omega,dt,dx).$$

Let now

$$K^{q,1}(\mu^n) = \{W*(\mu^n-\nu^n): W\in G^q(\mu^n)\},$$

$$K^{q,2}(\mu^n) = \{V*\mu^n: V\in H^q(\mu^n)\},$$

$$K^{q,3}(\mu^n) = (M\in H^q: \Delta M = 0 \text{ on } D^n\},\ q=1,2.$$

Suppose, that the family $\{N_\theta^n, \theta\in\Theta\}, n\geq 1$ is such that:

1) for any $\theta\in\Theta$ and $n\geq 1$ we have $N_\theta^n \in K^{1,2}(\mu^n)$ i.e. $N_\theta^n = V_\theta^n * \mu^n$, with some function $V_\theta^n \in H^1(\mu^n)$,
2) $E^n\mathcal{E}_T(N_\theta^n) = 1$ for all $n\geq 1$ and $\theta\in\Theta$,
3) $\int_E V_\theta^n \mu(\{t\},dx) > -1$.

For every $n\geq 1$ introduce the set H^n

$$H^n = \left\{\begin{array}{ll} W*(\mu^n-\nu^n): & \text{a) } W\in G^1(\mu^n), \\ & \text{b) } (W-\hat{W}^n)V_\theta^n \in H^1_{loc}(\mu^n) \text{ for all } \theta\in\Theta, \\ & \text{c) } \tilde{W}_t^n > -1 \text{ for all } t\in[0,T], \\ & \text{d) } E^n\mathcal{E}_T(M_\theta^{W,n}) = 1^{1)}, \text{ for all } \theta\in\Theta, \end{array}\right\}$$

where $M_\theta^{W,n} = W*(\mu^n-\nu^n) + (W-\hat{W}^n+1)V_\theta^n*\mu^n$.

Obviously, for every $n\geq 1$ the families $\{N_\theta^n, \theta\in\Theta\}$ and H^n satisfy conditions $(I_1)-(I_4)$

1) See [1] for sufficient conditions

Indeed, condition b) implies orthogonality of N_θ^n and H^n for all $\theta\in\Theta$, since for any $W*(\mu^n-\nu^n)\in H^n$ and N_θ^n the mutual square variation $[W*(\mu^n-\nu^n),\ V_\theta^n*\mu^n]=(W-\hat{W})V_\theta^n*\mu^n$ is a P^n-martingale; condition c) means that for any $X=W*(\mu^n-\nu^n)\in H^n$ we have $\Delta X>-1$. Condition d) coinsides with condition (I_3). Conditions 2) and 3) in the definition of the class $\{N_\theta^n,\theta\in\Theta\}_{n\geq 1}$ are equivalent to the condition (I_4).

The class of alternatives is defined as usual by the set of all measures $\{Q^X,X\in\mathbf{X}\}$, $Q^X=\{Q_\theta^{X,n},\theta\in\Theta\}_{n\geq 1}$ with $Q_\theta^{X,n}\sim P^n$. Consider

$$\frac{dQ_\theta^{X,n}}{dP^n}=\mathcal{E}(X_\theta^n)\,\mathcal{E}(N_\theta^n)$$

with $X_\theta^n=W_\theta^n*(\mu^n-\nu^n)\in H^n$ and construct stable estimators by the equation

$$L(m_\theta^n,N_\theta^n)=0$$

where for all $n\geq 1$ and $\theta\in\Theta$ we have $m_\theta^n\perp H^n$. For example, one can take $m_\theta^n=V_\theta^{m,n}*\mu^n$ where $V_\theta^{m,n}$ are such that $V_\theta^{m,n}*\mu^n\in K^{1,2}(\mu^n)$ and for any W with $W*(\mu^n-\nu^n)\in H^n$ we have $V_\theta^{m,n}(W-\hat{W}^n)\in H^1_{loc}(\mu^n)$.

Remark 6.1. Since any totally discontinuous martingale $M\in H^1$ admits the representation

$$M=M^1+M^2+M^3$$

with $M^i\in K^{1,i}(\mu^n)$ (see [3]), one can in general use for constructing stable estimators orthogonal w.r.t. H^n martingales m_θ^n, the decomposition of which involves the three components. But it may turn out that the families $\{N_\theta^n,\theta\in\Theta\},n\geq 1$ and H^n are such that $\mathcal{L}^1(H^n)=K^{11}(\mu^n)$. In this case the decomposition of orthogonal to H^n martingales may obviously involve only two components of the classes $K^{1,2}(\mu^n)$ and $K^{1,3}(\mu^n)$ (since $M\perp H^n\Rightarrow M\perp K^{1,1}(\mu^n)$).

If $N_\theta^n\in K^{2,2}(\mu^n)$ for all $\theta\in\Theta$, then for H^n one can take a subset of $K^{2,2}(\mu^n)$ (or $K^{2,2}\cup K^{2,3}$) satisfying only conditions $(I_2),(I_3)$, since in this case condition (I_1) is automatically satisfied (recall that classes $K^{2,i},i=1,3$ are mutually orthogonal.)

We consider the examples of the above scheme.

Example 1. *i.i.d. observations.*

Let $(y_1,z_1),\ldots,(y_n,z_n),\ldots$ be independent identically distributed (i.i.d) observations with distribution P_θ, equivalent to some probability measure $P(dy,dz)=P_1(dy)\times P_2(dz)$ with density $f_0(y,z,\theta)=\phi_0(y,\theta)\psi(z,\theta|y)>0$. Then the likelihood ratio process $\rho_n(\theta)=\prod_{i=1}^n f_0(y_i,z_i,\theta)$ has a multiplicative decomposition

$$\rho_n(\theta)=\prod_{i=1}^n\phi_0(y_i,\theta)\psi(z_i,\theta|y_i) \tag{6.5}$$

According to Cox [11] the second cofactor of this product is called partial likelihood based on z, and the estimator $\hat{\theta}_n$, defined by the relation

$$\hat{\theta}_n=\arg\max_\theta\prod_{i=1}^n\psi(z_i;\theta|y_i)$$

or by the equation

$$\sum_{i=1}^n\dot{\psi}(z_i;\theta|y_i)/\psi(z_i;\theta|y_i)=0,$$

is called the partial likelihood estimator (PLE).

Let P_θ^n and P^n be direct products of measures P_θ and P on $(\Omega^n,\mathcal{F}^n)=((R^2)^n,\ \mathcal{B}((R^2)^n))$, respectively, $\mathcal{F}_t^n=\sigma(x_i,y_i;\ i\leq[nt]),\ 0\leq t\leq 1$. Let $P_\theta^n(t)$ and $P^n(t)$ be the restrictions of measures P_θ^n and P^n to

$\mathcal{F}_t^n$. Then

$$\rho_t^n(\theta)=\frac{dP_\theta^n(t)}{dP^n(t)}=\mathcal{E}_t(M_\theta^n)=\mathcal{E}_t(X_{0,\theta}^n),\ \mathcal{E}_t(N_\theta^n), \tag{6.6}$$

where the processes M_θ^n, $X_{0,\theta}^n$, N_θ^n defined by relations

$$M_\theta^n(t)=\sum_{i=1}^{[nt]}(f_0(y_i,z_i;\theta)-1),$$

$$X_{0,\theta}^n(t)=\sum_{i=1}^{[nt]}(\phi_0(y_i;\theta)-1),$$

$$N_\theta^n(t)=\sum_{i=1}^{[nt]}(\psi(z_i;\theta|y_i)-1)$$

are P^n-martingales, and the martingales $X_{0,\theta}^n$ and N_θ^n are orthogonal for all $\theta\in\Theta$ and $n\geq 1$.

Comparing decompositions (6.5) and (6.6) we see that the partial likelihood introduced by Cox coincides with $\mathcal{E}(N_\theta^n)$. Thus our concept of partial likelihood applicable to general statistical experiments is reduced in the present special case to that of Cox.

Further, let μ^n be the jump measure of the process $\sum_{i=1}^{[nt]} Y_i$ i.e.

$$\mu^n((0,t]\times B)=\sum_{i=1}^{[nt]} I_{\{y_i\in B\}},B\in\mathcal{B}(R^1)\ (\mu(\{n\},B)=I_{\{y_n\in B\}}),$$

and ν^n its P^n-compensator, i.e. $\nu^n((0,t]\times B)=[nt]\ P_1(B)$ $(\nu^n(\{n\},B)=P_1(B))$. The martingales $X_{0,\theta}^n$ and N_θ^n can be written as

$$X_{0,\theta}^n=W_\theta^n*(\mu^n-\nu^n)$$

$$N_\theta^n=V_\theta^n*\mu^n,$$

where $W_\theta^n(\omega,t,y)=\phi_0(t;\theta)$ and $V_\theta^n(\omega,t,y)=\psi(z_{[nt]};\theta|y)-1$ (note that $\omega=(y_i,z_i)_{i\geq 1}$). It can be easily shown that $W_\theta^n\in G^1(\mu^n)$, besides

$$E(\psi(z_i;\theta|y_i)|\mathcal{F}_{i-1}U\sigma(Y_i))=0$$

and $E(V_\theta^2*\mu^n)^{1/2}<\infty$.

As in the general case we could consider all alternative measures $Q^X=\{Q_\theta^{X,n},\theta\in\Theta\}_{n\geq 1}$ corresponding to arbitrary $X=\{X^n\}_{n\geq 1}$, $X_n:\Theta\to H^n$ with usual H^n, however we restrict ourselves by natural alternatives in the sense that in the set H^n we single out only martingales of the form

$$X^n(t)=\sum_{i=1}^{[nt]}(\phi(y_i)-1)$$

i.e. $X^n=\phi*(\mu^n-\nu^n)$ with all possible one-dimensional densities ϕ $(\int\phi P(dy)=1,\ \phi>0)$. (We shall denote this set by H_1^n).

This means that alternative experiments are specified as i.i.d. observations $(y_i,z_i)_{i\geq 1}$ with densities $f(y,z;\theta)=\phi(y;\theta)\psi(z;\theta|y)$ and all possible $\phi(y;\theta)$ such that $\int\phi(y;\theta)\ P_1(dy)=1$, $\phi(y;\theta)>0$.

Stable estimators w.r.t. such alternatives are naturally constructed by the equation

$$L(m_\theta^n,N_\theta^n)=0$$

with $m_\theta^n\perp H_1^n$. In particular, we can take for m_θ^n

$$m_\theta^n(t)=\sum_{i=1}^{[nt]}g(y_i,z_i;\theta)$$

i.e. $m_\theta^n=V_\theta*\mu^n$, where $V_\theta(\omega,t,y)=g(y,z_{[nt]};\theta)$ with $g(y,z;\theta)$ such that for all $\theta\in\Theta$ the following conditions hold:

1) $\int g(y,z;\theta)P_2(dz)=0$ for all $y\in R^1, \theta\in\Theta$,
2) $\int\int \phi(y)|g(y,z;\theta|\ P(dy,dz)<\infty$, for any positive densities ϕ (including $\phi\equiv 1$),
3) $\int \dot{g}(y,z;\theta)\ P_2(dz)=0$ for $\forall y\in R^1$ ($\dot{g}=\frac{\partial}{\partial\theta}g$),
4) $\int\int \phi(y)|\dot{g}(y,z,\theta)|P(dy,dz)<\infty$ for all $\phi>0$, $\int\phi P_1(dy)=1$.

Clearly, if $\psi(z;\theta|y)$ is differentiable in θ and ψ and $\dot{\psi}$ satisfies conditions 1)-4), then under the conditions of Theorem 6.1 the PLE will be stable w.r.t. the considered class of alternatives.

In the examples below we give only the decomposition of the local density, specify measures μ^n and ν^n, a class of natural alternatives and in some cases a class of martingales, defining the stable estimators.

EXAMPLE 2. *Censored i.i.d. observations.*

Let $X_1, X_2, \ldots, X_n, \cdots$ be i.i.d. observations with a density (w.r.t. some measure μ) $f(x,\theta), \theta\in\Theta\subset R^1$, $f(x,\theta)>0$. Suppose the density $f(x,\theta)$ is known only on the set $\{x:|x|\leq c\}$ and θ is to be estimated.

We can transform observations $\{X_i\}_{i\geq 1}$ into a two-dimensional sequence $\{y_i,z_i\}_{i\geq 1}$, where $y_i=X_iI_{\{|X_i|>c\}}$, $z_i=X_iI_{\{|X_i|\leq c\}}$. Denote by $P_\theta^{y,z}$ the distribution of a couple (y_1,z_1) on $(R^2,\mathcal{B}(R^2))$, computed w.r.t. the distribution of X_1 (i.e. w.r.t. the P_θ^x with $dP_\theta^x/d\mu=f(x,\theta)$), and by $P^{Y,z}$ the distribution of the couple (y_1,z_1), computed w.r.t. μ. For every $\theta\in\Theta$ we obviously have $P_\theta^{y,z}\sim P^{y,z}$. Let

$$\frac{dP_\theta^{y,z}}{dP^{y,z}}=f(y,z;\theta).$$

Evidently, the density $f(y,z;\theta)$ can be factorized in the following manner

$$f(y,z;\theta)=\phi_0(y,;\theta)\ \psi(z;\theta|y),$$

where

$$\phi_0(y;\theta)=\begin{cases} f(y;\theta), & \text{if } |y|>c \\ \int\limits_{|u|\leq c} f(u;\theta)\mu(du)/\int\limits_{|u|>c}\mu(du), & \text{if } |y|\leq c,\end{cases}$$

$$\psi(z;\theta|y)=\begin{cases} f(z;\theta)\dfrac{\int\limits_{|u|>c}\mu(du)}{\int\limits_{|u|\leq c} f(u:\theta)\mu(du)}, & \text{if } y=0,\ |z|\leq c, \\ 1, & \text{otherwise.}\end{cases}$$

Here

$$\phi_0(y;\theta)=\frac{dP_\theta^y}{dP^y}$$

with the marginals P_θ^y and P^y of the distributions $P_\theta^{y,z}$ and $P^{y,z}$, respectively. Recall that

$$P^y(dy)=\begin{cases}\mu(dy), & |y|>c, \\ \delta_0(dy)\int\limits_{|u|\leq c}\mu(du), & |y|\leq c,\end{cases}$$

where $\delta_0(dy)$ is Diraque's measure.

Hence the initial problem is reduced to that of Example 1 as it can be formulated as the problem of estimating the parameter θ by means of two-dimensional observations $(y_i,z_i)_{i\geq 1}$ with factorized densities $f(y,z;\theta)=\phi(y:\theta)\psi(z;\theta|y)$. As above, we have

$$\rho_t^h(\theta)=\mathcal{E}_t(X_{0,\theta}^n)\,\mathcal{E}_t(N_\theta^n)$$

with $X_{0,\theta}^n=w*(\mu^n-\nu^n)$ and $N_\theta^n=V_\theta*\mu^n$ where μ^n is the jump measure of the process

$$\sum_{i=1}^{[nt]}Y_i=\sum_{i=1}^{[nt]}X_iI_{\{|X_i|>c\}},$$

and ν^n its compensator. The natural class of alternative experiments is described as the class of experiments corresponding to two-dimensional i.i.d. observations $(y_i,z_i)_{i\geq 1}$ with a fixed conditional density $\psi(z_i;\theta|y)$ and an arbitrary marginal density $\phi(y;\theta)$ (w.r.t. the measure P^y), which in the initial problem corresponds to the alternatives under which the density value on the set $\{X:|X|\geq c\}$ is fixed and on the set $\{X:|X|>c\}$ is arbitrary.

EXAMPLE 3. *Discrete time* (Cox, Wong).

Let $\{X_1=(y_1,z_1),\ldots,X_n=(y_n,z_n)\}_{n\geq 1}$ be a sequence of two-dimensional observations with a local density $f_n(x_1,\ldots,x_n;\theta)$, $\theta\in\Theta\subset R^d$ w.r.t. some probability measure P on $((R^2)^\infty,\mathcal{B}((R^2)^\infty)$)

Suppose that the full likelihood[1)] is factorized in the following manner

$$f(x_1,\ldots,x_n;\theta)=\prod_{i=1}^n\phi_0(y_i;\theta|d_i)\prod_{i=1}^n\psi(z_i;\theta|c_i),$$

where $\phi_0(y_i;\theta|d_i)$ and $\psi(z_i;\theta|c_i)$ are conditional densities with $d_i=(y_1,z_1,\ldots,y_{i-1},z_{i-1})$ and $c_i=(y_1,z_1,\ldots,y_{i-1},z_{i-1},y_i)$

According to Cox the conditional density $\phi_0(y;\theta|d_i)$ might be unknown or depend, apart from θ, on some nuisance parameter ξ. In this case Cox suggests to use the partial likelihood

$$\prod_{i=1}^n\psi(z_i;\theta|c_i)$$

instead of the full likelihood.

The local density can be written as

$$\rho_\theta^n(t)=\mathcal{E}_t(M_\theta^n)=\mathcal{E}_t(X_{\theta,n}^n)\,\mathcal{E}_t(N_\theta^n),$$

where the processes M_θ^n, $X_{0,\theta}^n$, N_θ^n, defined by the relations

$$M_\theta^n(t)=\sum_{i=1}^{[nt]}(\phi_0(y_i;\theta|d_i)\psi(z_i;\theta|c_i)-1),$$

$$X_{0,\theta}^n(t)=\sum_{i=1}^{[nt]}(\phi_0(y_i;\theta|d_i)-1),$$

$$N_\theta^n(t)=\sum_{i=1}^{[nt]}(\psi(z_i;\theta|c_i)-1)$$

are (P^n,F^n)-martingales. Here $F^n=(\mathcal{F}_t^n)$, $0\leq t\leq 1$, $\mathcal{F}_t^n=\mathcal{F}_{[nt]}$, $\mathcal{F}_n=\sigma(Y_i,z_i\leq n)$ and P^n is the restriction of the measure P to the σ-algebra $\mathcal{F}_n$. For all $n\geq 1$ and $\theta\in\Theta$ we have

$$X_{0,\theta}^n\perp N_\theta^n.$$

The partial likelihood has the form

1) Following Cox [11] we call $f_n(X_1,\ldots,X_n;\theta)$ the full likelihood.

$$\prod_{i=1}^{[nt]} \psi(z_i;\theta|c_i) = E_t(N_\theta^n).$$

Let μ^n be the jump measure of the process $\sum_{i=1}^{[nt]} y_i$ i.e. $\mu^n((0,t]\times B) = \sum_{i=1}^{[nt]} I_{\{y_i\in B\}}$, and ν^n its P^n-compensator i.e. $\nu^n((0,t]\times B) = \sum_{i=1}^{[nt]} P\{y_i\in B|\mathcal{F}_{i-1}\}$. Then we have

$$X_{0,\theta}^n = W_\theta^n * (\mu^n - \nu^n)$$

$$N_\theta^n = V_\theta^n * \mu^n,$$

where $W_\theta^n(\omega,t,y) \equiv \phi_0(y;\theta|d_{[nt]})$ and $V_\theta^n(\omega,t,y) = \psi(z_{[nt]};\theta|d_{[nt]},y) - 1$ (note that $\omega = ((y_i,z_i)_{i\geq 1})$. Hence $X_{0,\theta}^n \in K^{1,1}(\mu^n)$ and $N_\theta^n \in K^{1,2}(\mu^n)$. The class H^n of martingales determining the alternatives can be defined in the usual manner, but as above we shall restrict ourselves by natural alternatives and consider only the subclass H_1^n of H^n which consists of martingales of the form

$$X^n(t) = W_\theta^n * (\mu^n - \nu^n)$$

with $W^n(\omega,t,y) = \phi(y|d_{[nt]})$, where $\phi(y|d_i)$ is some conditional density i.e.

$$\int \phi(y|d_i) P(dy|d_i) = 1.$$

This means that alternatives are specified as measures Q_θ^n on $((R^2)^\infty, \mathcal{B}(R^2)^\infty)$ for which the densities of conditional distributions of z_i w.r.t. c_i are fixed and coincide with $\psi(z_i;\theta|c_i)$ and conditional densities of y_i w.r.t. d_i are allowed to be arbitrary. Stable M-estimators will be defined by martingales of the form

$$m_\theta^n = V_\theta^{m,n} * \mu^n$$

with $V_\theta^{m,n}(\omega,t,y) = g_n(y_1,z_n,\ldots,y_{[nt]-1}, z_{[nt]-1}, yz_{[nt]})$ where functions $g_i(y_1,z_1,\ldots,y_i,z_i)$ are chosen so that $m_\theta^n \perp H_1^n$. The necessary conditions on g_i include

$$\int g_i(y_1,z_1,\ldots,y_i,z)\, P(dz/c_i) = 0 \text{ (for all } c_i)$$

$$\int\int \phi(y/d_i) |g(c_i,z)|\, P(dy/d_i) P(dz/c_i) < \infty,$$

for all c_i, d_i, ϕ etc.

EXAMPLE 4. *Multivariate Point process.*

Let $(\Omega,\mathcal{F},F,P)$ be a probability space with filtration satisfying the usual condition.

A multivariate point process on $((R_+\times E), \mathcal{B}(R_+)\times\mathcal{B}(E))$ (E is some Luzin space) is assumed to be a sequence $(T_n,X_n)_{n\geq 1}$, where T_n are Markov moments such that $T_1>0$, $T_n<T_{n+1}$ on $[T_n<\infty]$, while X_n are random elements with the following property: $X_n\in E$ on $(T_n<\infty)$, $X_n=\delta$ on $[T_n=\infty]$ where δ is a "marginal point", $\delta\notin E$, and $\{X_n\in C\}\in\mathcal{F}_{T_n}$, $\forall C\in\mathcal{B}(E)$, $n\geq 1$.

The multivariate point process is complitely specified by the integer-valued random measure

$$\mu(\omega;dt,dx) = \sum_{n\geq 1} I_{\{T_n<\infty\}}\, \epsilon_{(T_n,X_n)}(dt,dx),$$

where $\epsilon_{(T_n,X_n)}(dt,dx)$ is Diraque's measure. It is well known [1,2] that the random measure μ has the compensator ν w.r.t. (P,F) (a version can be chosen such that $\tilde{\nu}\{\omega,\{t\},E\}\leq 1$).

Let $G = (G_t)_{t\geq 0}$ be the natural flow of σ-algebras corresponding to the measure μ.

Let $\mathbf{F} = \{\mathcal{F}_0 V G_t\}_{t\geq 0}$ and assume that $\mathbf{F}$ satisfies usual conditions.

Consider the stochastic basis $(\Omega,F,\mathbf{F},P)$. In this case the compensator ν of the random measure μ w.r.t. $(P,\mathbf{F})$ can be explicitly defined, and it will uniquely define the measure P in the sense that if $\bar{P}$ is some other measure and $\bar{\nu}$ is the $(\bar{P},F)$-compensator of μ, then

$$\bar{P}_0 = P_0,\ \nu = \bar{\nu} \Rightarrow \bar{P} = P.$$

For every $\theta\in\Theta\subset R^d$ on $(\Omega,F,\mathbf{F})$ let a measure $P_\theta \overset{loc}{\sim} P$ be defined, and let ν_θ be the P_θ-compensator of the measure μ.

We will assume below that

1) $(T_n<\infty)=\Omega \quad \forall n\geqslant 1$,

2) $\nu(\{t\},E)=\nu_\theta(\{t\},E)=0$.

Assumption 1) implies that for any $t>0$ we have

$$\mu(t)=\mu(\omega;(0,t]\times E)=\int_0^t\int_E \mu(\omega,ds,dx)<\infty,$$

which in turn implies that for all $\theta\in\Theta$ and $t>0$

$$\nu_\theta(t)=\nu_\theta(\omega;(0,t]\times E)<\infty,\quad \forall\omega,$$

$$\nu(t)=\nu(\omega;(0,t]\times E)<\infty.$$

Note that $\mu(\omega;dt)$ is a counting process

$$\mu(\omega;(0,t]\times E)=\sum_n I_{\{T_n<t\}},$$

and $\nu_\theta(t)$ and $\nu(t)$ are its compensators w.r.t. P_θ and P respectively.

From the local equivalence of the measures P_θ and P it follows that $\nu_\theta\sim\nu$ with

$$\frac{\nu_\theta(ds,dx)}{\nu(ds,dx)}=\lambda(s,x,\theta)>0$$

(the argument ω is omitted), and

$$\rho_\theta(t)=\frac{dP_\theta}{dP}=\mathcal{E}(M_\theta), \tag{6.7}$$

where M_θ is a P-local martingale $(\forall\theta\in\Theta)$[1)]

$$M_\theta(t)=\int_0^t\int_E(\lambda(s,x,\theta)-1)(\mu-\nu)(ds,dx)$$

by assumption 2).

In this example we give only the multiplicative decomposition of the exponential (or, which is the same, the decomposition of the martingale M_θ) and specify the measure μ (with respect to which the classes $K^{1,i}(\mu)$, $i=1,2,3$, are introduced).

We begin with the decomposition of the martingale M_θ. The measures $\nu_\theta(ds,dx)$ and $\nu(ds,dx)$ can be written in the factorized form

$$\nu_\theta(ds,dx)=\nu_\theta(ds)\,q_s(dx,\theta),$$

$$\nu(ds,dx)=\nu(ds)\,q_s(dx),$$

where, roughly speaking, $q_s(dx)$ is a conditional distribution of the jump under the condition that the jump take place at the moment s.

Denote

$$\lambda_\theta^0(t)=\frac{\nu_\theta(dt)}{\nu(dt)},$$

1) We assume below that all martingales considered belong to $H^1(\mathbf{F},P)$.

$$g_t(x,\theta)=\frac{q_t(dx,\theta)}{q_t(dx)}$$

Then

$$\lambda(t,x,\theta)=\lambda_\theta^0(t)\,g_t(x,\theta)$$

This results in the following decomposition of the martingale M_θ:

$$M_\theta(t)=\int_0^t\int_E(\lambda(s,x,\theta)-1)(\mu-\nu)(ds,dx)=\int_0^t\int_E(\lambda_\theta^0(s)-1)(\mu-\nu)(ds,dx)$$

$$+\int_0^t\int_E(g_s(x,\theta)-1)(\mu-\nu)(ds,dx)+\int_0^t\int_E(\lambda_\theta^0(s)-1)(g_s(x,\theta)-1)(\mu-\nu)(ds,dx)$$

The first term can be rewritten in the following way

$$X_{0,\theta}(t)=\int_0^t\int_E(\lambda_\theta^0(s)-1)(\mu-\nu)(ds,dx)=\int_0^t(\lambda_\theta^0(s)-1)(\mu(ds)-\nu(ds)).$$

As for the second term, with the equality

$$\int_E(g_t(x,\theta)-1)q_t(dx)=0,\quad \forall t>0 \tag{6.8}$$

in mind, it can be written as

$$N_\theta(t)=\int_0^t\int_E(g_s(x,\theta)-1)\,\mu(ds,dx)$$

The martingale N_θ can be written also as a stochastic integral w.r.t. the measure $\mu(\omega;dt)$

$$N_\theta(t)=\int_0^t(g_s(\beta_s,\theta)-1)\,\mu(\omega,ds),$$

where $\beta_s(\omega)\in E$ is such that

$$\mu(\omega;\{s\},dx)=\epsilon_{\beta_s(\omega)}(dx).$$

Indeed,

$$\Delta N_\theta(t)=(g_t(\beta_t,\theta)-1)I_D$$

and, besides,

$$\Delta_t\Big(\int_0^{\cdot}(g_s(\beta_s,\theta)-1)\,\mu(\omega,ds)\Big)=(g_t(\beta_t)-1)(\mu(\omega,\{t\},E)=(g_t(\beta_t)-1)I_D.$$

Therefore, the martingales N_θ and $X_{0,\theta}$ are orthogonal since

$$[X_{0,\theta},N_\theta]=\int_0^t\int_E(\lambda_\theta^0(s)-1)(g_s(x,\theta)-1)\mu(ds,dx)=\int_0^t\int_E(\lambda_\theta^0(s)-1)(g_s(x,\theta)-1)(\mu-\nu)(ds,dx)$$

is a P-martingale. (The last equality is true by virtue of (6.8)).

Hence, we obtain the decomposition

$$M_\theta=X_{0,\theta}+N_\theta+[X_{0,\theta},N_\theta]$$

with orthogonal N_θ and $X_{0,\theta}$, and the martingales can be written as

$$X_{0,\theta}=\int_0^t\int_E(\lambda_\theta^0(s)-1)(\mu-\nu)(ds,dx),$$

$$N_\theta = \int_0^t \int_E (g_s(x,\theta)-1)\mu(ds,dx),$$

or, alternatively,

$$X_{0,\theta} = \int_0^t (\lambda_\theta^0(s)-1)(\mu(\omega,ds)-\nu(\omega,ds)),$$

$$N_\theta = \int_0^t (g_s(\beta_s,\theta)-1)\mu(\omega,ds).$$

It is evident now how to introduce the classes $K^{1,i}(\mu)$ (with two different $\mu(dt,dx)$ and $\mu(dt)$) H_0,H etc.

Observe that the classes of natural alternatives are obtained by choosing all possible $\lambda_\theta(s)$ with fixed $g_s(x,\theta)$ (certainly $\lambda_\theta(s)$ defines the martingale

$$X_\theta(t) = \int_0^t (\lambda_s(\theta)-1)(\mu-\nu)(ds,dx)$$

such that $E_t(M_\theta^X)=1$, $M_\theta^X = X_\theta + N_\theta + [X_\theta,N_\theta]$).

K. Dzhaparidze [13] considered the special case of $E=\{1,2,...,r\}$ and $N_t=\{N_t^1,\ldots,N_t^r\}$ with $N_t^i=\mu(\omega:(0,t],i)$ which is called the r-variate counting process. It has the P-compensator $A_t=\{A_t^1,\ldots,A_t^r\}$ with $A_t^i=\nu((0,t],i)$, and the P_θ-compensator $A_t(\theta)=(A_t^1(\theta),\ldots,A_t^r(\theta)\}$ with $A_t^i(\theta)=\nu_\theta((0,t],i)$. In this case

$$\lambda_s^0(\theta) = \frac{\nu_\theta(dt,E)}{\nu(dt,E)} = \frac{d\bar{A}_t(\theta)}{d\bar{A}_t}$$

where

$$\bar{A}_t(\theta) = \sum_{i=1}^r A_t^i(\theta), \quad \bar{A}_t = \sum_{r=1}^n A_t^i(\theta)$$

and

$$g_s(i,\theta) = \frac{d\bar{A}_t^i(\theta)}{dA_t(\theta)} \frac{d\bar{A}_t}{dA_t^i}.$$

The martingale N_θ in this case has the form

$$N_\theta = \int_0^t (g_s(x,\theta)-1)\,\mu(ds,dx) = \sum_{i=1}^r \int_0^t (g_s(i,\theta)-1)dN_t^i.$$

EXAMPLE 5. *Two-dimensional diffusion.*

Let $\xi=(\eta,\zeta)$ be a two-dimentional diffusion type process satisfying the following stochastic differential equations

$$d\eta_t = \phi^0(t,\eta,\xi,\theta)dt + dW_t^1, \tag{6.9}$$

$$d\xi_t = \psi(t,\eta,\xi,\theta)dt + dW_t^2, \ \zeta_0=\eta_0=0$$

where W^1 and W^2 are independent Wiener processes and coefficients ϕ^0 and ψ are such that there exists the unique weak solution of equation (6.9) and $P_\theta^{\eta,\zeta} \overset{loc}{\sim} P^{w^1,w^2}$. Then

$$\frac{dP_\theta^{\eta,\zeta}}{dP^{w^1,w^2}} = \mathcal{E}(M_\theta)$$

where

$$M_\theta(t)=\int_0^t \phi^0(s,y,z,\theta)dy_s+\int_0^t \psi(s,y,z,\theta)dz_s.$$

Note that measures $P_\theta^{y,z}$ and P^{w^1,w^2} are defined on $(C^2_{[0,T]},\mathcal{B}(C^2_{[0,T]})$ with filtration $\mathcal{F}_t=\{y,z:y_s,z_s,s\leq t\}$ and $y=(y_s)_{s\geq 0},z=(z_s)_{s\geq 0}$ are coordinate processes.

Denote $X_{0,\theta}(t)=\int_0^t \phi^0(s,y,z,\theta)ds_s$ and $N_\theta(t)=\int_0^t \psi(s,y,z,\theta)dz_s$. The processes $X_{0,\theta}$ and N_θ are P^{w^1,w^2} martingales and $\langle X_{0,\theta},N_\theta\rangle=0$. Hence

$$\frac{dP_\theta^{\eta,\zeta}}{dP^{w^1,w^2}}=\mathcal{E}(M_\theta)=\mathcal{E}(X_{0,\theta})\,\mathcal{E}(N_\theta).$$

In this case the classes of natural alternatives are given by means of variation of ϕ with given ψ and the stable estimators with respect to this class are constructed by means of P_θ-martingales of the following form

$$m_\theta(t)=\int_0^t g(s,y,z,\theta)dz_s.$$

Example 6. *Censored Diffusion.*

Let ξ be a diffusion process with the differential

$$d\xi(t)=I_{\{|\xi_t|>c\}}\phi_0(t,\xi_t,\theta)dt+I_{\{|\xi_t|\leq c\}}\psi(t,\xi_t,\theta)dt+dw_t \tag{6.10}$$

where functionals $\phi_0(t,x,\theta)$ and $\psi(t,x,\theta)$ are such that for every $\theta\in\Theta\subset R^d$ the unique weak solution of this equation exists. Let P_θ^ξ be the distribution of ξ with given θ. Suppose that $P_\theta^\xi\overset{loc}{\sim}P^w$.

Then

$$\frac{dP^\theta}{dP}=\mathcal{E}(M_\theta)=\mathcal{E}(X_{0,\theta})\,\mathcal{E}(N_\theta)$$

where

$$M_\theta=X_{0,\theta}+N_\theta,$$

$$X_{0,\theta}(t)=\int_0^t I_{\{|x_s|>c\}}\phi_0(s,x_s,\theta)dx_s,$$

$$N_\theta(t)=\int_0^t I_{\{|x_s|\leq c\}}\,\psi(s,x_s,\theta)dx_s$$

and $\langle X_{0,\theta},N_\theta\rangle=0$.

Evidently the class of martingales defining alternatives consists of martingales of the following form

$$X(t)=\int_0^t I_{\{|x_s|>c\}}\,\phi(s,x_s,\theta)dx_s,$$

and the stable estimates are obtained by means of m_θ martingales of the form

$$m_\theta(t)=\int_0^t I\{|x_s|\leq c\}\,g(s,x_s,\theta)dx_s.$$

7. Regularity of M-estimator

Recall, that the estimator $\tilde\theta=(\tilde\theta_n)_{n\geq 1}$ is called regular at θ, if for some nondegenerate distribution $F(x)$, $x\in R^d$ we have the convergence

$$\mathcal{L}(c_n^{-1}(\tilde\theta_n-\theta)-u\,|\,P^n_{\theta+c_nu})\Rightarrow F$$

for any $u\in R^d$ such that $\theta+c_nu\in\Theta$.

We will show that the estimator $\hat\theta=(\hat\theta_n)_{n\geq 1}$ constructed in Theorem 5.1 is regular.

THEOREM 7.1. *Let the conditions of Theorems 3.1 and 5.1 hold. Then for any $\theta\in\Theta$ there exists M-estimator $\hat\theta=(\hat\theta_n)_{n\geq1}$ regular at θ with $F=N(0,\gamma^{-1}(\theta)\Gamma(\theta)(\gamma^{-1}(\theta))')$.*

PROOF. Denote $L_n(\theta)=L(m^n_\theta,M^n_\theta)$. By Taylor's formula we have

$$L_n(\theta+c_nu)=L_n(\theta)+c_n(\dot L_n(\theta)u)+c_n(\dot L_n(\bar\theta)-\dot L_n(\theta))u$$

where $\bar\theta=\theta+\alpha(\theta)c_nu$, $\alpha(\theta)\in[0,1]$ and $\dot L_n(\theta)=(\frac{\partial}{\partial\theta_i}L^j_n(\theta))_{i,j=\overline{1,d}}$.

Consequently

$$c^2_nL_n(\theta+c_nu)=c^2_nL_n(\theta)-\gamma(\theta)u+\epsilon_n(\bar\theta,\theta)u \tag{7.1}$$

where

$$\epsilon_n(y,\theta)=\{c^2_n(\dot L_n(y)-\dot L_n(\theta))+(c^2_n\dot L_n(\theta)-\gamma(\theta))\},\ y\in\Theta.$$

By conditions (b) and (e) of Theorem 5.1 we have

$$\lim_{r\to0}\overline{\lim_{n\to\infty}}\,P^n_\theta\{\sup_{y:\|y-\theta\|<r}\|\epsilon_n(y,\theta)\|>\rho\}=0,\ \text{for }\forall\rho>0 \tag{7.2}$$

By (7.1)

$$c_nL_n(\hat\theta_n)=c_nL_n(\theta)-\gamma(\theta)c_n^{-1}(\hat\theta_n-\theta)+\epsilon_n(\bar\theta_n,\theta)c_n^{-1}(\hat\theta_n-\theta) \tag{7.3}$$

where $(\hat\theta_n)_{n\geq1}$ is the estimator constructed in Theorem 5.1 and $\bar\theta_n=\theta+\alpha(\hat\theta_n-\theta)$ with $\alpha\in[0,1]$.

Now, by virtue of property II of $(\hat\theta_n)_{n\geq1}$ (see assertion of Theorem 5.1) and (7.2) we have

$$\overline{\lim_{n\to\infty}}\,P^n_\theta\{\|\epsilon_n(\bar\theta_n,\theta)\|>\rho\}=0,\ \text{ for }\forall\rho>0. \tag{7.4}$$

Indeed

$$\{\|\hat\theta_n-\theta\|\leq r\}\cap\{\sup_{y:\|y-\theta\|<r}\|\epsilon_n(y,\theta)\|\leq\rho\}\subset\{\epsilon_n(\bar\theta_n,\theta)\|\leq\rho\}.$$

Therefore

$$\overline{\lim_{n\to\infty}}P^n_\theta\{\|\epsilon_n(\bar\theta_n,\theta)\|>\rho\}\leq\overline{\lim_{n\to\infty}}P^n_\theta\{\sup_{y:\|y-\theta\|<r}\|\epsilon_n(y,\theta)\|>\rho\}\to0 \text{ as } r\to0.$$

Now by (7.3) we have

$$\|c_nL_n(\hat\theta_n)-c_nL_n(\theta)+\gamma(\theta)c_n^{-1}(\hat\theta_n-\theta)\|\leq\|\epsilon_n(\bar\theta_n,\theta)\gamma^{-1}(\theta)\|\cdot\|\gamma(\theta)c_n^{-1}(\hat\theta_n-\theta)\|$$

which easily leads to an inequality

$$\begin{aligned}&\|c_nL_n(\hat\theta_n)-c_nL_n(\theta)-c^2_n\langle L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,M^n_\theta),u)\rangle-(\gamma(\theta)c_n^{-1}(\hat\theta_n-\theta)-c^2_n\langle L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,M^n_\theta),u)\rangle\|\\&\leq\|\epsilon_n(\bar\theta_n,\theta)\gamma^{-1}(\theta)\|\cdot\|\gamma(\theta)c_n^{-1}(\hat\theta_n-\theta)-c^2_n\langle L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,M^n_\theta),u)\rangle\|\\&+\|\epsilon_n(\bar\theta_n,\theta)\gamma^{-1}(\theta)\|c^2_n\|\langle L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,M^n_\theta),u)\rangle\|\end{aligned} \tag{7.5}$$

Introduce the notations: $X^n=c_nL_n(\theta)$,

$$Y^n=c_nL_n(\hat\theta_n)-(X^n+c^2_n\langle L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,m^n_\theta),u)\rangle),$$

$$V^n=\gamma(\theta)c_n^{-1}(\hat\theta_n-\theta)-c^2_n<L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,M^n_\theta),u)>$$

$$Z^n=\|\epsilon_n(\bar\theta_n,\theta)\gamma^{-1}(\theta)\|,\ \beta^n=c^2_n\|\langle L(m^n_\theta,M^n_\theta),(L(\dot M^n_\theta,M^n_\theta),u)\rangle\|.$$

Rewrite (7.5) in this terms. We have

$$\|Y^n-V^n\|\leq z^n\|V^n\|+z^n\beta^n \tag{7.6}$$

and by (7.4)

$$P_\theta^n - \lim_{n\to\infty} z^n = 0. \tag{7.7}$$

By virtue of conditions (b) of Theorem 5.1, the sequence $(\beta^n)_{n\geq 1}$ is bounded w.r.t. P_θ^n-probability and, hence,

$$P_\theta^n - \lim_{n\to\infty} z^n\beta^n = 0. \tag{7.8}$$

Consider now the sequence of measures $(P_{\theta+c_nu}^n)$. By Theorem 1.6, [3], Sec V, and LAN property (see Theorem 3.1) it follows that for any u

$$(P_{\theta+c_nu}^n) \lhd (P_\theta^n)$$

Consequently, (7.7) and (7.8) imply

$$P_{\theta+c_nu}^n - \lim_{n\to\infty} z^n = 0, \tag{7.9}$$

and

$$P_{\theta+c_nu} - \lim_{n\to\infty} z^n\beta^n = 0. \tag{7.10}$$

Now, as in the proof of Lemma 4.1, section 4°, we have from (7.6), in view of (7.9) and (7.10), that

$$\lim_{n\to\infty} \mathcal{L}\{\gamma(\theta)c_n^{-1}(\hat{\theta}_n-\theta)-c_n^2\langle L(m_\theta^n,M_\theta^n),(L(\dot{M}_\theta^n,M_\theta^n),u)\rangle|P_{\theta+c_nu}^n\}$$
$$= \lim_{n\to\infty} \mathcal{L}\{c_nL_n(\hat{\theta}_n)-c_nL_n(\theta)+c_n^2\langle L(m_\theta^n,M_\theta^n),(L(\dot{M}_\theta^n,M_\theta^n),u)\rangle|P_{\theta+c_nu}^n\}.$$

Note that

$$P_{\theta+c_nu}^n - \lim c_n^2\langle L(m_\theta^n,M_h^n),(L(\dot{M}_\theta^n,M_\theta^n),u)\rangle = \gamma(\theta)u,$$

and

$$P_{\theta+c_nu}^n - \lim_{n\to\infty} c_nL_n(\hat{\theta}_n) = 0.$$

Therefore

$$\lim_{n\to\infty} \mathcal{L}\{\gamma(\theta)(c_n^{-1}(\hat{\theta}_n-\theta)-u)|P_{\theta+c_nu}^n\}$$
$$= \lim_{n\to\infty} \mathcal{L}\{-c_nL_n(\theta)+c_n^2\langle L(m_\theta^n,M_\theta^n),(L(\dot{M}_\theta^n,M_\theta^n),u)\rangle|P_{\theta+c_nu}^n\}.$$

It remains to show that

$$\lim_{n\geq\infty} \mathcal{L}\{c_nL_n(\theta)-c_n^2\langle L(m_\theta^n,M_\theta^n),(L(\dot{M}_\theta^n,M_\theta^n),u)\rangle|P_{\theta+c_nu}^n\} = N(0,\Gamma(\theta)).$$

We will apply Theorem 3.9, section X in [3], to the case in which

$$\tilde{P}^n = P_{\theta+c_nu}^n,\ P^n = P_\theta^n,\ X_n = c_nL_n(\theta),\ M = \Gamma^{1/2}(\theta)\frac{1}{\sqrt{T}}W,$$

W is an d-dimensional Wiener process and $D=\{T\}$.

The process X_n is a square integrable martingale w.r.t. the measure P_θ^n with the triplet

$$(B_n = -I_{\{\|x\|>1\}}x * \nu_{X_n}^{P_\theta^n},\ C_n = c_n^2\langle m_\theta^{n,c}\rangle,\ \nu_n = \nu_{X_n}^{P_\theta^n})$$

so that by our assumptions the conditions $(\hat{\delta}_5 - D)$ and $(\gamma_5 - D)$ in [3], section X.3.7, are satisfied.

Therefore

$$X_n - B_n \xrightarrow{\mathcal{L}(\{T\}|P_\theta^n)} \Gamma^{1/2}(\theta)W/\sqrt{T},$$

and hence by Theorem X.3.9, assertion a) in [3],

$$X_n-\tilde{B}_n \xrightarrow{\mathcal{L}(\{T\}|P^n_{\theta+c_nu})} \Gamma^{1/2}(\theta)W/\sqrt{T}$$

where $\tilde{B}_n$ is the first characteristic of the semimartingale X_n w.r.t. the measure $P^n_{\theta+c_nu}$.

We must show now that

$$P^n_{\theta+c_nu}-\lim_{n\to\infty}|\tilde{B}_n(T)-c_n^2\langle L(m^n_\theta,M^n_\theta),(L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T|=0.$$

By the transformation formula for triplets under the absolutely continuous change of a measure with

$$dP^n_{\theta+c_nu}=\mathcal{E}(L(M^n_{\theta+c_nu}-M^n_\theta,M^n_\theta))dP^n_\theta$$

we obtain that

$$\tilde{B}_n=B_n-\langle\overline{M}^n,L(M^n_{\theta+c_nu}-M^n_\theta,M^n_\theta)\rangle^{P,P^n_\theta} \tag{7.12}$$

where

$$\overline{M}^n=X_n-xI_{\{\|x\|>1\}}*(\mu_{X_n}-\nu^{P^n_\theta}_{X_n}).$$

Consequently

$$\tilde{B}_n(T)-c_n^2\langle L(m^n_\theta,M^n_\theta),(L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T$$
$$=B_n(T)+c_n\langle X_n-\overline{M}^n,(L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T+c_n\langle\overline{M}^n,(L(\dot{M}^n_y,M^n_\theta)-L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T$$

where $y=\theta+\alpha c_nu$, $\alpha\in[0,1]$.

By Lindeberg's condition and contiguity $(P_{\theta+c_nu^\cdot})\lhd(P^n_\theta)$ we have

$$P^n_{\theta+c_nu}-\lim_{n\to\infty}B_n(T)=0.$$

We show now that

$$P^n_{\theta+c_nu}-\lim c_n\langle X_n-\overline{M}^n,(L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T=0. \tag{7.13}$$

Indeed, (7.12) implies

$$\langle X_n-\overline{M}^n\rangle_T=xx'I_{\{\|x\|>1\}}*V^{P^n_\theta}_{X_n}(T)-\sum_{s\leq T}(\textstyle\int xI_{\{\|x\|>1\}}\nu^{P^n_\theta}_{X_n}(\{s\},dx))(\int xI_{\{\|x\|>1\}}\nu^{P^n_\theta}_{X_n}(\{s\},dx))'.$$

Again by Lindeberg's condition and contiguity $(P^n_{\theta+c_nu})\lhd(P^n_\theta)$ we have

$$P^n_{\theta+c_nu}-\lim_{n\to\infty}\langle X_n-\overline{M}^n\rangle_T=0. \tag{7.14}$$

Consequently, the Kunita-Watanabe inequality gives (7.13).

It remains to show that

$$P^n_{\theta+c_nu}-\lim_{n\to\infty}c_n\langle\overline{M}^n,(L(\dot{M}^n_y,M^n_\theta)-L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T=0.$$

We have

$$c_n^2|\langle\overline{M}^n,(L(\dot{M}^n_y,M^n_\theta)-L(\dot{M}^n_\theta,M^n_\theta),u)\rangle_T|^2\leq 2c_n^2|\langle -X_n+\overline{M}^n,(L(\dot{M}^n_y-\dot{M}^n_\theta,M^n_\theta),u)\rangle_T|^2$$
$$+2c_n^2|\langle X_n,(L(\dot{M}^n_y-\dot{M}^n_\theta,M^n_\theta),u)\rangle_T|^2$$

By condition (c) of Theorem 3.1 and the contiguity $(P^n_{\theta+c_nu})\lhd(P^n_\theta)$ we have

$$P^n_{\theta+c_nu}-\lim_{n\to\infty}c_n^2\langle(L(\dot{M}^n_y-\dot{M}^n_\theta,M^n_\theta),u)\rangle_T=0.$$

On the other hand, by (7.14) and condition (a) of Theorem 5.1, we arrive at the desired result by using the Kunita-Watanabe inequality. Hence

$$\lim_{n\to\infty} \mathfrak{L}\{\gamma(\theta)(c_n^{-1}(\hat{\theta}_n-\theta)-u)|P^n_{\theta+c_n u}\}=N(0,\Gamma(\theta)).$$

Consequently

$$\lim_{n\to\infty} \mathfrak{L}(c_n^{-1}(\hat{\theta}_n-\theta)-u|P^n_{\theta+c_n u}\}=N(0,\gamma^{-1}(\theta)\Gamma(\theta)(\gamma^{-1}(\theta))').$$ □

8. One-step approximation technique

We will consider regular ergodic experiments and M-estimators defined by martingales $(m^n_\theta)_{n\geq 1}$.

For convenience introduce the notations:

$$l_i^n(\theta)=L(\frac{\partial}{\partial\theta_i}M^n_\theta,M^n_\theta),\ i=\overline{1,d},\ h_i^n(\theta)=L(m^{n,i}_\theta,M^n_\theta),i=\overline{1,d},$$

$$\hat{I}(\theta)=(\hat{I}^n_{ij}(\theta))=(\langle l_i^n(\theta),l_j^n(\theta)\rangle),$$

$$\hat{H}^n(\theta)=(\hat{H}^n_{ij}(\theta))=(\langle h_i^n(\theta),h_j^n(\theta)\rangle),$$

$$\hat{A}^n(\theta)=(\hat{A}^n_{ij}(\theta))=(\langle h_i^n(\theta),l_j^n(\theta)\rangle),\ i,j=\overline{1,d}.$$

The ergodicity conditions imposed on the matrices $\hat{I}^n(\theta)$, $\hat{H}^n(\theta)$ and $\hat{A}^n(\theta)$ may be written as follows:

$$c_n^2\hat{I}^n_T(\theta)\xrightarrow{P^n_\theta}I(\theta),$$

$$c_n^2\hat{H}^n_T(\theta)\xrightarrow{P^n_\theta}\Gamma(\theta),$$

$$c_n^2\hat{A}^n_T(\theta)\xrightarrow{P^n_\theta}\gamma(\theta),$$

where the matrices $I(\theta)$, $\Gamma(\theta)$ and $\gamma(\theta)$ are defined by relations (i) in section 2.4, and (a) and (b) in Theorem 5.1. Assume that $\gamma(\theta)$ is θ-continuous.

Let $\bar{\theta}=(\bar{\theta}_n)_{n\geq 1}$ be some c_n-consistent estimator of the parameter θ, i.e.

$$c_n^{-1}(\bar{\theta}_n-\theta)=O_{P^n_\theta}(1).$$

Define the estimator $\hat{\theta}=(\hat{\theta}_n)_{n\geq 1}$ by the relation

$$\hat{\theta}_n=\bar{\theta}_n+c_n^2\gamma^{-1}(\bar{\theta}_n)h^n(\bar{\theta}_n)$$

The last relation is the one-step modification of Newton's method for successive approximation to the solution of

$$h^n(\theta)=0.$$

Theorem 8.1. *Let conditions of Theorem 5.1 hold. Then*

$$P^n_\theta-\lim_{n\to\infty}\hat{\theta}_n=\theta,$$

$$\mathfrak{L}(c_n^{-1}(\hat{\theta}_n-\theta)|P^n_\theta)\Rightarrow N(0,\gamma^{-1}(\theta)\Gamma(\theta)(\gamma^{-1}(\theta))').$$

Proof. By Taylor's formula we have

$$\begin{aligned}(\hat{\theta}_n-\theta)&=(\bar{\theta}_n-\theta)+c_n^2\gamma^{-1}(\bar{\theta}_n)h^n(\bar{\theta}_n)=(\bar{\theta}_n-\theta)+c_n^2\gamma^{-1}(\theta)h^n(\bar{\theta}_n)+c_n^2(\gamma^{-1}(\bar{\theta}_n)-\gamma^{-1}(\theta))h^n(\bar{\theta}_n)\\&=(\bar{\theta}_n-\theta)+c_n^2\gamma^{-1}(\theta)h^n(\theta)+\gamma^{-1}(\theta)c_n^2\dot{h}^n(\theta)(\bar{\theta}_n-\theta)+\gamma^{-1}(\theta)[c_n^2(\dot{h}^n(y_n)-\dot{h}^n(\theta))](\bar{\theta}_n-\theta)\\&\quad+c_n^2(\gamma^{-1}(\bar{\theta}_n)-\gamma^{-1}(\theta))h^n(\bar{\theta}_n)=c_n^2\gamma^{-1}(\theta)h^n(\theta)+\epsilon_n(y_n,\theta)(\bar{\theta}_n-\theta)+\delta_n(\bar{\theta}_n,\theta),\end{aligned}$$

where

$$\dot{h}^n(\theta)=(\frac{\partial}{\partial\theta_i}h_j^n(\theta)), i,j=\overline{1,d},\ y_n=\theta+\alpha(\overline{\theta}_n-\theta),\ \alpha\in[0,1],$$

$$\epsilon_n(y_n,\theta)=\gamma^{-1}(\theta)[c_n^2(-\dot{h}^n(\theta)+\dot{h}^n(y_n))+(c_n^2\dot{h}^n(\theta)-\gamma(\theta))],$$

$$\delta_n(\overline{\theta}_n,\theta)=c_n^2(\gamma^{-1}(\overline{\theta}_n)-\gamma^{-1}(\theta))h^n(\overline{\theta}_n).$$

It can be easily seen that

$$\overline{\lim_{n\to\infty}}\ P_\theta^n\{\|\epsilon_n(y_n,\theta)\|>\rho\}=0,\ \ \forall\rho>0.$$

Now we will show that

$$\overline{\lim_{n\to\infty}}\ P_\theta^n\{c_n^{-1}\|\delta_n(\overline{\theta}_n,\theta)\|>\rho\}=0,\ \ \forall\rho>0. \tag{8.1}$$

We have

$$c_n^{-1}\delta_n(\overline{\theta}_n,\theta)=c_n(\gamma^{-1}(\overline{\theta}_n)-\gamma^{-1}(\theta))h^n(\theta)+(\gamma^{-1}(\overline{\theta}_n)-\gamma^{-1}(\theta))c_n^2\dot{h}^n(\theta)c_n^{-1}(\overline{\theta}_n-\theta)$$
$$+(\gamma^{-1}(\overline{\theta}_n)-\gamma^{-1}(\theta))c_n^2(\dot{h}^n(\overline{\theta}_n)-\dot{h}^n(\theta))c_n^{-1}(\overline{\theta}_n-\theta). \tag{8.2}$$

By conditions (a) and (d) in Theorem 5.1 it follows that $\mathfrak{L}(c_nh^n(\theta)|P_\theta^n)\Rightarrow N(0,\Gamma(\theta))$. Therefore the sequence $c_nh^n(\theta)$ is bounded in probability and the matrix $\gamma^{-1}(\theta)$ is θ-continuous. So the first term of the last relation converges to 0. Further, by virtue of conditions (b) and (c) of Theorem 5.1 $c_n^2\dot{h}^n(\theta)\xrightarrow{P_\theta^n}\gamma(\theta)$, and, since the sequence $c_n^{-1}(\overline{\theta}_n-\theta)$ is bounded in probability, the second term in (8.2) tends to the zero. Finally, the third term converges to zero by the condition (e) of Theorem 5.1.

Now

$$\hat{\theta}_n\xrightarrow{P_\theta^n}\theta,$$

since by condition (a) of Theorem 5.1

$$c_n^2h^n(\theta)\xrightarrow{P_\theta^n}0,$$

and finally

$$\lim_{n\to\infty}\mathfrak{L}(c_n^{-1}(\hat{\theta}_n-\theta)|P_\theta^n)=\lim_{n\to\infty}\mathfrak{L}(c_n\gamma^{-1}(\theta)h^n(\theta)|P_\theta^n)=N(0,\gamma^{-1}(\theta)\Gamma(\theta)(\gamma^{-1}(\theta))'). \qquad \square$$

9. Finite-dimensional nuisanse parameter

For simplicity we will consider the case of a two-dimensional parameter (θ,η), with a parameter of interest θ and a nuisance parameter η. By "$\star$" and "$\cdot$" we will denote partial derivatives w.r.t. θ and η respectively. Denote

$$I_1^n(\theta,\eta)=L(\overset{\star}{M}{}_{\theta,\eta}^n,M_{\theta,\eta}^n),\ I_2^n(\theta,\eta)=L(\dot{M}_{\theta,\eta}^n,M_{\theta,\eta}^n),$$

$$h_i^n(\theta,\eta)=L(m_{\theta,\eta}^{n,i},M_{\theta,\eta}^n), i=1,2.$$

As above, the experiment will be assumed to be regular and ergodic.

9.1 *Pseudo M-estimators.*

Let $N_{\theta,\eta}^n$ be a one-dimentional martingale. Assume that $\hat{\eta}=(\hat{\eta}_n)_{n\geq1}$ is some estimator of the parameter η with the property

$$\mathfrak{L}(c_nL(N_{\theta,\eta}^n,M_{\theta,\eta}^n),\ c_n^{-1}(\hat{\eta}_n-\eta)|P_{\theta,\eta}^n)\Rightarrow N(0,\Sigma) \tag{9.1}$$

where Σ is a symmetric positive defined matrix. Consider the equation w.r.t. θ

$$h^n(\theta,\hat{\eta}_n)=0$$

where $h^n(\theta,\eta)=L(N_{\theta,\eta}^n,M_{\theta,\eta}^n)$.

THEOREM 9.1. *Let for any θ,η the following conditions be satisfied:*

(b[1]) $$P^n_{\theta,\eta}-\lim_{n\to\infty} c_n^2[L(N^n_{\theta,\eta},M^n_{\theta,\eta}),\ L(\overset{*}{M}{}^n_{\theta,\eta},M^n_{\theta,\eta})]=\sigma_{11}(\theta,\eta),$$
$$P^n_{\theta,\eta}-\lim_{n\to\infty} c_n^2[L(N^n_{\theta,\eta},M^n_{\theta,\eta}),\ L(\dot{M}^n_{\theta,\eta},M^n_{\theta,\eta})]=\sigma_{12}(\theta,\eta);$$

(c[1]) $$P^n_{\theta,\eta}-\lim_{n\to\infty} c_n^2L(\overset{*}{N}{}^n_{\theta,\eta},M^n_{\theta,\eta})=0,$$
$$P^n_{\theta,\eta}-\lim_{n\to\infty} c_n^2L(\dot{N}^n_{\theta,\eta},M^n_{\theta,\eta})=0;$$

(e[1]) $$\lim_{r\to 0}\overline{\lim_{n\to\infty}}\, P^n_{\theta,\eta}\{\sup_{x,y:|x-\theta|+|y-\eta|<r}\|\overset{\circ}{L}(N^n_{x,y},M^n_{x,y})-\overset{\circ}{L}(N^n_{\theta,\eta},M^n_{\theta,\eta})\|>\rho\}=0)$$

with $\overset{\circ}{L}=(\overset{*}{L},\dot{L})$.

Then for any θ,η there exists an asymptotically unique estimator $\tilde{\theta}=(\tilde{\theta}_n),n\geq 1$ such that

I. $\lim_{n\to\infty} P^n_{\theta,\eta}\{h^n(\tilde{\theta}_n,\hat{\eta}_n)=0\}=1,$

II. $\mathfrak{L}(c_n^{-1}(\tilde{\theta}_n-\theta)|P^n_{\theta,\eta})\Rightarrow N(0,\sigma^2),$
where
$$\sigma^2=\sigma_{11}^{-2}(\Sigma_{11}-2\Sigma_{12}\sigma_{12}+\Sigma_{22}\sigma_{12}^2).$$

PROOF. Under our assumptions the standard technique using Taylor's formula, described in details in Lemma 4.1, leads to the existence of the estimator $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ with property I. In this case we have the decomposition

$$c_nh^n(\tilde{\theta},\hat{\eta}_n)=c_nh^n(\theta,\eta)-c_n^{-1}(\tilde{\theta}_n-\theta)\sigma_{11}(\theta,\eta)-c_n^{-1}(\hat{\eta}_n-\eta)\sigma_{12}(\theta,\eta)+c_n^{-1}(\tilde{\theta}_n-\theta)\epsilon_1^n+c_n^{-1}(\hat{\eta}_n-\eta)\epsilon_2^n,$$

where ϵ_1^n and ϵ_2^n are small in the usual sense (see, e.g. the proof of Lemma 4.1, 1°).

Hence, following standard arguments, we can conclude that

$$\lim_{n\to\infty}\mathfrak{L}(c_n^{-1}(\tilde{\theta}_n-\theta)|P^n_{\theta,\eta})=\lim_{n\to\infty}\mathfrak{L}(\sigma_{11}^{-1}(\theta,\eta)(c_nh^n(\theta,\eta)-c_n^{-1}(\hat{\eta}_n-\eta)|P^n_{\theta,\eta})=N(0,\sigma^2).\qquad\square$$

REMARK. It seems interesting to construct an estimator of the parameter θ by means of an M-estimator of the nuisance parameter η, for which a non-constructive condition (9.1) can be expressed in terms of the characteristics of a martingale defining this M-estimator.

With this aim in view we will consider an M-estimator $(\hat{\theta}_n,\hat{\eta}_n),n\geq 1$ defined by a system of equations

$$h_i^n(\theta,\eta)=0,\quad i=1,2.$$

Under the assumptions of Theorem 5.1 we see that the limit distribution of the vector $c_n^{-1}\{(\hat{\theta}_n-\theta),(\hat{\eta}_n-\eta)\}'$ coincides with that of the vector

$$c_n\gamma^{-1}(\theta,\eta)\begin{bmatrix}h_1^n(\theta,\eta)\\ h_2^n(\theta,\eta)\end{bmatrix}.$$

Therefore[1])

$$\lim_{n\to\infty}\mathfrak{L}(c_nL(N^n_{\theta,\eta},M^n_{\theta,\eta}),\ c_n^{-1}(\hat{\eta}_n-\eta)|P^n_{\theta,\eta})=\lim_{n\to\infty}\mathfrak{L}(c_nL(N^n_{\theta,\eta},M^n_{\theta,\eta}),\ c_n\gamma^{21}(\theta,\eta)h_1^n(\theta,\eta)$$
$$+c_n\gamma^{22}(\theta,\eta)h_2^n(\theta,\eta)|P^n_{\theta,\eta}).$$

Suppose now that the $P^n_{\theta,\eta}$-martingale $c_nL(N^n_{\theta,\eta},M^n_{\theta,\eta})$ satisfies the following conditions:

1) $\gamma^{ij}(\theta,\eta)$, $i,j=1,2$ are elements of the matrix $\gamma^{-1}(\theta,\eta)$.

$$P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2 \langle L(N^n_{\theta,\eta}, M^n_{\theta,\eta})\rangle_T = \alpha,$$

$$P^n_{\theta,\eta} - \lim_{n\to\infty} \int_0^T \int_{|x|>\epsilon} x^2 \nu^n_L(ds,dx) = 0, \quad \epsilon\in(0,1],$$

where ν^n_L is the compensator of the jump measure of the process $c_n L(N^n_{\theta,\eta}, M^n_{\theta,\eta})$, and besides,

$$P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2 \langle L(N^n_{\theta,\eta}, M^n_{\theta,\eta}),\ h^n_i,(\theta,\nu)\rangle_T = \beta_i,\ i=1,2.$$

Then[2)]

$$P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2 \begin{bmatrix} \langle L(N_{\theta,\eta}, M^n_{\theta,\eta})\rangle & ,\langle L(N^n_{\theta,\eta}, M^n_{\theta,\eta}), \gamma^{21}(\theta,\eta)h^n_1(\theta,\eta) + \gamma^{22}(\theta,\eta)h^n_2(\theta,\eta)\rangle \\ & ,\langle \gamma^{21}(\theta,\eta)h^n_1(\theta,\eta) + \gamma^{22}(\theta,\eta)h^n_2(\theta,\eta)\rangle \end{bmatrix}$$

$$= \begin{bmatrix} \alpha, \beta_1\gamma^{21}(\theta,\eta) + \beta_2\gamma^{22}(\theta,\eta) \\ ,(\gamma^{21}(\theta,\eta))^2\Gamma_{11}(\theta,\eta) + 2\gamma^{21}(\theta,\eta)\gamma^{22}(\theta,\eta) + (\gamma^{22}(\theta,\eta))^2\Gamma_{22}(\theta,\eta) \end{bmatrix} = \Sigma,$$

and since for the vector

$$(c_n L(N^n_{\theta,\eta}, M^n_{\theta,\eta}),\ c_n(\gamma^{21}_{(\theta,\eta)} h^n_1(\theta,\eta) + \gamma^{22}(\theta,\eta)h^n_2(\theta,\eta)))'$$

Lindeberg's condition is satisfied, we have (9.1).

9.2. Skew Projection Technique

Here we use the known projection technique [15], applied for constructing an estimator of the structural parameter θ efficient in Fisher's sense in presence of a nuisance parameter η. This method allows to construct an estimator of the structural parameter θ with the same asymptotic variance as for the first component of a two-dimensional M-estimator (obtained by solving the complete system of equations defining M-estimator). For this we need projection of $h^n_1(\theta,\eta)$ onto $h^n_2(\theta,\eta)$ in the direction defined by the relation

$$P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2 \langle h^n_1(\theta,\eta) - b h^n_2(\theta,\eta),\ l^n_2(\theta,\eta)\rangle_T = 0, \tag{9.2}$$

where $b\in R^1$. Note that this direction is orthogonal to the linear space spanned by $l^n_2(\theta,\eta)$ rather then to that of spanned by $h^n_2(\theta,\eta)$.

Relation (9.2) implies that $b = \gamma_{12}(\theta,\eta)\gamma_{22}^{-1}(\theta,\eta)$. In the case when $h^n_1(\theta,\eta) = l^n_1(\theta,\eta)$, $h^n_2(\theta,\eta) = l^n_2(\theta,\eta)$ (i.e. MLE is considered), we have $b = I_{12}(\theta,\eta)I_{22}^{-1}(\theta,\eta)$ and (9.2) leads to a modified informant

$$\tilde{l}^n_1(\theta,\eta) = l^n_1(\theta,\eta) - I_{12}(\theta,\eta)I_{22}^{-1}(\theta,\eta)l^n_2(\theta,\eta).$$

Denote $\tilde{h}^n_1(\theta,\eta) = h^n_1(\theta,\eta) - \gamma_{12}(\theta,\nu)\gamma_{22}^{-1}(\theta,\eta)h^n_2(\theta,\eta)$. This is so colled modified influence martingale.

Let $\bar\eta = (\bar\eta_n)_{n\geq1}$ be a c_n-consistent estimator of η, i.e. $c_n^{-1}(\bar\eta_n - \eta) = O_{P^n_{\theta,\eta}}(1)$. Consider the equation

$$\tilde{h}^n_1(y,\bar\eta_n) = 0 \tag{9.3}$$

THEOREM 9.2. *Let the conditions of Theorems 3.1 and 5.1 be satisfied. Besides, let the function* $\gamma_{12}(\theta,\eta)\gamma_{22}^{-1}(\theta,\eta)$ *be continuously differentiable w.r.t.* (θ,η). *Then for any* θ,η *there exists an estimator* $\theta = (\theta_n)_{n\geq1}$ *of the parameter of interest* θ *such that*

I. $\lim_{n\to\infty} P^n_{\theta,\eta}\{\tilde{h}^n_1(\hat\theta_n,\bar\eta_n) = 0\} = 1,$

II. $P^n_{\theta,\eta} - \lim \hat\theta_n = \theta$

2) The matrix $\begin{bmatrix} a,b \\ ,c \end{bmatrix} := \begin{bmatrix} a,b \\ b,c \end{bmatrix}$

III. *If there exists another estimator $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ with properties I and II, then*

$$\lim_{n\to\infty} P^n_{\theta,\eta}\{\tilde{\theta}_n=\hat{\theta}_n\}=1$$

IV. $\mathcal{L}\{c_n^{-1}(\hat{\theta}_n-\theta)|P^n_{\theta,\eta}\}\Rightarrow N(0,(\gamma^{-1}(\theta,\eta)\Gamma(\theta,\eta)(\gamma^{-1}(\theta,\eta))')_{11})$.

PROOF. Denote $n^n_{\theta,\eta}=m^{n,1}_{\theta,\eta}-\gamma_{12}(\theta,\eta)\gamma_{22}^{-1}(\theta,\eta)m^{n,2}_{\theta,\eta}$. Then $\tilde{h}^n_1(\theta,\eta)=L(n^n_{\theta,\eta},M^n_{\theta,\eta})$.
By Taylor's formula we have

$$\tilde{h}^n_1(y,\bar{\eta}_n)=\tilde{h}^n_1(\theta,\eta)+(y-\theta)\overset{*}{L}(n^n_{\theta,\eta},M^n_{\theta,\eta})+(\bar{\eta}_n-\eta)\dot{L}(n^n_{\theta,\eta},M^n_{\theta,\eta})$$
$$+(y-\theta)(\overset{*}{L}(n^n_{\tilde{\theta},\tilde{\eta}},M^n_{\tilde{\theta},\tilde{\eta}})-\overset{*}{L}(n^n_{\theta,\eta},M^n_{\theta,\eta}))+(\bar{\eta}_n-\eta)(\dot{L}(n^n_{\tilde{\theta},\tilde{\eta}},M^n_{\tilde{\theta},\tilde{\eta}})), \tag{9.4}$$

where $(\tilde{\theta},\tilde{\eta})'=(\theta,\eta)'+\alpha(y-\theta,\bar{\eta}_n-\eta)'$, $\alpha\in[0,1]$.
Further (for simplicity we will omit the arguments (θ,η)), we have

$$\overset{*}{L}(n^n,M^n)=L(\overset{*}{n}{}^n,M^n)-[\tilde{h}^n_1,l^n_1],$$
$$\dot{L}(n^n,M^n)=L(\dot{n}^n,M^n)-[\tilde{h}^n_1,l^n_2].$$

The ergodicity conditions imply that

$$c_n^2L(\overset{*}{n}{}^n,M^n)\xrightarrow{P^n_{\theta,\eta}}0,$$
$$c_n^2L(\dot{n}^n,M^n)\xrightarrow{P^n_{\theta,\eta}}0.$$

By Lindeberg's condition and (9.2), we have

$$c_n^2[\tilde{h}^n_1,l^n_1]\xrightarrow{P^n_{\theta,\eta}}\gamma_{11}-\gamma_{12}\gamma_{22}^{-1}:=\beta,$$
$$c_n^2[\tilde{h}^n_1,l^n_2]\xrightarrow{P^n_{\theta,\eta}}0.$$

Consequently

$$c_n^2\overset{*}{L}(n^n,M^n)\xrightarrow{P^n_{\theta,\eta}}-\beta, \tag{9.5}$$
$$c_n^2\dot{L}(n^n,M^n)\xrightarrow{P^n_{\theta,\eta}}0. \tag{9.6}$$

Taking into account the last relations we can rewrite (9.4) as

$$c_n^2\tilde{h}^n_1(y,\bar{\eta}_n)=c_n^2\tilde{h}^n_1(\theta,\eta)-\beta(\theta,\eta)(y-\theta)+(y-\theta)\{c_n^2(\overset{*}{L}(n^n_{\tilde{\theta},\tilde{\eta}},M^n_{\tilde{\theta},\tilde{\eta}})-\overset{*}{L}(n^n_{\theta,\eta},M^n_{\theta,\eta}))$$
$$+(c_n^2\overset{*}{L}(n^n_{\theta,\eta},M_{\theta,\eta})-\beta(\theta,\eta)\}$$
$$+(\bar{\eta}_n-\eta)\{c_n^2(\dot{L}(n^n_{\tilde{\theta},\tilde{\eta}},M^n_{\tilde{\theta},\tilde{\eta}})-\dot{L}(n^n_{\theta,\eta},M^n_{\theta,\eta}))+c_n^2\dot{L}(n^n_{\theta,\eta},M^n_{\theta,\eta})\}$$
$$=c_n^2\tilde{h}^n_1(\theta,\eta)-\beta(\theta,\eta)(y-\theta)+(y-\theta)\epsilon^n_1(\tilde{\theta},\tilde{\eta},\theta,\eta)+(\bar{\eta}_n-\eta)\epsilon^n_2(\tilde{\theta},\tilde{\eta},\theta,\eta), \tag{9.7}$$

where

$$\epsilon^n_1(u,v,\theta,\eta)=c_n^2(\overset{*}{L}(n^n_{u,v},M^n_{u,v})-\overset{*}{L}(n^n_{\theta,\eta},M^n_{\theta,\eta}))+(c_n^2L^n_{\theta,\eta},M^n_{\theta,\eta})-\beta(\theta,\eta)),$$
$$\epsilon^n_2(u,v,\theta,\eta)=c_n^2(\dot{L}(n^n_{u,v},M^n_{u,v})-\dot{L}(n^n_{\theta,\eta},M^n_{\theta,\eta}))+c_n^2\dot{L}(n^n_{\theta,\eta},M^n_{\theta,\eta}),\quad u,v\in\Theta.$$

By condition e) of Theorem 5.1, the continuous differentiability of $\gamma_{12}(\theta,\eta)\gamma_{22}^{-1}(\theta,\eta)$ and properties (9.5) and (9.6) it can be easily seen that

$$\lim_{r\to 0}\overline{\lim_{n\to\infty}}P^n_{\theta,\eta}\{\sup_{u,v:|u-\theta|+|v-\eta|<r}(\|\epsilon^n_1(u,v,\theta,\eta)\|+\|\epsilon^n_2(u,v,\theta,\eta)\|)>\rho\}=0 \tag{9.8}$$

for any $\rho>0$.

To proceed further we will follow the same scheme as in proving Lemma 4.1 by constructing first a set $\Omega(n,r)$ with properties 1), 2) (see the proof of Lemma 4.1, section 2°) and property 3): for any $n\geq 1$, $r>0$ and $\omega\in\Omega(u,r)$ the equation (9.3) has the unique solution in the ball $|y-\theta|\leq r$.

The following set proves to possess the properties indicated above:

$$\Omega(n,r)=\{\omega: c_n^2|\tilde{h}_1^n(\theta,\eta)|<\frac{\beta(\theta,\eta)r}{2}, \sup_{u,v:|u-\theta|+|v-\eta|<r}|\epsilon_i^n(u,v,\theta,\eta)|<\frac{\beta(\theta,\eta)}{4}, i=1,2, |\bar{\eta}_n-\eta|<r\}.$$

Note that from (9.7) we have

$$c_n^2\tilde{h}_1^n(y,\bar{\eta}_n)<\beta r/2-\beta r+\frac{\beta r}{4}+\frac{\beta r}{4}=0 \text{ if } y-\theta=r \text{ and } \omega\in\Omega(n,r)$$

and

$$c_n^2\tilde{h}_1^n(y,\bar{\eta}_n)>0 \text{ if } y-\theta=-r$$

Assertions I-II are proved exactly as the corresponding assertions in Lemma 4.1. The proof of IV is based on the decomposition

$$c_n\tilde{h}_1^n(\hat{\theta}_n,\bar{\eta}_n)=c_n\tilde{h}_1^n(\theta,\eta)-\beta(\theta,\eta)(\hat{\theta}_n-\theta)c_n^{-1}+c_n^{-1}(\hat{\theta}_n-\theta)\epsilon_1(\theta,\eta,\theta,\eta)$$
$$+c_n^{-1}(\bar{\eta}_n-\eta)\epsilon_2^n(\hat{\theta}_n,\bar{\eta}_n,\theta,\eta), \tag{9.9}$$

where $\hat{\theta}_n$ is the estimator constructed in I. As in the course of proving Lemma 4.1, we can use (9.9) to see that

$$\lim_{n\to\infty}\mathcal{L}\{\beta(\theta,\eta)c_n^{-1}(\hat{\theta}_n-\theta)|P_{\theta,\eta}^n\}=\lim_{n\to\infty}\mathcal{L}\{c_n\tilde{h}_1^n(\theta,\eta)|P_{\theta,\eta}^n\}$$

which implies assertion IV. □

9.3. Limit Distribution Characterization

An estimator $\tilde{\theta}=(\tilde{\theta}_n), n\geq 1$ of the parameter of interest θ is usually called regular at point (θ,η) if for some nondegenerate distribution $F(x)$, $x\in R^1$ we have

$$\mathcal{L}\{c_n^{-1}(\tilde{\theta}_n-(\theta+c_nu))|P_{\theta+c_nu,\eta+c_nv}\}\Rightarrow F$$

for any u,v, such that $\theta+c_nu, \eta+c_nv\in\Theta$. (Note, that F is free of u,v)

THEOREM 9.3. *Let a family of measures $\{P_{\theta,\eta}^n\}$ be such that the conditions of Theorem 3.1 are satisfied at point (θ,η). Let $\tilde{\theta}=(\tilde{\theta}_n), n\geq 1$ be an estimator of the parameter θ regular at (θ,η). Then:*

1) *the limit law for the random variable $\xi_n=c_n^{-1}(\tilde{\theta}_n-\theta)$ is represented as the convolution of the normal law[1)] $N(0,I_{11}^*(\theta,\eta)^{-1})$ and some other distribution law $G(x)$;*
2) *random variables $\xi_n-\tilde{\Delta}_1^n/I_{11}^*$ and $\tilde{\Delta}_1^n/I_{11}^*$ are asymptotically independent in the sense that*

$$P_{\theta,\eta}^n\{\xi_n-\frac{\tilde{\Delta}_1^n}{I_{11}^*}<x, \tilde{\Delta}_1^n/I_{11}^*<y\}\Rightarrow G(x)\Phi(y),$$

where Φ is a distribution function of the law $N(0,(I_{11}^*(\theta,\eta))^{-1})$, $\tilde{\Delta}_1^n=c_n\tilde{I}_1^n$, $\tilde{I}_1^n=l_1^n-I_{12}I_{22}^{-1}l_2^n$.

PROOF. Denote by $f(s)$ the characteristic function of the distribution $F(x)$. Since the estimator $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ is regular, for $n\to\infty$ we have

$$E_{\theta+c_nu,\eta+c_nv}^n e^{isc_n^{-1}(\tilde{\theta}_n-(\theta+c_nu))}=E_{\theta,\eta}^n e^{isc_n^{-1}[(\tilde{\theta}_n-\theta)+\ln z_{\theta,\eta}^n(u,v)]-isu}\to f(s). \tag{9.10}$$

1) $I_{11}^*(\theta,\eta)=I_{11}(\theta,\eta)-I_{12}^2(\theta,\eta)I_{22}^{-1}$, where $I(\theta,\eta)=(I_{ij}(\theta,\eta))_{i,j=1,2}$.

For each fixed u we can choose $v=v(u)$ such that

$$I_{12}^2 I_{22}^{-1} u^2 + 2I_{12}uv + v^2 I_{22} = 0,$$

i.e. $v(u) = -I_{12}I_{22}^{-1}u$ (here and below arguments (θ,η) are omited). Then

$$\ln z_{\theta,\eta}^n(u, v(u)) = \tilde{\Delta}_1^n u - \frac{1}{2} I_{11}^* u^2 + \tilde{\psi}_n$$

where

$$\tilde{\Delta}_1^n = c_n \tilde{l}_1^n,\ \tilde{\psi}_n = \psi_n + (\Delta_1^n - c_n \tilde{l}_1^n)u + \Delta_2^n v(u)$$

(see Theorem 3.1 for notation) and

$$\mathfrak{L}(\tilde{\Delta}_1^n | P_{\theta,\eta}^n) \Rightarrow N(0, I_n^*),$$

$$P_{\theta,\eta}^n - \lim_{n\to\infty} \tilde{\psi}_n = 0.$$

Indeed, $\tilde{\Delta}_1^n$ is a $P_{\theta,\eta}^n$-martingale with the characteristic

$$\langle \tilde{\Delta}_1^n \rangle = c_n^2 \langle l_1^n - I_{12} I_{22}^{-1} l_2^n \rangle = c_n^2 \langle l_1^n \rangle - 2I_{12}I_{22}^{-1} c_n^2 \langle l_1^n, l_2^n \rangle + c_n^2 I_{12}^2 I_{22}^{-2} \langle l_2^n \rangle$$

and by ergodicity

$$P_{\theta,\eta}^n - \lim_{n\to\infty} \langle \tilde{\Delta}_1^n \rangle_T = I_{11} - I_{12}^2 I_{22}^{-1} = I_{11}^*.$$

Hence, by condition (b) of Theorem 3.1, it follows that

$$\mathfrak{L}(\tilde{\Delta}_1^n | P_{\theta,\eta}^n) \Rightarrow N(0, I_{11}^*).$$

Further, $\psi_n \xrightarrow{P_{\theta,\eta}^n} 0$ and $(\Delta_1^n - c_n \tilde{l}_1^n)u + \Delta_2^n\, v(u)$ is a $P_{\theta,\eta}^n$-martingale with the property

$$\langle (\Delta_1^n - c_n \tilde{l}_1^n)u + \Delta_2^n v(u) \rangle_T = \langle (\Delta_1^n - c_n \tilde{l}_1^n)u - I_{12} I_{22}^{-1} \Delta_2^n u \rangle_T$$

$$= u^2 c_n^2 (I_{12}^2 I_{22}^{-2} \langle l_2^n \rangle_T - 2 I_{12}^2 I_{22}^{-2} \langle l_2^n \rangle_T + I_{12}^2 I_{22}^{-2} \langle l_2^n \rangle_T) \xrightarrow{P_{\theta,\eta}^n} 0.$$

By the fact that the limit law F is indepent of u and v, from (9.10) we have for $n\to\infty$

$$E_{\theta,\eta}^n e^{isc_n^{-1}(\tilde{\theta}_n - \theta)} + \tilde{\Delta}_1^n u - \frac{1}{2} I_{11}^* u^2 - isu \to f(s). \tag{9.11}$$

Theorem 11.8.1 in [5] and its corollary tells us that there exists a truncation $\hat{\Delta}_1^n$ of $\tilde{\Delta}_1^n$ such that

$$\lim_{n\to\infty} E_{\theta,\eta}^n e^{isc_n^{-1}(\tilde{\theta}_n - \theta)} + \tilde{\Delta}_1^n u - \frac{1}{2} I_{11}^* u^2 - isu = \lim_{n\to\infty} E_{\theta,\eta}^n e^{isc_n^{-1}(\tilde{\theta}_n - \theta)}$$

$$+ \hat{\Delta}_1^n u - \frac{1}{2} I_n^* u^2 - isu = f(s).$$

Therefore, for any u we have

$$\lim_{n\to\infty} E_{\theta,\eta}^n\, e^{isc_n^{-1}(\tilde{\theta}_n - \theta) + \hat{\Delta}_1 u} = f(s) e^{\frac{1}{2} I_{11}^* u^2 + isu}. \tag{9.12}$$

By standard arguments (see [5], p. 213) it can be shown that relation (9.12) is valid for complex values of u too. Put $u = -is/I_{11}^*$. Then by (9.11) we have

$$\lim_{n\to\infty} E_{\theta,\eta}^n\, e^{is(c_n^{-1}(\tilde{\theta}_n - \theta) - \hat{\Delta}_1 / I_{11}^*)} = f(s) e^{\frac{1}{2}s^2 / I_{11}^*}.$$

Hence, by the continuous correspondence between distribution functions and characteristic functions it follows that the random variable $c_n^{-1}(\tilde{\theta}_n - \theta) - \hat{\Delta}_1^n / I_{11}^*$, as well as $c_n^{-1}(\tilde{\theta}_n - \theta) - \tilde{\Delta}_1^n / I_{11}^*$, has the limit distribution G with the characteristic function g such that

$$f(s)=g(s)e^{\frac{1}{2}s^2/\Gamma^*_{11}}. \tag{9.13}$$

Thus assertion 1) of the theorem is proved.

Choosing now $u=\dfrac{i(\sigma-s)}{\Gamma^*_{11}}$ and using (9.11) and (9.13) we obtain

$$\lim_{n\to\infty} E^n_{\theta,\eta}\exp\{is(c_n^{-1}(\tilde{\theta}_n-\theta)-\hat{\Delta}^n_1/\Gamma^*_{11})+i\sigma\hat{\Delta}^n_1/\Gamma^*_{11}\}=g(s)e^{-\frac{1}{2}\sigma^2/\Gamma^*_{11}}$$

as required. □

9.4. Regularity of estimators constructed by Skew projection technique.

The estimator constructed in Theorem 9.2 is irregular, in general. But the estimator $\tilde{\theta}$ solving the equation (9.3), where $\bar{\eta}=(\bar{\eta}_n)_{n\geq 1}$ is an estimator of the parameter η such that for all u,v $c_n^{-1}(\bar{\eta}_n-\eta)=0_{P^n_{\theta+c_nu,\eta+c_nv}}(1)$, is regular under the conditions of Theorems 3.1 and 5.1.

THEOREM 9.4. *Let the conditions of Theorems 3.1 and 5.1 be satisfied. Then for every θ there exists an estimator $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ of the parameter of interest θ regular at (θ,η), with*

$$F=N(0,(\gamma^{-1}(\theta,\eta)\Gamma(\theta,\eta)(\gamma^{-1}(\theta,\eta))')_{11}).$$

PROOF. As in the course of proving Theorem 9.2, we can easily obtain the decomposition

$$c_n\tilde{h}^n_1(\tilde{\theta}_n,\bar{\eta}_n)=c_n\tilde{h}^n_1(\theta,\eta)-\beta(\theta,\eta)c_n^{-1}(\tilde{\theta}_n-\theta)+c_n^{-1}(\tilde{\theta}_n-\theta)\epsilon^n_1+ + c_n^{-1}(\bar{\eta}_n-\eta)\epsilon^n_2,$$

where $\tilde{\theta}=(\tilde{\theta}_n)_{n\geq 1}$ is the estimator constructed by the equation (9.3) in the same manner as in Theorem 9.2.

Now we can easily obtain

$$\begin{aligned}
&|c_n\tilde{h}^n_1(\tilde{\theta}_n,\bar{\eta}_n)+c_n\tilde{h}^n_1(\theta,\eta)-c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v\rangle\\
&\quad-\beta(\theta,\eta)c_n^{-1}(\tilde{\theta}_n-\theta)-c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v\rangle|\\
&\quad\leq|\epsilon^n_1\beta^{-1}(\theta,\eta)||\beta(\theta,\eta)c_n^{-1}(\tilde{\theta}_n-\theta)-c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_1(\theta,\eta)v\rangle|\\
&\quad+|\epsilon^n_1\beta^{-1}(\theta,\eta)|\cdot|c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v\rangle|+|\epsilon^n_2||c_n^{-1}(\bar{\eta}_n-\eta)|.
\end{aligned}$$

Consequently, taking into consideration the properties of $\epsilon^n_1,\epsilon^n_2$ and $(\bar{\eta}_n)$, we get as in the course of proving Theorem 9.2 that

$$\begin{aligned}
&\lim_{n\to\infty}\mathcal{L}\{\beta(\theta,\eta)c_n^{-1}(\tilde{\theta}_n-\theta)-c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v\rangle|P^n_{\theta+c_nu,\eta+c_nv}\}\\
&\qquad=\lim_{n\to\infty}\mathcal{L}\{c_n\tilde{h}^n_1(\theta,\eta)-c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v|P^n_{\theta+c_nu,\eta+c_nv}\}.
\end{aligned}$$

By the definition of $\tilde{h}^n_1(\theta,\eta)$, and the contiguity $(P^n_{\theta+c_nu,\eta+c_nv})\lhd(P^n_{\theta,\eta})$ we have

$$P^n_{\theta+c_nu,\eta+c_nv}-\lim_{n\to\infty}c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v\rangle_T=-\beta(\theta,\eta)u.$$

To complete the proof it remains to find the limit distribution (calculated w.r.t. the measure $P^n_{\theta+c_nu,\eta+c_nv}$) of the expression

$$\beta^{-1}(\theta,\eta)(c_n\tilde{h}^n_1(\theta,\eta)-c_n^2\langle\tilde{h}^n_1(\theta,\eta),l^n_1(\theta,\eta)u+l^n_2(\theta,\eta)v\rangle.$$

As in the course of proving Theorem 7.1, we use the transformation formula for the triplet under the absolutely continuous change of a measure to arrive at the desired result. □

9.5. Partial likelihood estimator and comparison of variances in presence of a nuisance parameter.

We consider the situation described in § 6 in presence of a nuisance parameter. Assume that

$$\frac{dP^n_{\theta,\eta}}{dP^n} = \mathcal{E}(M^n_{\theta,\eta}) = \mathcal{E}(X^n_{\theta,\eta})\,\mathcal{E}(N^n_\theta),$$

where $X^n_{\theta,\eta}$ is a P^n-martingale depending on (θ,η), while a P^n-martingale N^n_θ depends only on θ and $X^n_{\theta,\eta} \perp N^n_\theta$ for all $\theta,\eta \in R^1$. Suppose also that N_θ is θ-differentiable (denote $\overset{*}{N}{}^n_\theta = \frac{\partial}{\partial\theta} N^n_\theta$) and $\overset{*}{N}{}^n_\theta \perp X^n_{\theta,\eta}$ for all $\theta,\eta \in \underline{R}^1$.

The PLE $\hat{\theta} = (\hat{\theta}_n)_{n \geq 1}$ of the parameter θ is defined as usual by the equation

$$L(\overset{*}{N}{}^n_\theta, N^n_\theta) = 0.$$

Note that the process $L(\overset{*}{N}{}^n_\theta, N^n_\theta)$ is a $P^n_{\theta,\eta}$-martingale for all $\eta \in R^1$.

Applying now Theorem 6.1 to this case in which $Q^{X,n}_\theta = P^n_{\theta,\eta}$, $m^n_\theta = \overset{*}{N}{}^n_\theta$ we get

$$\mathcal{L}\{c_n^{-1}(\hat{\theta}_n - \theta) | P^n_{\theta,\eta}\} \Rightarrow N(0, \sigma^{-1}, (\theta,\eta))$$

where

$$\sigma^{-1}(\theta,\eta) = P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2 \langle L(\overset{*}{N}{}^n_\theta, N^n_\theta)\rangle_T.$$

If $\tilde{\theta} = (\tilde{\theta}_n)_{n \geq 1}$ is the estimator constructed by means of the classical projection technique i.e. by the equation

$$\tilde{l}^n_1(\theta, \bar{\eta}_n) = 0$$

when $\bar{\eta}$ is a c_n-consistent estimator of η and $\tilde{l}^n_1(\theta,\eta)$ is a modified informant, then

$$\mathcal{L}\{c_n^{-1}(\tilde{\theta}_n - \theta) | P^n_{\theta,\eta}\} \Rightarrow N(0, \sigma_2(\theta,\eta))$$

where

$$\sigma_2^{-1}(\theta,\eta) = I_{11}(\theta,\eta) - I^2_{12}(\theta,\eta) I^{-1}_{22}(\theta,\eta) = I^*_{11}(\theta,\eta) \tag{9.14}$$

(see Theorem 9.2).

Finally if $\bar{\theta} = (\bar{\theta}_n)_{n \geq 1}$ is MLE of parameter θ with known η, then

$$\mathcal{L}\{c_n^{-1}(\hat{\theta}_n - \theta) | P^n_{\theta,\eta}\} \Rightarrow N(0, \sigma_3(\theta,\eta))$$

where

$$\sigma_3^{-1}(\theta,\eta) = I_{11}(\theta,\eta)$$

(see Theorem 5.1).

We will compare these variances. Moreover, we will establish a functional relation between them.

It can be easily seen that the $P^n_{\theta,\eta}$-martingale $L(\overset{*}{M}{}^n_{\theta,\eta}, M^n_{\theta,\eta})$ can be decomposed at least in two ways

$$L(\overset{*}{M}{}^n_{\theta,\eta}, M^n_{\theta,\eta}) = \tilde{l}^n_1(\theta,\eta) - I_{12}(\theta,\eta) I^{-1}_{22}(\theta,\eta) L(\dot{M}^n_{\theta,\eta}, M^n_{\theta,\eta}) \tag{9.15}$$

and

$$L(\overset{*}{M}{}^n_{\theta,\eta}, M^n_{\theta,\eta}) = L(\overset{*}{N}{}^n_\theta, N^n_\theta) + L(m^n_1(\theta,\eta), M^n_{\theta,\eta}) \tag{9.16}$$

where

$$m^n_1(\theta,\eta) = \overset{*}{M}{}^n_{\theta,\eta} - m^n_{\theta,\eta}, m^n_{\theta,\eta} = \overset{*}{N}{}^n_{\theta,\eta} + [\overset{*}{N}{}^n_{\theta,\eta}, X^n_{\theta,\eta}].$$

By definition of the modified informant $\tilde{l}_1^n(\theta,\eta)$ we have

$$c_n^2\langle \tilde{l}_1(\theta,\eta), L(\dot{M}^n_{\theta,\eta}, M^n_{\theta,\eta})\rangle_T \xrightarrow{P^n_{\theta,\eta}} 0. \tag{9.17}$$

Besides it is easily shown that

$$\langle L(\overset{*}{M}{}^n_\theta, N^n_\theta), L(m_1^n(\theta,\eta), M^n_{\theta,\eta})\rangle_T = 0. \tag{9.18}$$

Indeed, by definition of $m_1^n(\theta,\eta)$ and L-transformation we have

$$[L(\overset{*}{N}{}^n_\theta, N^n_\theta), L(m_1^n(\theta,\eta), M^n_{\theta,\eta})] = \langle \overset{*}{N}{}^{n,c}_\theta, \overset{*}{X}{}^{n,c}_{\theta,\eta}\rangle + \sum \frac{\Delta \overset{*}{N}{}^n_\theta \Delta X^n_{\theta,\eta}}{(1+\Delta N^n_\theta)(1+\Delta X^n_{\theta,\eta})} = L([\overset{*}{N}{}^n_\theta, \overset{*}{X}{}^n_{\theta,\eta}], M^n_{\theta,\eta}).$$

But the process $[\overset{*}{N}{}^n_\theta, \overset{*}{X}{}^n_{\theta,\eta}]$ is a P^n-martingale (if we assume that the following natural condition holds: $\overset{*}{N}{}^n_\theta \perp \overset{*}{X}{}^n_{\theta,\eta}$) and so the process $L([\overset{*}{N}{}^n_\theta, \overset{*}{X}{}^n_{\theta,\eta}], M^n_{\theta,\eta})$ is a $P^n_{\theta,\eta}$-martingale, which means that the process $[L(\overset{*}{N}{}^n_\theta, N^n_\theta), L(m_1^n(\theta,\eta), M^n_{\theta,\eta})]$ is also a $P^n_{\theta,\eta}$-martingale and consequently

$$L(\overset{*}{N}{}^n_\theta, N^n_\theta) \perp L(m_1^n(\theta,\eta), M^n_{\theta,\eta}).$$

Since $L(\overset{*}{N}{}^n_\theta, N^n_\theta)$, $L(m_1^n(\theta,\eta), M^n_{\theta,\eta}) \in M^2(P^n_{\theta,\eta})$, we get (9.18)
The decompositions (9.16) and (9.18) lead to

$$\sigma_3^{-1}(\theta,\eta) = \sigma_1^{-1}(\theta,\eta) + P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2\langle L(m_1^n(\theta,\eta), M^n_{\theta,\eta})\rangle_T.$$

On the other hand, (9.14) gives

$$\sigma_3^{-1}(\theta,\eta) = \sigma_2^{-1}(\theta,\eta) + \frac{(P^n_{\theta,\eta} - \lim\limits_{n\to\infty} c_n^2\langle L(m_1^n, M^n_{\theta,\eta}), L(\dot{M}^n_{\theta,\eta})\rangle)^2}{P^n_{\theta,\eta} - \lim\limits_{n\to\infty} c_n^2\langle L(\dot{M}^n_{\theta,\eta}, M^n_{\theta,\eta})\rangle}.$$

The last relations imply

$$\sigma_2^{-1}(\theta,\eta) = \sigma_1^{-1}(\theta,\eta)$$

$$+(P^n_{\theta,\eta} - \lim_{n\to\infty} c_n^2\langle L(m_1^n, M^n_{\theta,\eta})\rangle)\left[1 - \frac{(P^n_{\theta,\eta} - \lim\limits_{n\to\infty} c_n^2\langle L(m_1^n, M^n_{\theta,\eta}), L(\dot{M}^n_{\theta,\eta}, M^n_{\theta,\eta})\rangle)^2}{(P^n_{\theta,\eta} - \lim\limits_{n\to\infty} c_n^2\langle L(m_1^n, M^n_{\theta,\eta})\rangle)(P^n_{\theta,\eta} - \lim\limits_{n\to\infty} c_n^2\langle L(\dot{M}^n_{\theta,\eta}, M^n_{\theta,\eta})\rangle)}\right].$$

It is easily seen from the obtained relations for $\sigma_1^{-1}, \sigma_2^{-1}, \sigma_3^{-1}$ that

$$\sigma_3 \leqslant \sigma_2 \leqslant \sigma_1.$$

As usual, the condition for adaptation i.e. for $\sigma_3 = \sigma_2$, is expressed as $I_{12}(\theta,\eta) = 0$. For $\sigma_2 = \sigma_1$ we have the following condition:

$$P^n_{\theta,\eta} - \lim_{n\to\infty} \frac{(\langle L(m_1^n, M^n_{\theta,\eta})\rangle)^2}{\langle L(m_1^n, M^n_{\theta,\eta}\rangle \langle L(\dot{M}^n_{\theta,\eta}, M^n_{\theta,\eta})\rangle} = 1.$$

In the i.i.d. case the relations between σ_1, σ_2 and σ_3 take the following simple form:

$$\sigma_1^{-1} = \int \left[\frac{\overset{*}{\psi}}{\psi}\right]^2 \psi\phi\, d\mu,$$

$$\sigma_3^{-1} = \int \left[\frac{\overset{*}{\psi}}{\psi} + \frac{\overset{*}{\phi}}{\phi}\right]^2 \psi\phi\, d\mu = \sigma_1^{-1} + \int \left[\frac{\overset{*}{\phi}}{\phi}\right]^2 \psi\phi\, d\mu,$$

$$\sigma_2^{-1} = \sigma_1^{-1} + \left(\int \left[\frac{\overset{*}{\phi}}{\phi}\right]^2 \psi\phi\, d\mu\right)\left(1 - \frac{\left(\int \frac{\overset{*}{\phi}}{\phi}\frac{\dot{\phi}}{\phi}\psi\phi\, d\mu\right)^2}{\left(\int \left[\frac{\overset{*}{\phi}}{\phi}\right]^2 \psi\phi\, d\mu\right)\left(\int \left[\frac{\dot{\phi}}{\phi}\right]^2 \phi\psi\, d\mu\right)}\right).$$

Bibliographical notes

The local asymptotic normality (LAN) of distributions was studied by various authors. In particular, the diffusion case was studied by Yu. Kutojants, the case of point processes by K. Dzhaparidze, the case of semimartingales with an integral representation by A. Taraskin and Yu. Lin'kov. In [9] and [10] (univariate and multivariate cases, respectively) an integral representation of semimartingales is not required.

Partial likelihood technique based on the factorization of the full likelihood was introduced by Cox [11]. Partial likelihood theory in the discrete time scheme was considered by Wong in [12], where detailed references are given. For point processes the Cox model has been generalized by Dzhaparidze [13]. The partial likelihood scheme for the general statistical model was investigated by J. Jacod [17]. The asymptotic behaviour of an estimator w.r.t. asymptotically distingwishable alternatives have not been studied.

The scheme described in section 5 extends the Cox-Wong method to the general case; see also [14].

Wellner [15] have discussed projection methods, one-step approximation methods, etc, for the i.i.d. case. In section 6 we extend some of these methods to the general case; see [15] for bibliography.

References

1. Jacod, J. (1979). Calcul stochastique et problemes de martingales. *Lect. Notes in Math.,* 714, New York, Springer.
2. Liptser, R.Sh., Shiryayev, A.N. (1989). *Theory of Martingales,* Dordrecht, Kluwer Acad. Publ.
3. Jacod, J., Shyryayev, A.N. (1987). *Limit Theorems for Stochastic Processes,* New York, Springer.
4. Le Breton, A. (1976). *Dissertation*, Grenoble.
5. Ibragimov, I.A., Khas'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory,* New York, Springer.
6. Billingsly, P. (1968). *Convergence of Probability Measures,* New York, Willey.
7. Dellacherie, C. (1972). *Capacites et Processus Stochastiques,* Springer-Verlag, Berlin.
8. Lazrieva, N.L., Toronjadze, T.A., Chitashvili R.J. (1989). On asymptotic estimation theory under model disturbance in a general scheme of statistical models. *Theory of probability and mathematical statistics, Tbilisi* (Russian).
9. Lazrieva, N.L., Toronjadze, T.A. (1986). Asymptotic properties of the maximal likelihood estimator in a sequence of general statistical models. *Bulletin Academy Science Georgian SSR,* 133, No. 1.
10. Kordzakhia, N., Nadaraya, L. (1989). Asymptotic properties of projection estimators under nuisance parameter in a general scheme of statistical models. *Bulletin Academy Science Georgian SSR,* No 1. (Russian).
11. Cox, D.H. (1975). Partial likelihood. *Biometrica,* 64, 269-276.
12. Wing Hung Wong (1986). Theory of partial likelihood, AS,v.14, No 1.
13. Dzhaparidze, K. (1986). On asymptotic inference about intensity parameters of a counting process. *Centre for Mathematics and Computer Science,* Report MS-R8614.
14. Lazrieva, N.L., Toronjadze, T.A. (1989). Asymptotic properties of M-estimator in a general scheme of statistical models. *Statistics and control of random processes. Transections of Steklov institute of mathematics,* Moskow (Russian).

15. Wellner, J.A. (1986). Semiparametric Models: Progress and Problems. *Centre for Mathematics and Computer Science,* Amsterdam, Report MS-R8614.
16. Klebanov, L., Melamed, J. (1976). Fisher information under nuisance parameters and some applications to statistical estimation theory. *Bulletin Academy Science Georgian SSR, 81, 1.*
17. Jacod, J. (1988). Regularity, Partial regularity, Partial Information Processes, for a filtered statistical Model. *Laboratoire de Probabilites Universite Pierre et Marie Curie,* Preprint.