Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 171 (2017) 57–75
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Recursive estimation procedures for one-dimensional parameter of statistical models associated with semimartingales

Nanuli Lazrieva*, Temur Toronjadze

Business School, Georgian–American University, 8 M. Aleksidze Str., Tbilisi 0160, Georgia
A. Razmadze Mathematical Institute of I. Javakhishvili Tbilisi State University, 6 Tamarashvili Str., Tbilisi 0177, Georgia

Available online 22 December 2016

Abstract

The recursive estimation problem of a one-dimensional parameter for statistical models associated with semimartingales is considered. The asymptotic properties of recursive estimators are derived, based on the results on the asymptotic behavior of a Robbins–Monro type SDE. Various special cases are considered.

Keywords: Stochastic approximation; Robbins–Monro type SDE; Semimartingale statistical models; Recursive estimation; Asymptotic properties

0. Introduction

Beginning from the paper [1] of A. Albert and L. Gardner a link between Robbins–Monro (RM) stochastic approximation algorithm (introduced in [2]) and recursive parameter estimation procedures was intensively exploited. Later on recursive parameter estimation procedures for various special models (e.g., i.i.d. models, non i.i.d. models in discrete time, etc.) have been studied by a number of authors using methods of stochastic approximation (see, e.g., [3–12]). It would be mentioned the fundamental book [13] by M.B. Nevelson and R.Z. Khas'minski (1972) between them.

In 1987 by N. Lazrieva and T. Toronjadze a heuristic algorithm of a construction of the recursive parameter estimation procedures for statistical models associated with semimartingales (including both discrete and continuous time semimartingale statistical models) was proposed [14]. These procedures could not be covered by the generalized stochastic approximation algorithm with martingale noises (see, e.g., [15]), while in discrete time case the classical RM algorithm contains recursive estimation procedures.

To recover the link between the stochastic approximation and recursive parameter estimation in [16–18] by Lazrieva, Sharia and Toronjadze the semimartingale stochastic differential equation was introduced, which naturally

* Corresponding author at: Business School, Georgian–American University, 8 M. Aleksidze Str., Tbilisi 0160, Georgia.
E-mail addresses: laz@rmi.ge (N. Lazrieva), toronj333@yahoo.com (T. Toronjadze).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

http://dx.doi.org/10.1016/j.trmi.2016.12.001

includes both generalized RM stochastic approximation algorithms with martingale noises and recursive parameter estimation procedures for semimartingale statistical models.

In the present work we are concerning with the construction of recursive estimation procedures for semimartingale statistical models asymptotically equivalent to the MLE and M-estimators, embedding these procedures in the Robbins–Monro type equation. For this reason in Section 1 we shortly describe the Robbins–Monro type SDE and give necessary objects to state results concerning the asymptotic behavior of recursive estimator procedures.

In Section 2 we give a heuristic algorithm of constructing recursive estimation procedures for one-dimensional parameter of semimartingale statistical models. These procedures provide estimators asymptotically equivalent to MLE. To study the asymptotic behavior of these procedures we rewrite them in the form of the Robbins–Monro type SDE. Besides, we give a detailed description of all objects presented in this SDE, allowing us separately study special cases (e.g. discrete time case, diffusion processes, point processes, etc.).

In Section 4 we formulate main results concerning the asymptotic behavior of recursive procedures, asymptotically equivalent to the MLE.

In Section 5, we develop recursive procedures, asymptotically equivalent to M-estimators.

Finally, in Section 6, we give various examples demonstrating the usefulness of our approach.

1. The Robbins–Monro type SDE

Let on the stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions the following objects be given:

(a) the random field $H = \{H_t(u), t \geq 0, u \in R^1\} = \{H_t(\omega, u), t \geq 0, \omega \in \Omega, u \in R^1\}$ such that for each $u \in R^1$ the process $H(u) = (H_t(u))_{t\geq 0} \in \mathcal{P}$ (i.e. is predictable);

(b) the random field $M = \{M(t, u), t \geq 0, u \in R^1\} = \{M(\omega, t, u), \omega \in \Omega, t \geq 0, u \in R^1\}$ such that for each $u \in R^1$ the process $M(u) = (M(t, u))_{t\geq 0} \in \mathcal{M}^2_{\rm loc}(P)$;

(c) the predictable increasing process $K = (K_t)_{t\geq 0}$ (i.e. $K \in \mathcal{V}^+ \cap \mathcal{P}$).

In the sequel we restrict ourselves to the consideration of the following particular case: for each $u \in R^1 M(u) = \varphi(u) \cdot m + W(u) * (\mu - \nu)$, where $m \in \mathcal{M}^c_{\rm loc}(P)$, μ is an integer-valued random measure on $(R \times E, \mathcal{B}(R_+) \times \mathcal{E})$, ν is its P-compensator, $(E, \mathcal{E})$ is the Blackwell space, $W(u) = (W(t, x, u), t \geq 0, x \in E) \in \mathcal{P} \otimes \mathcal{E}$. Here we also mean that all stochastic integrals are well-defined.[1]

Later on by the symbol $\int_0^t M(ds, u_s)$, where $u = (u_t)_{t\geq 0}$ is some predictable process, we denote the following stochastic line integrals:

$$\int_0^t \varphi(s, u_s)\, dm_s + \int_0^t \int_E W(s, x, u_s)(\mu - \nu)(ds, dx)$$

provided the latters are well-defined.

Consider the following semimartingale stochastic differential equation

$$z_t = z_0 + \int_0^t H_s(z_{s-})\, dK_s + \int_0^t M(ds, z_{s-}), \quad z_0 \in \mathcal{F}_0. \tag{1.1}$$

We call SDE (1.1) the Robbins–Monro (RM) type SDE if the drift coefficient $H_t(u)$, $t \geq 0$, $u \in R^1$ satisfies the following conditions: for all $t \in [0, \infty)$ P-a.s.

$$\text{(A)} \begin{array}{l} H_t(0) = 0, \\ H_t(u)u < 0 \quad \text{for all } u \neq 0. \end{array}$$

The question of strong solvability of SDE (1.1) is well-investigated (see, e.g., [20]).

We assume that there exists a unique strong solution $z = (z_t)_{t\geq 0}$ of Eq. (1.1) on the whole time interval $[0, \infty)$ and such that $\widetilde{M} \in \mathcal{M}^2_{\rm loc}(P)$, where

$$\widetilde{M}_t = \int_0^t M(ds, z_{s-}).$$

Sufficient conditions for the latter can be found in [20].

[1] See [19] for basic concepts and notations.

The unique solution $z = (z_t)_{t\geq 0}$ of RM type SDE (1.1) can be viewed as a semimartingale stochastic approximation procedure.

In [16,17], the asymptotic properties of the process $z = (z_t)_{t\geq 0}$ as $t \to \infty$ are investigated, namely, convergence ($z_t \to 0$ as $t \to \infty$ P-a.s.), rate of convergence (that means that for all $\delta < \frac{1}{2}$, $\gamma_t^\delta z_t \to 0$ as $t \to \infty$ P-a.s., with the specially chosen normalizing sequence $(\gamma_t)_{t\geq 0}$) and asymptotic expansion

$$\chi_t^2 z_t^2 = \frac{L_t}{\langle L\rangle_t^{1/2}} + R_t$$

with the specially chosen normalizing sequence χ_t^2 and martingale $L = (L_t)_{t\geq 0}$, where $R_t \to 0$ as $t \to \infty$ (see [16,17] for definition of objects χ_t^2, L_t and R_t).

2. Basic model and regularity

Our object of consideration is a parametric filtered statistical model

$$\mathcal{E} = (\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \{P_\theta; \theta \in R\})$$

associated with one-dimensional $\mathbb{F}$-adapted RCLL process $X = (X_t)_{t\geq 0}$ in the following way: for each $\theta \in R^1$ P_θ is assumed to be the unique measure on $(\Omega, \mathcal{F})$ such that under this measure X is a semimartingale with predictable characteristics $(B(\theta), C(\theta), \nu_\theta)$ (w.r.t. standard truncation function $h(x) = xI_{\{|x|\leq 1\}}$). For simplicity assume that all P_θ coincide on $\mathcal{F}_0$.

Suppose that for each pair (θ, θ') $P_\theta \overset{\text{loc}}{\sim} P_{\theta'}$. Fix some $\theta_0 \in R$ and denote $P = P_{\theta_0}$, $B = B(\theta_0)$, $C = C(\theta_0)$, $\nu = \nu_{\theta_0}$.

Let $\rho(\theta) = (\rho_t(\theta))_{t\geq 0}$ be a local density process (likelihood ratio process)

$$\rho_t(\theta) = \frac{dP_{\theta,t}}{dP_t},$$

where for each θ $P_{\theta,t} := P_\theta|\mathcal{F}_t$, $P_t := P|\mathcal{F}_t$ are restrictions of measures P_θ and P on $\mathcal{F}_t$, respectively.

As it is well-known (see, e.g., [21, Ch. III, §3d, Th. 3.24]) for each θ there exists a $\widetilde{\mathcal{P}}$-measurable positive function

$$Y(\theta) = \{Y(\omega, t, x; \theta), (\omega, t, x) \in \Omega \times R_+ \times R\},$$

and a predicable process $\beta(\theta) = (\beta_t(\theta))_{t\geq 0}$ with

$$|h(Y(\theta) - 1)| * \nu \in \mathcal{A}^+_{\text{loc}}(P), \qquad \beta^2(\theta) \circ C \in \mathcal{A}^+_{\text{loc}}(P),$$

and such that

$$\begin{aligned} &(1)\ B(\theta) = B + \beta(\theta) \circ C + h(Y(\theta) - 1) * \nu, \\ &(2)\ C(\theta) = C, \qquad (3)\ \nu_\theta = Y(\theta) \cdot \nu. \end{aligned} \tag{2.1}$$

In addition, the function $Y(\theta)$ can be chosen in such a way that

$$a_t := \nu(\{t\}, R) = 1 \iff a_t(\theta) := \nu_\theta(\{t\}, R) = \int Y(t, x; \theta)\nu(\{t\})dx = \widehat{Y}_t(\theta) = 1.$$

We give a definition of the regularity of the model based on the following representation of the density process as exponential martingale:

$$\rho(\theta) = \mathcal{E}(M(\theta)),$$

where

$$M(\theta) = \beta(\theta) \cdot X^c + \left(Y(\theta) - 1 + \frac{\widehat{Y}(\theta) - a}{1 - a} I_{\{0<a<1\}}\right) * (\mu - \nu) \in \mathcal{M}_{\text{loc}}(P), \tag{2.2}$$

$\mathcal{E}_t(M)$ is the Dolean exponential of the martingale M (see, e.g., [19]). Here X^c is a continuous martingale part of X under measure P.

We say that the model is regular if for almost all (ω, t, x) the functions $\beta : \theta \to \beta_t(\omega; \theta)$ and $Y : \theta \to Y(\omega, t, x; \theta)$ are differentiable (notation $\dot\beta(\theta) := \frac{\partial}{\partial\theta}\beta(\theta)$, $\dot Y(\theta) := \frac{\partial}{\partial\theta}Y(\theta)$) and differentiability under integral sign is possible. Then

$$\frac{\partial}{\partial\theta} \ln \rho(\theta) = L(\dot M(\theta), M(\theta)) := L(\theta) \in \mathcal{M}_{\mathrm{loc}}(P_\theta),$$

where $L(m, M)$ is the Girsanov transformation defined as follows: if $m, M \in \mathcal{M}_{\mathrm{loc}}(P)$ and $Q \ll P$ with $\frac{dQ}{dP} = \mathcal{E}(M)$, then

$$L(m, M) := m - (1 + \Delta M)^{-1} \circ [m, M] \in \mathcal{M}_{\mathrm{loc}}(Q).$$

It is not hard to verify that

$$L(\theta) = \dot\beta(\theta) \cdot (X^c - \beta(\theta) \circ C) + \Phi(\theta) * (\mu - \nu_\theta), \tag{2.3}$$

where

$$\Phi(\theta) = \frac{\dot Y(\theta)}{Y(\theta)} + \frac{\dot a(\theta)}{1 - a(\theta)}$$

with $I_{\{a(\theta)=1\}}\dot a(\theta) = 0$, and $0/0 = 0$ (recall that $\frac{\partial}{\partial\theta}\widehat Y(\theta) = \dot a(\theta)$).

Indeed, due to the regularity of the model, we have

$$\dot M(\theta) = \dot\beta(\theta) \cdot X^c + \left(\dot Y(\theta) - \frac{\dot a(\theta)}{1 - a} I_{(0<a<1)}\right) * (\mu - \nu)$$

and (2.3) simply follows from (1.16)–(1.18) of [22, Part I] with

$$g(\theta) = Y(\theta) - 1 + \frac{a(\theta) - a}{1 - a} I_{(0<a<1)},$$
$$\psi(\theta) = \dot Y(\theta) - \frac{\dot a(\theta)}{1 - a} I_{(0<a<1)}.$$

The empirical Fisher information process is $\widehat I_t(\theta) = [L(\theta), L(\theta)]_t$ and if we assume that for each $\theta \in R^1 L(\theta) \in \mathcal{M}^2_{\mathrm{loc}}(P_\theta)$, then the Fisher information process is

$$I_t(\theta) = \langle L(\theta), L(\theta)\rangle_t.$$

3. Recursive estimation procedure for MLE

In [14], a heuristic algorithm was proposed for the construction of recursive estimators of unknown parameter θ asymptotically equivalent to the maximum likelihood estimator (MLE).

This algorithm was derived using the following reasons:

Consider the MLE $\widehat\theta = (\widehat\theta_t)_{t\ge0}$, where $\widehat\theta_t$ is a solution of estimational equation

$$L_t(\theta) = 0.$$

The question of solvability of this equation is considered in [22, Part II].

Assume that

(1) for each $\theta \in R^1$, $I_t(\theta) \to \infty$ as $t \to \infty$, P_θ-a.s., the process $(\widehat I_t(\theta))^{1/2}(\widehat\theta_t - \theta)$ is P_θ-stochastically bounded and, in addition, the process $(\widehat\theta_t)_{t\ge0}$ is a P_θ-semimartingale;
(2) for each pair (θ', θ) the process $L(\theta') \in \mathcal{M}^2_{\mathrm{loc}}(P_{\theta'})$ and is a P_θ-special semimartingale;
(3) the family $(L(\theta), \theta \in R^1)$ is such that the Itô–Ventzel formula is applicable to the process $(L(t, \widehat\theta_t))_{t\ge0}$ w.r.t. P_θ for each $\theta \in R^1$;
(4) for each $\theta \in R^1$ there exists a positive increasing predictable process $(\gamma_t(\theta))_{t\ge0}$, $\gamma_0 > 0$, asymptotically equivalent to $\widehat I_t^{-1}(\theta)$, i.e.

$$\gamma_t(\theta)\widehat I_t(\theta) \xrightarrow{P_\theta} 1 \quad \text{as } t \to \infty.$$

Under these assumptions using the Ito–Ventzel formula for the process $(L(t,\widehat{\theta}_t))_{t\ge 0}$ we get an "implicit" stochastic equation for $\widehat{\theta}=(\widehat{\theta}_t)_{t\ge 0}$. Analyzing the orders of infinitesimality of terms of this equation and rejecting the high order terms we get the following SDE (recursive procedure)

$$d\theta_t = \gamma_t(\theta_{t-})L(dt,\theta_{t-}), \tag{3.1}$$

where $L(dt,u_t)$ is a stochastic line integral w.r.t. the family $\{L(t,u), u\in R^1, t\in R_+\}$ of P_θ-special semimartingales along the predictable curve $u=(u_t)_{t\ge 0}$.

Note that in many cases under consideration one can choose $\gamma_t(\theta)=(I_t^{-1}(\theta)+1)^{-1}$, or in ergodic situations such as i.i.d. case, ergodic diffusion one can replace $I_t(\theta)$ by another process equivalent to them (see examples).

To give an explicit form to the SDE (3.1) for the statistical model associated with the semimartingale X assume for a moment that for each (u,θ) (including the case $u=\theta$)

$$|\Phi(u)| * \mu \in \mathcal{A}^+_{\mathrm{loc}}(P_\theta). \tag{3.2}$$

Then for each pair (u,θ) we have

$$\Phi(u)*(\mu-\nu_u) = \Phi(u)*(\mu-\nu_\theta) + \Phi(u)\left(1-\frac{Y(u)}{Y(\theta)}\right)*\nu_\theta .$$

Based on this equality one can obtain the canonical decomposition of P_θ-special semimartingale $L(u)$ (w.r.t. measure P_θ):

$$L(u) = \dot{\beta}(u)\circ(X^c-\beta(\theta)\circ C) + \Phi(u)*(\mu-\nu_\theta) + \dot{\beta}(u)(\beta(\theta)-\beta(u))\circ C + \Phi(u)\left(1-\frac{Y(u)}{Y(\theta)}\right)*\nu_\theta . \tag{3.3}$$

Now, using (3.3) the meaning of $L(dt,u_t)$ is

$$\begin{aligned}\int_0^t L(ds,u_{s-}) &= \int_0^t \dot{\beta}_s(u_{s-})d(X^c-\beta(\theta)\circ C)_s + \int_0^t\int \Phi(s,x,u_{s-})(\mu-\nu_\theta)(ds,dx)\\ &\quad + \int_0^t \dot{\beta}_s(u_s)(\beta_s(\theta)-\beta_s(u_s))dC_s + \int_0^t\int \Phi(s,x,u_{s-})\left(1-\frac{Y(s,x,u_{s-})}{Y(s,x,\theta)}\right)\nu_\theta(ds,dx).\end{aligned}$$

Finally, the recursive SDE (3.1) takes the form

$$\begin{aligned}\theta_t &= \theta_0 + \int_0^t \gamma_s(\theta_{s-})\dot{\beta}_s(\theta_{s-})d(X^c-\beta(\theta)\circ C)_s + \int_0^t\int \gamma_s(\theta_{s-})\Phi(s,x,\theta_{s-})(\mu-\nu_\theta)(ds,dx)\\ &\quad + \int_0^t \gamma_s(\theta)\dot{\beta}_s(\theta_s)(\beta_s(\theta)-\beta_s(\theta_s))dC_s\\ &\quad + \int_0^t\int \gamma_s(\theta_{s-})\Phi(s,x,\theta_{s-})\left(1-\frac{Y(s,x,\theta_{s-})}{Y(s,x,\theta)}\right)\nu_\theta(ds,dx).\end{aligned} \tag{3.4}$$

Remark 3.1. One can give more accurate than (3.2) sufficient conditions (see, e.g., [21,19]) to ensure the validity of decomposition (3.3).

Assume that there exists a unique strong solution $(\theta_t)_{t\ge 0}$ of the SDE (3.4).

Fix arbitrary $\theta\in R^1$. To investigate the asymptotic properties, under measure P_θ, of recursive estimators $(\theta_t)_{t\ge 0}$ as $t\to\infty$, namely, a strong consistency, rate of convergence and asymptotic expansion we reduce the SDE (3.4) to the Robbins–Monro type SDE.

For this aim denote $z_t=\theta_t-\theta$. Then (3.4) can be rewritten as

$$\begin{aligned}z_t &= z_0 + \int_0^t \gamma_s(\theta+z_{s-})\dot{\beta}(\theta+z_{s-})(\beta_s(\theta)-\beta_s(\theta+z_{s-}))dC_s\\ &\quad + \int_0^t\int \gamma_s(\theta+z_{s-})\Phi(s,x,\theta+z_{s-})\left(1-\frac{Y(s,x,\theta+z_{s-})}{Y(s,x,\theta)}\right)\nu_\theta(ds,dx)\end{aligned}$$

$$+ \int_0^t \gamma_s(\theta + z_s)\dot{\beta}_s(\theta + z_s)d(X^c - \beta(\theta) \circ C)_s$$
$$+ \int_0^t \int \gamma_s(\theta + z_{s-})\Phi(s, x, \theta + z_{s-})(\mu - \nu_\theta)(ds, dx). \tag{3.5}$$

For the definition of the objects K^θ, $\{H^\theta(u),\ u \in R^1\}$ and $\{M^\theta(u),\ u \in R^1\}$ we consider such a version of characteristics (C, ν_θ) that

$$C_t = c^\theta \circ A_t^\theta,$$
$$\nu_\theta(\omega, dt, dx) = dA_t^\theta B_{\omega,t}^\theta(dx),$$

where $A^\theta = (A_t^\theta)_{t\ge 0} \in \mathcal{A}_{\mathrm{loc}}^+(P_\theta)$, $c^\theta = (c_t^\theta)_{t\ge 0}$ is a nonnegative predictable process, and $B_{\omega,t}^\theta(dx)$ is a transition kernel from $(\Omega \times R_+, \mathcal{P})$ in $(R, \mathcal{B}(R))$ with $B_{\omega,t}^\theta(\{0\}) = 0$ and

$$\Delta A_t^\theta B_{\omega,t}^\theta(R) \le 1$$

(see [21, Ch. 2, §2, Prop. 2.9]).

Put $K_t^\theta = A_t^\theta$,

$$H_t^\theta(u) = \gamma_t(\theta + u)\left\{\dot{\beta}_t(\theta + u)(\beta_t(\theta) - \beta_t(\theta + u))c_t^\theta + \int \Phi(t, x, \theta + u)\left(1 - \frac{Y(t, x, \theta + u)}{Y(t, x, \theta)}\right)B_{\omega,t}^\theta(dx)\right\}, \tag{3.6}$$

$$M^\theta(t, u) = \int_0^t \gamma_s(\theta + u)\dot{\beta}_s(\theta + u)d(X^c - \beta(\theta) \circ C)_s + \int_0^t \int \gamma_s(\theta + u)\Phi(s, x, \theta + u)(\mu - \nu_\theta)(ds, dx). \tag{3.7}$$

Assume that for each u, $u \in R$, $M^\theta(u) = (M^\theta(t, u))_{t\ge 0} \in \mathcal{M}_{\mathrm{loc}}^2(P_\theta)$. Then

$$\langle M^\theta(u)\rangle_t = \int_0^t (\gamma_s(\theta + u)\dot{\beta}_s(\theta + u))^2 c_s^\theta dA_s^\theta + \int_0^t \gamma_s^2(\theta + u)\left(\int \Phi^2(s, x, \theta + u)B_{\omega,s}^\theta(dx)\right)dA_s^{\theta,c}$$
$$+ \int_0^t \gamma_s^2(\theta + u)B_{\omega,t}^\theta(R)\left\{\int \Phi^2(s, x, \theta + u)q_{\omega,s}^\theta(dx)\right.$$
$$\left. - a_s(\theta)\left(\int \Phi(s, x, \theta + u)q_{\omega,s}^\theta(dx)\right)^2\right\}dA_s^{\theta,d},$$

where $a_s(\theta) = \Delta A_s^\theta B_{\omega,s}^\theta(R)$, $q_{\omega,s}^\theta(dx)I_{\{a_s(\theta)>0\}} = \frac{B_{\omega,s}^\theta(dx)}{B_{\omega,s}^\theta(R)} I_{\{a_s(\theta)>0\}}$.

Now we give a more detailed description of $\Phi(\theta)$, $I(\theta)$, $H^\theta(u)$ and $\langle M^\theta(u)\rangle$. This allows us to study the special cases separately (see Remark 3.2 below). Denote

$$\frac{d\nu_\theta^c}{d\nu^c} := F(\theta), \qquad \frac{q_{\omega,t}^\theta(dx)}{q_{\omega,t}(dx)} := f_{\omega,t}(x, \theta) \quad (:= f_t(\theta)).$$

Then

$$Y(\theta) = F(\theta)I_{\{a=0\}} + \frac{a(\theta)}{a} f(\theta)I_{\{a>0\}}$$

and

$$\dot{Y}(\theta) = \dot{F}(\theta)I_{\{a=0\}} + \left(\frac{\dot{a}(\theta)}{a} f(\theta) + \frac{a(\theta)}{a}\dot{f}(\theta)\right)I_{\{a>0\}}.$$

Therefore

$$\Phi(\theta) = \frac{\dot{F}(\theta)}{F(\theta)} I_{\{a=0\}} + \left\{\frac{\dot{f}(\theta)}{f(\theta)} + \frac{\dot{a}(\theta)}{a(\theta)(1 - a(\theta))}\right\}I_{\{a>0\}} \tag{3.8}$$

with $I_{\{a(\theta)>0\}} \int \frac{\dot{f}(\theta)}{f(\theta)} q^\theta(dx) = 0$.

Remark 3.2. Denote $\dot{\beta}(\theta) = \ell^c(\theta)$, $\frac{\dot{F}(\theta)}{F(\theta)} := \ell^\pi(\theta)$, $\frac{\dot{f}(\theta)}{f(\theta)} := \ell^\delta(\theta)$, $\frac{\dot{a}(\theta)}{a(\theta)(1-a(\theta))} := \ell^b(\theta)$.

Indices $i = c, \pi, \delta, b$ carry the following loads: "c" corresponds to the continuous part, "π" to the Poisson type part, "δ" to the predictable moments of jumps (including a main special case—the discrete time case), "b" to the binomial type part of the likelihood score $\ell(\theta) = (\ell^c(\theta), \ell^\pi(\theta), \ell^\delta(\theta), \ell^b(\theta))$.

In these notations we have for the Fisher information process:

$$I_t(\theta) = \int_0^t (\ell_s^c(\theta))^2 dC_s + \int_0^t \int (\ell_s^\pi(x;\theta))^2 B^\theta_{\omega,s}(dx) dA_s^{\theta,c} + \int_0^t B^\theta_{\omega,s}(R)\Big[\int (\ell_s^\delta(x;\theta))^2 q^\theta_{\omega,s}(dx)\Big] dA_s^{\theta,d} + \int_0^t (\ell_s^b(\theta))^2(1-a_s(\theta)) dA_s^{\theta,d}. \tag{3.9}$$

For the random field $H^\theta(u)$ we have

$$\begin{aligned} H_t^\theta(u) = \gamma_t(\theta+u)\Big\{&\ell_t^c(\theta+u)(\beta_t(\theta)-\beta_t(\theta+u))c_t^\theta \\ &+ \int \ell_t^\pi(x;\theta+u)\Big(1-\frac{F_t(x;\theta+u)}{F_t(x;\theta)}\Big)\Big\} B^\theta_{\omega,t}(dx) I_{\{\Delta A_t^\theta=0\}} \\ &+ \Big\{\int \ell_t^\delta(x;\theta+u) q^\theta_{\omega,t}(dx) \ell_t^b(\theta+u)\,\frac{a_t(\theta)-a_t(\theta+u)}{a_t(\theta)}\Big\} B^\theta_{\omega,t}(R) I_{\{\Delta A_t^\theta>0\}}. \end{aligned} \tag{3.10}$$

Finally, we have for $\langle M^\theta(u)\rangle$:

$$\begin{aligned} \langle M^\theta(u)\rangle_t = \big(\gamma(\theta+u)\ell^c(\theta+u)\big)^2 c^\theta \circ A_t^\theta &+ \int_0^t \gamma_s^2(\theta+u)\int (\ell_s^\pi(x;\theta+u))^2 B^\theta_{\omega,s}(dx) dA_s^{\theta,c} \\ &+ \int_0^t \gamma_s^2(\theta+u) B^\theta_{\omega,s}(R)\Big\{\int (\ell_s^\delta(x;\theta+u)+\ell_s^b(\theta+u))^2 q^\theta_{\omega,s}(dx) \\ &- a_s(\theta)\Big(\int (\ell_s^\delta(x;\theta+u)+\ell_s^b(\theta+u)) q^\theta_{\omega,s}(dx)\Big)^2\Big\} dA_s^{\theta,d}. \end{aligned} \tag{3.11}$$

Thus, we reduced SDE (3.5) to the Robbins–Monro type SDE with $K_t^\theta = A_t^\theta$, and $H^\theta(u)$ and $M^\theta(u)$ defined by (3.6) and (3.7), respectively.

As it follows from (3.6), (3.10)

$$H_t^\theta(0) = 0 \quad \text{for all } t \ge 0,\ P_\theta\text{-a.s.}$$

As for condition (A) to be satisfied it is enough to require that for all $t \ge 0$, $u \ne 0$ P_θ-a.s.

$$\begin{aligned} &\dot{\beta}_t(\theta+u)(\beta_t(\theta)-\beta_t(\theta+u)) < 0, \\ &\Big(\int \frac{\dot{F}(t,x,\theta+u)}{F(t,x,\theta+u)}\Big(1-\frac{F(t,x;\theta+u)}{F(t,x;\theta)}\Big) B^\theta_{\omega,t}(dx)\Big) I_{\{\Delta A_t^\theta=0\}} u < 0, \\ &\Big(\int \frac{\dot{f}(t,x;\theta+u)}{f(t,x;\theta+u)} q_t^\theta(dx)\Big) I_{\{\Delta A_t^\theta>0\}} u < 0, \\ &\dot{a}_t(\theta+u)(a_t(\theta)-a_t(\theta+u))u < 0, \end{aligned}$$

and the simplest sufficient conditions for the latter ones are the strong monotonicity (P-a.s.) of functions $\beta(\theta)$, $F(\theta)$, $f(\theta)$ and $a(\theta)$ w.r.t. θ.

4. Main results

We are ready to formulate main results about asymptotic properties of recursive estimators $\{\theta_t,\ t \ge 0\}$ as $t \to \infty$, (P_θ-a.s.), which is the same of solution z_t, $t \ge 0$, of Eq. (3.5).

For simplicity we restrict ourselves by the case when semimartingale $X = (X_t)_{t\geq 0}$ is left quasi-continuous, so $\nu(\omega; \{t\}, R) = 0$ for all $t \geq 0$, P-a.s., and $A^\theta = (A_t^\theta)_{t\geq 0}$ is a continuous process. In this case

$$H_t^\theta(u) = \gamma_t(\theta+u)\left\{\dot{\beta}_t(\theta+u)(\beta_t(\theta)-\beta_t(\theta+u))c_t^\theta + \int \frac{\dot{F}_t(x;\theta+u)}{F_t(x;\theta+u)}\left(1-\frac{\dot{F}_t(x;\theta+u)}{F_t(x;\theta)}\right)B_{\omega,t}^\theta(dx)\right\}, \tag{4.1}$$

$$\langle M^\theta(u)\rangle_t = \int_0^t (\gamma_s(\theta+u)\dot{\beta}_s(\theta+u))^2 dA_s^\theta + \int_0^t \gamma_s^2(\theta+u)\left(\int\left(\frac{\dot{F}_s(x;\theta+u)}{F_s(x;\theta+u)}\right)^2 B_{\omega,s}^\theta(dx)\right)dA_s^\theta, \tag{4.2}$$

$$I_t(\theta) = \int_0^t (\dot{\beta}_s(\theta))^2 c_s^\theta dA_s^\theta + \int_0^t \int \left(\frac{\dot{F}_s(x;\theta)}{F_s(x;\theta)}\right)^2 B_{\omega,s}(dx)dA_s^\theta. \tag{4.3}$$

Theorem 4.1 (*Strong Consistency*)**.** *Let for all $t \geq 0$, P_θ-a.s. the following conditions be satisfied:*

(A) $H_t^\theta(0) = 0$, $H_t^\theta(u)u < 0$, $u \neq 0$,
(B) $h_t^\theta(u) \leq B_t^\theta(1+u^2)$, *where $B^\theta = (B_t^\theta)_{t\geq 0}$ is a predictable process, $B_t^\theta \geq 0$, $B^\theta \circ A_\infty^\theta < \infty$,*

$$h_t^\theta(u) = \frac{d\langle M^\theta(u)\rangle_t}{dA_t^\theta}, \tag{4.4}$$

(C) *for each ε, $\varepsilon > 0$,*

$$\inf_{\varepsilon\leq|u|\leq\frac{1}{\varepsilon}} |H^\theta(u)u| \circ A_\infty^\theta = \infty.$$

Then for each $\theta \in R^1$

$$\widehat{\theta}_t \to 0 \quad (or \quad z_t \to 0), \quad as\ t \to \infty,\ P_\theta\text{-a.s.}$$

Proof. Immediately follows from conditions of Theorem 3.1 of [16] applied to prespecified by (4.1)–(4.3) objects. □

In the sequel we assume that for each $\theta \in R^1$

$$P_\theta\left(\lim_{t\to\infty}\frac{\widehat{I_t}(\theta)}{I_t(\theta)} = 1\right) = 1,$$

from which it follows that $\gamma_t(\theta) = I_t^{-1}(\theta)$. Denote

$$g_t^\theta = \frac{dI_t(\theta)}{dA_t^\theta} = (\dot{\beta}_t(\theta))^2 c_t^\theta + \int\left(\frac{\dot{F}_t(x;\theta)}{F_t(x;\theta)}\right)^2 B_{\omega,t}(dx). \tag{4.5}$$

We assume also that $z_t \to 0$ as $t \to \infty$, P_θ-a.s.

Theorem 4.2 (*Rate of Convergence*)**.** *Suppose that for each δ, $0 < \delta < 1$, the following conditions are satisfied:*

$$\text{(i)} \quad \int_0^\infty \left[\delta\frac{g_t^\theta}{I_t^\theta} - 2\beta_t^\theta(z_t)\right]^+ dA_t^\theta < \infty, \quad P_\theta\text{-a.s., where } \beta_t^\theta(u) = \begin{cases} -\dfrac{H_t^\theta(u)}{u}, & u \neq 0,\\ -\lim\limits_{u\to 0}\dfrac{H_t^\theta(u)}{u}, & u = 0,\end{cases} \tag{4.6}$$

$$\text{(ii)} \quad \int_0^\infty (I_t(\theta))^\delta h_t^\theta(z_t)dA_t^\theta < \infty, \quad P_\theta\text{-a.s.}$$

Then for each $\theta \in R^1$, δ, $0 < \delta < 1$,

$$I_t^\delta(\theta)z_t^2 \to 0 \quad as\ t \to \infty,\ P_\theta\text{-a.s.}$$

Proof. It is enough to note that conditions (2.3) and (2.4) of Theorem 2.1 from [17] are satisfied with $I_t(\theta)$ instead of γ_t, $\delta g_t^\theta / I_t(\theta)$ instead of r_t^δ and $\beta_t^\theta(u)$ instead of $\beta_t(u)$. $\square$

In the sequel we assume that for all δ, $0 < \delta < \frac{1}{2}$,

$$I_t^\delta(\theta) z_t \to 0 \quad \text{as } t \to \infty, \ P_\theta\text{-a.s.}$$

It is not hard to verify that the following expansion holds true

$$I_t^{1/2}(\theta) z_t = \frac{L_t^\theta}{\langle L^\theta \rangle_t^{1/2}} + R_t^\theta, \tag{4.7}$$

where L_t^θ, R_t^θ will be specified below.

Indeed, according to "Preliminary and Notation" section of [17]

$$\overline{\beta}_t^\theta = -\lim_{u \to 0} \frac{H_t^\theta(u)}{u} = -I_t^{-1}(\theta) g_t^\theta.$$

Further,

$$-\overline{\beta}^\theta \circ A_t^\theta = \int_0^t I_s^{-1}(\theta) \frac{dI_s(\theta)}{dA_s(\theta)} \, dA_s^\theta = \ln I_t(\theta).$$

Therefore

$$\Gamma_t^\theta = \varepsilon_t^{-1}(-\overline{\beta}^\theta \circ A_t^\theta) = I_t(\theta) \tag{4.8}$$

and

$$L_t^\theta = \int_0^t \Gamma_s^\theta dM^\theta(s, 0)$$

with

$$\langle L^\theta \rangle_t = \int_0^t (\Gamma_s^\theta)^2 d\langle M^\theta(0) \rangle_s = \int_0^t I_s^2(\theta) I_s^{-2}(\theta) dI_s(\theta) = I_t(\theta). \tag{4.9}$$

Finally, we obtain

$$\chi_t^\theta = \Gamma_t^\theta \langle L^\theta \rangle_t^{-1/2} = I_t^{1/2}(\theta). \tag{4.10}$$

As for R_t^θ, one can use the definition of R_t from the same section by replacing of objects by the corresponding objects with upperscripts "θ", e.g. $\overline{\beta}_t$ by $\overline{\beta}_t^\theta$, L_t by L_t^θ, etc.

Theorem 4.3 (*Asymptotic Expansion*)**.** *Let the following conditions be satisfied:*

(i) $\langle L^\theta \rangle_t$ *is a deterministic process,* $\langle L^\theta \rangle_\infty = \infty$,

(ii) *there exists* ε, $0 < \varepsilon < \frac{1}{2}$, *such that*

$$\frac{1}{\langle L^\theta \rangle_t} \int_0^t |\beta_s^\theta - \beta_s^\theta(z_s)| I_s^{-\varepsilon}(\theta) \langle L^\theta \rangle_s dA_s^\theta \to 0 \quad \text{as } t \to \infty, \ P_\theta\text{-a.s.},$$

(iii)

$$\frac{1}{\langle L^\theta \rangle_t} \int_0^t I_t^2(\theta)(h_s^\theta(z_s, z_s) - 2h_s^\theta(z_s, 0) + h_s(0, 0)) dA_s^\theta \xrightarrow{P_\theta} 0 \quad \text{as } t \to \infty,$$

where

$$h_t^\theta(u, v) = \frac{d\langle M^\theta(u), M^\theta(v) \rangle}{dA_t^\theta}. \tag{4.11}$$

Then in Eq. (4.7) *for each* $\theta \in R$

$$R_t^\theta \xrightarrow{P_\theta} 0 \quad as\ t \to \infty.$$

Proof. It is not hard to verify that all conditions of Theorem 3.1 from [17] are satisfied with $\langle L^\theta \rangle_t$ instead of $\langle L \rangle_t$, $\beta_s^\theta(u)$ instead of $\beta_s(u)$, $I_\theta^{-1}(\theta)$ instead of γ_t, A_t^θ instead of χ_t, Γ_s^θ instead Γ_s, and $I_t^{1/2}(\theta)$ instead of χ_t, $h_t^\theta(u, v)$ instead of $h_t(u, v)$, and, finally, P^θ instead of P. $\square$

Remark. It follows from Eq. (4.7) and Theorem 4.3 that, using the Central Limit Theorem for martingales

$$I_t^{1/2}(\theta)(\theta_t - \theta) \xrightarrow{d} N(0, 1).$$

5. Recursive procedure for M-estimators

As stated in previous section the maximum likelihood equation has the form

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

This equation is the special member of the following family of estimational equations

$$L_t(m_\theta, M_\theta) = 0 \tag{5.1}$$

with certain P-martingales m_θ, $\theta \in R_1$. These equations are of the following sense: their solutions are viewed as estimators of unknown parameter θ, so-called M-estimators. 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.

As it is well known M-estimators play the important role in robust statistics, besides they are sources to obtain asymptotically normal estimators.

Since for each $\theta \in R_1 P_\theta$ is a unique measure such that under this measure $X = (X_t)_{t\geq 0}$ is a semimartingale with characteristics $(B(\theta), c(\theta), \nu_\theta)$ all P_θ-martingales admit an integral representation property w.r.t. continuous martingale part and martingale measure $(\mu - \nu_\theta)$ of X. In particular, the P-martingale M_θ has the form (see Eq. (2.2))

$$M_\theta = \beta(\theta) \circ X^s + \psi * (\mu - \nu), \tag{5.2}$$

where

$$\psi(s, x, \theta) = Y(t, x, \theta) - 1 + \frac{\widehat{Y}(t, \theta) - a}{1 - a} I_{(0<a<1)}$$

and $m_\theta \in \mathcal{M}_{\text{loc}}(P)$ can be represented as

$$m(\theta) = g(\theta) \circ X^c + G(\theta) * (\mu - \nu) \tag{5.3}$$

with certain functions $g(\theta)$ and $G(\theta)$.

It can be easily shown that P_θ-martingale $L(m_\theta, M_\theta)$ can be represented as

$$L(m_\theta, M_\theta) = \varphi_m(\theta) \cdot (X^c - \beta(\theta) \circ C) + \Phi_m(\theta) * (\mu - \nu_\theta), \tag{5.4}$$

where the functions φ_m and Φ_m are expressed in terms of functions $\beta(\theta)$, $\psi(\theta)$, $g(\theta)$ and $G(\theta)$.

On the other hand, it can be easily shown that each P_θ-martingale $\widetilde{M}_\theta$ can be expressed as $L(\widetilde{m}_\theta, M_\theta)$ with P-martingale $\widetilde{m}_\theta$ defined as

$$\widetilde{m}_\theta = L(\widetilde{M}_\theta, L(-M_\theta, M_\theta)) \in \mathcal{M}_{\text{loc}}(P)$$

(since $\frac{dP}{dP_\theta} = \mathcal{E}(L(-M_\theta, M_\theta))$, according to the generalized Girsanov theorem $L(\widetilde{M}_\theta, L(-M_\theta, M_\theta)) \in \mathcal{M}_{\text{loc}}(P)$).

Therefore without loss of generality one can consider the M-estimator associated with the parametric family $(\widetilde{M}_\theta,\ \theta \in R)$ of P_θ-martingale as the solution of the estimational equation

$$\widetilde{M}_t(\theta) = 0. \tag{5.5}$$

In the sequel we assume that for each $\theta \in R_1$, $\widetilde{M}_\theta \in \mathcal{M}^2_{\text{loc}}(P_\theta)$. Assume also that there exists a positive decreasing predictable process $\widetilde{\gamma}_t(\theta)$ with $\widetilde{\gamma}_0(\theta) = 1$ such that $\widetilde{\gamma}_t(\theta)\langle \widetilde{M}_\theta\rangle_t \xrightarrow{P_\theta} 1$ as $t \to \infty$.

Now using the same arguments as in Section 3 we introduce the following recursive procedure for constructing estimator $(\widetilde{\theta}_t,\ t \geq 0)$ asymptotically equivalent to the M-estimator defined by relation (5.5) as the solution of the following SDE

$$d\widetilde{\theta}_t = \widetilde{\gamma}_t(\theta)\widetilde{M}(dt, \widetilde{\theta}_{t-}). \tag{5.6}$$

To obtain the explicit form of the last SDE, recall that $\widetilde{M}_\theta$ has an integral representation property

$$\widetilde{M}_t(\theta) = \widetilde{\varphi}(\theta) \circ (X^c - \beta(\theta) \circ \langle X^c\rangle) + \widetilde{\Phi}(\theta) * (\mu - \nu_\theta).$$

We can obtain the canonical decomposition of P_θ-semimartingale $\widetilde{M}_t(u)$, $u \in R^1$ (w.r.t. measure P_θ)

$$\begin{aligned}\widetilde{M}(u) &= \widetilde{\varphi}(u) \circ (X^c - \beta(\theta) \circ C) + \widetilde{\Phi}(u) * (\mu - \nu_\theta)\\ &\quad + [\widetilde{\varphi}(u)(\beta(\theta) - \beta(u))] \circ C + \widetilde{\Phi}(u)\left(1 - \frac{y(u)}{y(\theta)}\right) * (\mu - \nu_\theta).\end{aligned}$$

Based on the last expression we can derive the explicit form of SDE (5.5)

$$\begin{aligned}\theta_t &= \theta_0 + \int_0^t \widetilde{\gamma}_s(\widetilde{\theta}_{s-})\widetilde{\varphi}(s, \theta_{s-})d(X^c - \beta(\theta) \circ C) + \int_0^t \int \widetilde{\gamma}_s(\theta_{s-})\widetilde{\Phi}(s, x, \widetilde{\theta}_{s-})(\mu - \nu_\theta)(ds, dx)\\ &\quad + \int_0^t \widetilde{\gamma}_s(\theta_{s-})\widetilde{\varphi}(s, \widetilde{\theta}_{s-})(\beta_s(\theta) - \beta_s(\theta_{s-}))dC_s\\ &\quad + \int_0^t \int \gamma_s(\theta_{s-})\widetilde{\Phi}(s, x, \widetilde{\theta}_{s-})\left(1 - \frac{Y(s, x, \widetilde{\theta}_{s-})}{Y(s, x, \theta)}\right)\nu_\theta(ds, dx).\end{aligned} \tag{5.7}$$

To study the asymptotic properties of the solution of this equation $(\widetilde{\theta}_t,\ t \geq 0)$ (e.g. consistency, rate of convergence, asymptotic normality) is more convenient to rewrite this equation as $(z_t = \widetilde{\widetilde{\theta}}_t - \theta)$

$$\begin{aligned}z_t &= z_0 + \int_0^t \widetilde{\gamma}_s(\theta + z_{s-})\widetilde{\varphi}(s, \theta + z_{s-})d(X^c - \beta(\theta) \circ C)\\ &\quad + \int_0^t \int \widetilde{\gamma}_s(\theta + z_{s-})\widetilde{\Phi}(s, x, \theta + z_{s-})(\mu - \nu_\theta)(ds, dx)\\ &\quad + \int_0^t \widetilde{\gamma}_s(\theta + z_{s-})\widetilde{\varphi}(s, \theta + z_{s-})(\beta_s(\theta) - \beta_s(\theta_s + z_{s-}))dC_s\\ &\quad + \int_0^t \int \widetilde{\gamma}_s(\theta + z_{s-})\widetilde{\Phi}(s, x, \theta + z_{s-})\left(1 - \frac{Y(s, x, \theta + z_{s-})}{Y(s, x, \theta)}\right)\nu_\theta(ds, dx).\end{aligned} \tag{5.8}$$

6. Examples

To make the things more clear let us begin with the simplest case of i.i.d. observations.

Example 1. Let $\{p_\theta,\ \theta \in R_1\}$ be the family of probability measures defined on some measurable space $(X, \mathcal{B})$ such that for each pair θ, θ', $p_\theta \sim p_{\theta'}$.

Put $\Omega = X^\infty$, $\mathcal{F}_n = \mathcal{B}(X^n)$, $\mathcal{F} = \mathcal{B}(X^\infty)$, $P_\theta = p_\theta \times p_\theta \times \cdots$. Then for θ, θ', $P_\theta \overset{\text{loc}}{\sim} P_{\theta'}$. Fix some $\theta_0 \in R_1$ and denote $p = p_{\theta_0}$. Let $dp_\theta/dp = f(x, \theta)$. Then the local density process

$$\rho_n(\theta) = \frac{dP_{n,\theta}}{dP_n} = \prod_{i=1}^n f(X_i, \theta) = \mathcal{E}_n(M_\theta), \tag{6.1}$$

where

$$M(\theta) = \sum_{i=1}^{n} (f(X_i, \theta) - 1)$$

is a P-martingale. Here $(X_n)_{n \geq 1}$ is a coordinate process, $X_n(\omega) = x_n$.

Assume that for all x, $f(x, \theta)$ is continuous differentiable in θ and denote $\frac{\partial}{\partial\theta} f(X, \theta) = \dot{f}(X, \theta)$. Assume also that $\frac{\partial}{\partial\theta} \int f(x, \theta) p(dx) = \int \dot{f}(x, \theta) p(dx)$. Then $\dot{M}_n(\theta) = \sum_{i=1}^{n} \dot{f}(X_i, \theta)$ is a P-martingale.

In these notation the MLE takes the form

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

The Fischer information process

$$I_n(\theta) = \langle L(\dot{M}_\theta, M_\theta) \rangle = nI(\theta), \tag{6.2}$$

where $I(\theta) = E_\theta \big(\frac{\dot{f}(\cdot, \theta)}{f(\cdot, \theta)} \big)^2$, assuming that the last integral is finite.

The recursive estimation procedure to obtain the estimator θ_n, asymptotically equivalent to MLE is well known:

$$\theta_n = \theta_{n-1} + \frac{1}{nI(\theta_{n-1})} \frac{\dot{f}(X_n, \theta_{n-1})}{f(X_n, \theta_{n-1})}. \tag{6.3}$$

Let us derive this equation from the general recursive SDE.

For this aim consider the process $S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$. This process is a semimartingale with the jump measure

$$\mu(\omega, [0, n] \times B) = \sum_{i \leq n} I_{\{X_i \in B\}}$$

and its P_θ-compensator is

$$\nu_\theta(\omega, [0, n] \times B) = \sum_{i \leq n} P_\theta(X_i \in B) = n \int_B f(x, \theta) p(dx).$$

Note that $a_n(\theta) = \nu(\omega, \{n\}; X) = 1$ for all $n \geq 1$ and $\theta \in R_1$.

It is obvious that $\nu_\theta = Y \cdot \nu$, where $Y_\theta(\omega, n, x) \equiv f(x, \theta)$. Besides,

$$\Phi(\theta) = \frac{\dot{Y}(\theta)}{Y(\theta)} + \frac{\dot{a}(\theta)}{1 - a(\theta)} = \frac{\dot{f}(\cdot, \theta)}{f(\cdot, \theta)}.$$

At the same time the general recursive SDE for this special case can be written as

$$\theta_n = \theta_{n-1} + \frac{1}{nI(\theta_{n-1})} \frac{\dot{f}(x_n, \theta_{n-1})}{f(x_n, \theta_{n-1})} - \frac{1}{nI(\theta_{n-1})} \int \frac{\dot{f}(x, u)}{f(x, u)} \frac{f(x, u)}{f(x, \theta)} f(x, \theta)\, d\mu|_{u=\theta_{n-1}}.$$

But $\int \dot{f}(x, u)\, d\mu = 0$ and thus the last term equals zero and we come to Eq. (6.3).

In terms of $z_n = \theta_n - \theta$ Eq. (6.3) takes the form

$$z_n = z_{n-1} + \frac{1}{nI(\theta + z_{n-1})} b(\theta, z_{n-1}) + \frac{1}{nI(\theta + z_{n-1})} \Delta m_n,$$

where

$$b(\theta, u) = \int \frac{\dot{f}(x, u)}{f(x, u)} f(x, \theta)\, d\mu, \qquad \Delta m_n = \Delta m_n(u), \qquad \Delta m_n = \frac{\dot{f}(x, u)}{f(x, u)} - b(\theta, u).$$

Concerning M-estimators recall that by the definition the estimational equation is

$$L_n(m(\theta), M(\theta)) = 0, \tag{6.4}$$

where $m(\theta)$ is some P-martingale, $m_n(\theta) = \sum_{i \leq n} g(X_i, \theta)$ with $\int g(x, \theta)\, dp = 0$.

Eq. (6.4) can be written as

$$\sum_{i\le n}\frac{g(X_i,\theta)}{f(X_i,\theta)}=0.$$

Thus, without loss of generality, we can define M-estimator as the solution of the equation

$$\widetilde{M}_n(\theta)=\sum_{i\le n}\psi(X_i,\theta)=0, \tag{6.5}$$

where

$$\int\psi(x_i,\theta)f(x_i,\theta)\,\mu(dx)=0,\qquad \langle\widetilde{M}(\theta)\rangle_n=n\int\psi^2(x,\theta)f(x,\theta)\,\mu(dx)=nI_\psi(\theta).$$

Now using the same arguments as in the case of MLE we obtain the following recursive procedure for constructing the estimator asymptotically equivalent to the M-estimator defined by (6.5)

$$\theta_n=\theta_{n-1}+\frac{1}{nI_\psi(\theta_{n-1})}\,\psi(X_n,\theta_{n-1}).$$

Example 2. Discrete time case.

Let $X_0,X_1,\dots,X_n,\dots$ be observations taking values in some measurable space $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ such that the regular conditional densities of distributions (w.r.t. some measure p) $f_i(x_i,\theta|x_{i-1},\dots,x_0)$, $i\le n$, $n\ge 1$ exist, $f_0(x_0,\theta)\equiv f_0(x_0)$, $\theta\in R^1$ is the parameter to be estimated. Denote P_θ corresponding distribution on $(\Omega,\mathcal{F}):=(\mathcal{X}^\infty,\mathcal{B}(\mathcal{X}^\infty))$. Identify the process $X=(X_i)_{i\ge 0}$ with coordinate process and denote $\mathcal{F}_0=\sigma(X_0)$, $\mathcal{F}_n=\sigma\,(X_i,\ i\le n)$. If $\psi=\psi(X_i,X_{i-1},\dots,X_0)$ is a r.v., then under $E_\theta(\psi|\mathcal{F}_{i-1})$ we mean the following version of conditional expectation

$$E_\theta(\psi\mid\mathcal{F}_{i-1}):=\int\psi(z,X_{i-1},\dots,X_0)f_i(z,\theta\mid X_{i-1},\dots,X_0)\mu(dz),$$

if the last integral exists.

Assume that the usual regularity conditions are satisfied and denote

$$\frac{\partial}{\partial\theta}f_i(x_i,\theta\mid x_{i-1},\dots,x_0):=\dot{f}_i(x_i,\theta\mid x_{i-1},\dots,x_0),$$

the maximum likelihood scores

$$l_i(\theta):=\frac{\dot{f}_i}{f_i}(X_i,\theta\mid X_{i-1},\dots,X_0)$$

and the empirical Fisher information

$$I_n(\theta):=\sum_{i=1}^n E_\theta(l_i^2(\theta)\mid\mathcal{F}_{i-1}).$$

Denote also

$$b_n(\theta,u):=E_\theta(l_n(\theta+u)\mid\mathcal{F}_{n-1})$$

and indicate that for each $\theta\in R^1$, $n\ge 1$

$$b_n(\theta,0)=0\quad(P_\theta\text{-a.s.}). \tag{6.6}$$

Using the same arguments as in the case of i.i.d. observations we come to the following recursive procedure

$$\theta_n=\theta_{n-1}+I_n^{-1}(\theta_{n-1})l_n(\theta_{n-1}),\qquad \theta_0\in\mathcal{F}_0.$$

Fix θ, denote $z_n=\theta_n-\theta$ and rewrite the last equation in the form

$$\begin{aligned}&z_n=z_{n-1}+I_n^{-1}(\theta+z_{n-1})b_n(\theta,z_{n-1})+I_n^{-1}(\theta+z_{n-1})\Delta m_n,\\&z_0=\theta-\theta,\end{aligned} \tag{6.7}$$

where $\Delta m_n=\Delta m(n,z_{n-1})$ with $\Delta m(n,u)=l_n(\theta+u)-E_\theta(l_n(\theta+u)|\mathcal{F}_{n-1})$.

Note that the algorithm (6.7) is embedded in SDE (1.1) with

$$H_n(u) = I_n^{-1}(\theta+u)b_n(\theta,u) \in \mathcal{F}_{n-1}, \quad \Delta K_n = 1,$$
$$\Delta M(n,u) = I_n^{-1}(\theta+u)\Delta m(n,u).$$

This example clearly shows the necessity of consideration of random fields $H_n(u)$ and $M(n,u)$.

The discrete time case was considered by T. Sharia in [10,11].

Example 3. Recursive parameter estimation in the trend coefficient of a diffusion process.

Here we consider the problem of recursive estimation of the one-dimensional parameter in the trend coefficient of a diffusion process $\xi = \{\xi_t,\ t \geq 0\}$ with

$$d\xi_t = a(\xi_t,\theta)\,dt + \sigma(\xi_t)\,dw_t, \quad \xi_0, \tag{6.8}$$

where $w = \{w_t,\ t \geq 0\}$ is a standard Wiener process, $a(\cdot,\theta)$ is the known function, $\theta \in \Theta \subseteq R$ is a parameter to be estimated, Θ is some open subset of R, $\sigma^2(\cdot)$ is the known diffusion coefficient.

We assume that there exists a unique weak solution of Eq. (6.8).

For each $\theta \in \Theta$ denote by P^θ the distribution of the process ξ on $(C_{[0,\infty)}, \mathcal{B})$.

Let $X = \{X_t,\ t \geq 0\}$ be the coordinate process, that is, for each $x = \{x_t, t \geq 0\} \in C_{[0,\infty)},\ X_t(x) = x_t, t \geq 0$.

Fix some $\theta \in \Theta$ and assume that for each $\theta' \in \Theta$, $P^\theta \overset{(loc)}{\sim} P^{\theta'}$. Then the density process $\rho_t(X,\theta)$ can be written as

$$\rho_t(X,\theta) := \frac{dP_t^\theta}{dP_t^{\theta'}}(X) = \exp\left\{\int_0^t \frac{a(X_s,\theta)-a(X_s,\theta')}{\sigma(X_s)}\,\frac{(dX_s - a(X_s,\theta')ds)}{\sigma(X_s)}\right\}$$
$$- \frac{1}{2}\int_0^t \left(\frac{a(X_s,\theta)-a(X_s,\theta')}{\sigma(X_s)}\right)^2 ds.$$

Recall that if for all $t \geq 0$ P^θ-a.s.

$$\int_0^1 \sigma^2(X_s)\,ds < \infty, \tag{6.9}$$

then the process $\left\{X_t - \int_0^t a(X_s,\theta)\,ds,\ t \geq 0\right\} \in M^2_{\text{loc}}(P^\theta)$ with the square characteristic $\int_0^t \sigma^2(X_s)\,ds$.

Under suitable regularity conditions if we assume that for all $t \geq 0$ P^θ-a.s.

$$\int_0^t \left(\frac{\dot a(X_s,\theta)}{\sigma(X_s)}\right)^2 ds < \infty, \tag{6.10}$$

we will have

$$\left\{\frac{\partial}{\partial\theta}\ln\rho_t(X,\theta) = \int_0^t \left(\frac{\dot a(X_s,\theta)}{\sigma(X_s)}\right) d(X_s - a(X_s,\theta)ds),\ t \geq 0\right\} \in M^2_{\text{loc}}(P^\theta),$$

where $\dot a(\cdot,\theta)$ denotes the derivative of $a(\cdot,\theta)$ w.r.t. θ.

Below we assume that conditions (6.9) and (6.10) are satisfied.

Introduce the Fisher information process

$$I_t(\theta) = \int_0^t \left(\frac{\dot a(X_s,\theta)}{\sigma(X_s)}\right)^2 ds.$$

Then, according to Eq. (3.4), the SDE for constructing the recursive estimator $(\theta_t, t \geq 0)$ has the form

$$d\theta_t = I_t(\theta_t)\left[\frac{\dot a(X_t,\theta_t)}{\sigma^2(X_s)}\,dX_t^c + \frac{\dot a(X_t,\theta_t)}{\sigma^2(X_t)}\,(a(X_t,\theta)-a(X_t,\theta_t))\,dt\right]. \tag{6.11}$$

Fix some $\theta \in \Theta$. To study the asymptotic properties of the recursive estimator $\{\theta_t, t \geq 0\}$ as $t \to \infty$ under measure P^θ let us denote $z_t = \theta_t - \theta$ and rewrite (6.11) in the following form:

$$dz_t = I_t(\theta + z_t)\left[\frac{\dot{a}(X_t, \theta + z_t)}{\sigma^2(X_s)}\, dX_t^c + \frac{\dot{a}(X_t, \theta + z_t)}{\sigma^2(X_t)}\,(a(X_t, \theta) - a(X_t, \theta + z_t))\, dt\right]. \tag{6.12}$$

In the sequel we assume that there exists a unique strong solution of Eq. (6.12) such that

$$\left\{\int_0^t I_s(\theta + z_s)\,\frac{\dot{a}(X_s, \theta + z_s)}{\sigma^2(X_s)}\, dX_s^c,\ t \geq 0\right\} \in M_{\mathrm{loc}}^2(P_\theta),$$

that is, for each $t \geq 0$ P^θ-a.s.

$$\int_0^t I_s^2(\theta + z_s)\left(\frac{\dot{a}(X_s, \theta + z_s)}{\sigma(X_s)}\right)^2 ds < \infty.$$

To study the asymptotic properties of the process $z = \{z_t,\ t \geq 0\}$ as $t \to \infty$ (under the measure P^θ) one can use the results of Theorems 4.1–4.3 concerning the asymptotic behavior of solutions of the Robbins–Monro type SDE

$$z_t = z_0 + \int_0^t H_s(z_{s-})\, dK_s + \int_0^t M(ds, z_{s-}). \tag{6.13}$$

Note that Eq. (6.13) covers Eq. (6.12) with $K_t = t$,

$$H_t(u) := H_t^\theta(u) = I_t(\theta + u)\,\frac{\dot{a}(X_t, \theta + u)}{\sigma^2(X_t)}\,(a(X_t, \theta) - a(X_t, \theta + u))\,, \quad H_t^\theta(0) = 0, \tag{6.14}$$

$$M(u) := M^\theta(u) = \left\{M^\theta(t, u) = \int_0^t I_s(\theta + u)\,\frac{\dot{a}(X_t, \theta + u)}{\sigma^2(X_t)}\, dX_s^c,\ t \geq 0\right\}. \tag{6.15}$$

Let for each $u \in R$ the process $M^\theta(u) \in M_{\mathrm{loc}}^2(P^\theta)$. Then

$$\langle M^\theta(u), M^\theta(v)\rangle_t = \int_0^t h_s(u, v)\, ds,$$

where

$$h_t(u, v) = h_t^\theta(u, v) = I_t(\theta + u) I_t(\theta + v)\,\frac{\dot{a}(X_t, \theta + u)\dot{a}(X_t, \theta + v)}{\sigma^2(X_t)}. \tag{6.16}$$

This problem is fully studied by Lazrieva and Toronjadze in [14].

Example 4. Let $(\Omega, \mathcal{F} = (\mathcal{F}_t)_{t\geq 0}, P, P_\theta, \theta \in R_1)$ be filtered probability space and $M = (M_t)_{t\geq 0}$ be a P-martingale with the deterministic characteristic $\langle M\rangle_t$, $\langle M\rangle_\infty = \infty$. Let for each $\theta \in R_1$ P_θ be unique measure on $(\Omega, \mathcal{F})$ such that the process $X(t)$ follows the equation

$$X_t = X_0 + a(\theta)\langle M\rangle_t + M_t,$$

where $a(\theta)$ is known function depending on the unknown parameter θ. Then for each pair (θ, θ'), $P_\theta \overset{\mathrm{loc}}{\sim} P_{\theta'}$. Fix some $\theta_0 \in R_1$. Then the local density process

$$\rho_t(\theta) = \frac{dP_{\theta,t}}{dP_{\theta_0,t}} = \mathcal{E}_t(M(\theta)),$$

where

$$M_t(\theta) = (a(\theta) - a(\theta_0))(X_t - a(\theta_0)\langle M\rangle_t). \tag{6.17}$$

Assume that $a(\theta)$ is strongly monotone function continuously differentiable in θ. Then

$$L_t(\theta) = \frac{\partial}{\partial\theta}\ln\rho_t(\theta) = L_t(\dot{M}(\theta), M(\theta)) = \dot{a}(\theta)(X_t - a(\theta)\langle M\rangle_t)$$

and the Fischer information process is

$$I_t(\theta) = \langle L(\theta), L(\theta)\rangle_t = [\dot a(\theta)]^2 \langle M\rangle_t.$$

Put $\gamma_t(\theta) = [\dot a(\theta)]^{-2} \frac{1}{\langle M\rangle_t + 1} = [\dot a(\theta)]^{-2}\gamma_t^{-1}$ (with the obvious notation $\gamma_t = \langle M\rangle_t + 1$). Therefore the recursive estimation procedure to obtain estimator asymptotically equivalent to the MLE θ_t is

$$\theta_t = \theta_0 + \int_0^t \frac{1}{\langle M\rangle_s + 1} \frac{a(\theta) - a(\theta_s)}{\dot a(\theta_s)} d\langle M\rangle_s + \int_0^t \frac{1}{1 + \langle M\rangle_s} \frac{1}{\dot a(\theta_s)} d(X_s - a(\theta)\langle M\rangle_s). \tag{6.18}$$

Denote $z_t = \theta_t - \theta$ and rewrite the last equation

$$dz_t = \frac{1}{\langle M\rangle_t + 1} \frac{a(\theta) - a(\theta + z_t)}{\dot a(\theta + z_t)} d\langle M\rangle_t + \frac{1}{\langle M\rangle_t + 1} \frac{1}{\dot a(\theta + z_t)} d(X_t - a(\theta)\langle M\rangle_t). \tag{6.19}$$

Further, denote

$$H_t(\theta, u) = \frac{1}{\langle M\rangle_t + 1} \frac{a(\theta) - a(\theta + z_t)}{\dot a(\theta + z_t)},$$

$$M_t(\theta, u) = \int_0^t \frac{1}{\langle M\rangle_s + 1} \frac{1}{\dot a(\theta + u)} d(X_s - a(\theta)\langle M\rangle_t).$$

In these notation Eq. (6.19) is the Robbins–Monro type equation

$$dz_t = H_t(\theta, z_t) d\langle M\rangle_t + dM_t(\theta, z_t). \tag{6.20}$$

Indeed, condition (A) of Theorem 4.1 is satisfied since

$$H_t(\theta, 0) = 0 \quad \text{and} \quad H_t(\theta, u)u < 0 \quad \text{for all } u \neq 0.$$

We study the asymptotic behavior of z_t as $t \to \infty$ under measure P_θ.

(1) Convergence: $z_t \to 0$ as $t \to \infty$ P_θ-a.s. or $\theta_t \to \theta$ as $t \to \infty$ P_θ-a.s. (strong consistency).

Proposition 6.1. *Let the following condition be satisfied*

$$[\dot a(\theta + u)]^2 (1 + u^2) \geq c, \tag{6.21}$$

where c is some constant depending on θ. Then

$$z_t \to 0 \quad \text{as } t \to \infty \ P_\theta\text{-a.s.}$$

Proof. Let us check conditions (A), (B), (C) of Theorem 4.1. (A) is evident. Concerning condition (B) note that

$$\langle M(\theta, u)\rangle_t = \frac{1}{(\dot a(\theta + u))^2} \int_0^t \frac{1}{(\langle M\rangle_s + 1)^2} d\langle M\rangle_s$$

and

$$h_t(\theta, u) = \frac{1}{(\dot a(\theta + u))^2} \frac{1}{(\langle M\rangle_t + 1)^2}.$$

Then if we denote $B_t = \frac{1}{(\langle M\rangle_t + 1)^2}$, taking into account Eq. (6.21) we simply obtain

$$h_t(\theta, u) \leq B_t(1 + u^2) \quad \text{with } B \circ \langle M\rangle_\infty < \infty.$$

As for condition (C), we have to verify that for each $\varepsilon > 0$

$$\inf_{\varepsilon \leq u \leq \frac{1}{\varepsilon}} \left| \frac{a(\theta) - a(\theta + u)}{\dot a(\theta + u)} \right| \int_0^\infty \frac{d\langle M\rangle_t}{\langle M\rangle_t + 1} = \infty.$$

The last condition is satisfied if for each $\varepsilon > 0$

$$\inf_{\varepsilon \leq |u| \leq \frac{1}{\varepsilon}} \left| \frac{a(\theta) - a(\theta + u)}{\dot a(\theta + u)} \right| > 0,$$

which holds since $\dot a(\theta)$ is continuous. $\square$

(2) Rate of convergence. Here we assume that $z_t \to 0$ as $t \to \infty$ P_θ-a.s.

Proposition 6.2. *For all δ, $0 < \delta < \frac{1}{2}$, we have*

$$\gamma_t^\delta z_t = (\langle M\rangle_t + 1)^\delta z_t \to 0 \quad \text{as } t \to \infty, \ P_\theta\text{-a.s.}$$

Proof. We have to check conditions (i) and (ii) of Theorem 4.2.

Condition (ii) is satisfied. Indeed, for all $0 < \delta < 1$

$$\int_0^\infty (\langle M\rangle_t + 1)^\delta [\dot a(\theta + u)]^{-2} \frac{1}{(\langle M\rangle_t + 1)^2} \, d\langle M\rangle_t < \infty.$$

As for condition (i), it is enough to verify that for all δ, $0 < \delta < \frac{1}{2}$,

$$\int_0^\infty \frac{1}{\langle M\rangle_t + 1} \left[\delta - I_{(z_t=0)} - \frac{a(\theta) - a(\theta + z_t)}{z_t \, \dot a(\theta + z_t)}\right]^+ d\langle M\rangle_t < \infty.$$

But $\left[\delta - I_{(z_t=0)} - \frac{a(\theta)-a(\theta+z_t)}{z_t \, \dot a(\theta+z_t)} I_{\{z_t \neq 0\}}\right]^+ = 0$ eventually since $z_t \to 0$. $\square$

(3) Asymptotic expansion. Here we assume that for all δ, $0 < \delta < \frac{1}{2}$, $\gamma_t^\delta z_t \to 0$ as $t \to \infty$ P_θ-a.s.

Proposition 6.3. *Let there exist some $\varepsilon > 0$, $\gamma > 0$ and $c(\theta)$ such that*

$$|\dot a(\theta + u) - \dot a(\theta + v)| \le c|u - v|^\gamma \tag{6.22}$$

for all $(u, v) \in O_\varepsilon(0)$, then all conditions of Theorem 4.3 are satisfied and the following asymptotic expansion holds true

$$(1 + \langle M\rangle_t)^{1/2} \dot a(\theta) z_t = \frac{L_t}{\langle L\rangle_t^{1/2}} + R_t,$$

where $R_t \to 0$ as $t \to \infty$ P-a.s., $L_t = [\dot a(\theta)]^{-1}(X_t - a(\theta)\langle M\rangle_t)$.

Example 5 (*Point Process with Continuous Compensator*)**.** Let Ω be a space of piecewise constant functions $x = (x_t)_{t\ge 0}$ such that $x_0 = 0$, $x_t = x_{t-} + (0 \text{ or } 1)$, $\mathcal{F} = \sigma\{x : x_s, s \ge 0\}$ and $\mathcal{F}_t = \sigma\{x : x_s, 0 < s \le t\}$. Let for $x \in \Omega$

$$\tau_n(x) = \inf\{s : s > 0, \ x_s = n\}$$

setting $\tau_n(\infty) = \infty$ if $\lim_{t\to\infty} x_t < n$. Let $\tau_\infty(x) = \lim_{n\to\infty} \tau_n(x)$.

Note that $x = (x_t)_{t\ge 0}$ can be written as

$$x_t = \sum_{n\ge 1} I_{\{\tau_n(x) \le t\}},$$

and so $(x_t)_{t\ge 0}$ and the family of σ-algebras $(\mathcal{F}_t)_{t\ge 0}$ are right-continuous.

Let for each $\theta \in R_1$ P_θ be a probability measure on $(\Omega, \mathcal{F})$ such that under this measure the coordinate process $X_t(\omega) = x_t$ if $\omega = (x_t)_{t\ge 0}$ is a point process with compensator $A_t(\theta) = A(\theta)A(t)$, where $A(t) = A(t, \omega)$ is an increasing process with continuous trajectories (P_θ-a.s.), $A(0) = 0$, $P_\theta\{A_\infty = \infty\} = 1$, and for each $t > 0$ $P_{\dot\theta}(A_t < \infty) = 1$, $A(\theta)$ is a strongly monotone deterministic function, $A(\theta) > 0$, and $A(\theta)$ is continuously differentiable (denote $\dot A(\theta) = \frac{d}{d\theta} A(\theta)$).

Assume that for each pair (θ, θ'), $P_\theta \overset{loc}{\sim} P_{\theta'}$. Fix as usual some $\theta_0 \in R_1$. Then the local density process $\rho_t(\theta) = \frac{dP_{\theta,t}}{dP_{\theta_0,t}}$ can be represented as

$$\rho_t(\theta) = \mathcal{E}_t(M(\theta)),$$

where

$$M_t(\theta) = \left(\frac{A(\theta)}{A(\theta_0)} - 1\right)(X_t - A(\theta_0)A_t).$$

Therefore $L_t(\theta) = \frac{\partial}{\partial\theta} \ln \rho_t(\theta)$ has the form

$$L_t(\theta) = L_t(\dot{M}(\theta), M(\theta)) = \frac{\dot{A}(\theta)}{A(\theta)} (X_t - A(\theta)A(t)).$$

The Fisher information process is

$$I_t(\theta) = \langle L(\dot{M}(\theta), M(\theta))\rangle_t = \left[\frac{\dot{A}(\theta)}{A(\theta)}\right]^2 A(\theta)A(t).$$

Put $\gamma_t(\theta) = \frac{A(\theta)}{[\dot{A}(\theta)]^2} \frac{1}{A(t)+1}$. It is evident that

$$\lim_{t\to\infty} \gamma_t(\theta)I_t(\theta) = 1.$$

Note that the process $(X_t)_{t\geq 0}$ is a P_θ-semimartingale with the triplet of characteristics $(A(\theta)A(t), 0, A(\theta)A(t))$. Therefore, according to Section 3,

$$F(\theta) = F(\omega, t, x, \theta) = \frac{A(\theta)}{A(\theta_0)}, \quad \Phi(\theta) = \frac{\dot{A}(\theta)}{A(\theta)},$$
$$\ell^c(\theta) = \ell^\delta(\theta) = \ell^b(\theta) = 0, \quad \ell^\pi(\theta) = \frac{\dot{A}(\theta)}{A(\theta)}.$$

Thus from (3.10) we obtain

$$H_t^\theta(u) = \frac{1}{A(t)+1} \frac{A(\theta) - A(\theta+u)}{\dot{A}(\theta+u)},$$
$$M^\theta(t, u) = \frac{1}{\dot{A}(\theta+u)} \int_0^t \frac{1}{A(s)+1} d(X_s - A(\theta)A(s)),$$

and the equation for $z_t = \theta_t - \theta$ is

$$dz_t = \frac{1}{A(t)+1} \frac{A(\theta) - A(\theta+z_t)}{\dot{A}(\theta+z_t)} dA(t) + \frac{1}{A(t)+1} \frac{1}{\dot{A}(\theta+z_t)} d(X_t - A(\theta)A(t)), \tag{6.23}$$

where $(\theta_t)_{t\geq 0}$ is recursive estimation satisfying the equation

$$d\theta_t = \frac{1}{A(t)+1} \frac{A(\theta) - A(\theta_t)}{\dot{A}(\theta_t)} dA(t) + \frac{1}{A(t)+1} \frac{1}{\dot{A}(\theta_t)} d(X_t - A(\theta)A(t)).$$

As one can see Eq. (6.23) is quite similar to (6.19) with $A(\theta)$ instead of $a(\theta)$ and $A(t)$ instead of $\langle M\rangle_t$.

Now if conditions (6.21) and (6.22) with $A(\theta)$ instead of $a(\theta)$ and $A(t)$ instead of $\langle M\rangle_t$ are satisfied, then the asymptotic expansion holds true

$$(A(t) + 1)^{1/2} \dot{A}(\theta) z_t = \frac{L_t}{\langle L\rangle_t^{1/2}} + R_t,$$

where $R_t \to 0$ as $t \to \infty$ P_θ-a.s., $L_t = [\dot{A}(\theta)]^{-1}(X_t - A(\theta)A(t))$.

References

[1] A.E. Albert, L.A. Gardner Jr., Stochastic Approximations and Nonlinear Regression, in: M.I.T. Press Research Monograph, No. 42, The M.I.T. Press, Cambridge, Mass., 1967.

[2] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Statist. 22 (1951) 400–407.

[3] A. Le Breton, About Gaussian schemes in stochastic approximation, Stochastic Process. Appl. 50 (1) (1994) 101–115.

[4] H.F. Chen, Asymptotically efficient stochastic approximation, Stoch. Stoch. Rep. 45 (1–2) (1993) 1–16.

[5] H.F. Chen, W. Zhao, Recursive Identification and Parameter Estimation, CRC Press, Boca Raton, FL, 2014.

[6] H.J. Kushner, G.G. Yin, Stochastic Approximation Algorithms and Applications, in: Applications of Mathematics (New York), vol. 35, Springer-Verlag, New York, 1997.

[7] H.J. Kushner, G.G. Yin, Stochastic approximation and recursive algorithms and applications, in: Applications of Mathematics (New York), second ed., in: Stochastic Modelling and Applied Probability, vol. 35, Springer-Verlag, New York, 2003.

[8] D. Levanony, A. Shwartz, O. Zeitouni, Recursive identification in continuous-time stochastic processes, Stochastic Process. Appl. 49 (2) (1994) 245–275.
[9] Y. Liang, A. Tovaneswaran, B. Abraham, Recent developments in recursive estimation for time-series models, Int. J. Adv. Stat. Probab. 5 (2) (2016).
[10] T. Sharia, On the recursive parameter estimation in the general discrete time statistical model, Stochastic Process. Appl. 73 (2) (1998) 151–172.
[11] T. Sharia, Recursive parameter estimation: convergence, Stat. Inference Stoch. Process. 11 (2) (2008) 157–175.
[12] P. Spreij, Recursive approximate maximum likelihood estimation for a class of counting process models, J. Multivariate Anal. 39 (2) (1991) 236–245.
[13] M.B. Nevelson, R.Z. Khas'minskiĭ, Stochastic Approximation and Recurrent Estimation, in: Monographs in Probability Theory and Mathematical Statistics, Nauka, Moscow, 1972 (in Russian).
[14] N.L. Lazrieva, T.A. Toronjadze, Ito–Ventzel's formula for semimartingales, asymptotic properties of MLE and recursive estimation, Stochastic differential systems, in: Proc. IFIP-WG 7/1 Work. Conf., Eisenach/GDR 1986, in: Lect. Notes Control Inf. Sci., vol. 96, 1987, pp. 346–355.
[15] A.V. Melnikov, È. Valkeĭla, Martingale models of stochastic approximation and their convergence, Teor. Veroyatn. Primen. 44 (2) (1999) 278–311 (in Russian); translation in Theory Probab. Appl. 44 (2000), no. 2, 330–360.
[16] N. Lazrieva, T. Sharia, T. Toronjadze, The Robbins–Monro type stochastic differential equations. I. Convergence of solutions, Stoch. Stoch. Rep. 61 (1–2) (1997) 67–87.
[17] N. Lazrieva, T. Sharia, T. Toronjadze, The Robbins–Monro type stochastic differential equations. II. Asymptotic behaviour of solutions, Stoch. Stoch. Rep. 75 (3) (2003) 153–180.
[18] N. Lazrieva, T. Sharia, T. Toronjadze, Semimartingale stochastic approximation procedure and recursive estimation. Martingale theory and its application, J. Math. Sci. (NY) 153 (3) (2008) 211–261.
[19] R.Sh. Liptser, A.N. Shiryayev, Martingale Theory, in: Probability Theory and Mathematical Statistics, Nauka, Moscow, 1986 (in Russian).
[20] L.I. Gal'čuk, On the existence and uniqueness of solutions of stochastic equations with respect to semimartingales, Teor. Veroyatn. Primen. 23 (4) (1978) 782–795 (in Russian).
[21] J. Jacod, A.N. Shiryaev, Limit Theorems for Stochastic Processes, in: Grundlehren der Mathematischen Wissenschaften ([Fundamental Principles of Mathematical Sciences]), vol. 288, Springer-Verlag, Berlin, 1987.
[22] N. Lazrieva, T. Toronjadze, Asymptotic theory of M-estimators in general statistical models. Parts I and II, Reports BS-R9010-20, Centrum Voor Wiskunde en Informatica, 1990.