Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

P.J.C. Spreij

Recursive parameter estimation for counting processes with linear intensity

Department of Operations Research and System Theory Report OS-R8513 November

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 Pure Research (Z.W.O.).

Recursive Parameter Estimation for Counting Processes with Linear Intensity

P.J.C. Spreij
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Recursive estimation algorithms are presented for counting processes that have an intensity process which is linear in the parameter. Strong consistency and asymptotic normality of the estimators generated by the algorithms are proved.

1980 Mathematics Subject Classification: 62F12, 93E12.
Note: This report will be submitted for publication elsewhere.

INTRODUCTION

This paper deals with the problem of recursively estimating a d-dimensional parameter that occurs in the intensity process of a counting process. Off-line estimation procedures, such as maximum likelihood estimation, have been analyzed in a number of papers, for instance those written by LIN'KOV [8], KUTOYANTS [5], SAGALOVSKY [11], KONECNY [4] or OGATA [9]. They proved that under certain conditions that differ from paper to paper the maximum likelihood estimator has desirable properties, such as consistency, asymptotic normality and efficiency. For recursive estimators these properties remain to be investigated. We will establish these in a rather specific situation namely that where the intensity process λ has the form $\lambda_t = \theta^T \phi_t$, where ϕ is some other observed process with values in $\mathbb{R}^d$ and $\theta \in \mathbb{R}^d$ the parameter. Two algorithms are presented that generate recursive estimates and their asymptotic behaviour is analyzed. Both the issues of almost sure convergence and the asymptotic distrubution of the estimators are treated. The first one is attacked by means of a stochastic Lyapunov techinque while for the second one we use central limit theorems for martingales. Some examples illustrate the theory.

1. NOTATION AND CONVENTIONS

We assume that all the stochastic processes that will appear in the sequel are defined on some fixed complete probability space $(\Omega, \mathcal{F}, P)$. We also assume that all these processes are adapted to a complete right continuous filtration $\{\mathcal{F}_t\}_{t \geq 0}$ generated by the observations. With respect to this filtration a counting process $n: \Omega \times [0,\infty) \to \mathbb{N}_0$ is a submartingale that enjoys the Doob-Meyer decomposition

$$n = A + m \tag{1.1}$$

where A is a predictable increasing process and m a local martingale [1].
We assume that A is an absolutely continuous process of the form

$$A_t = \int_0^t \phi_s^T \theta \, ds \tag{1.2}$$

where $\theta \in \mathbb{R}_+^d$ and $\phi: \Omega \times [0,\infty) \to \mathbb{R}_+^d$ is some other observed process which is assumed to be predictable. The non random parameter θ is unknown and is to be estimated on the basis of the observations n_t and ϕ_t. We will denote by θ_0 the "true" parameter value. We will often use instead of (1.1),

Report OS-R8513
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

(1.2) the differential notation

$$dn_t = \phi_t^T\theta dt + dm_t, \;\; n_0 = 0 \tag{1.3}$$

2. The Algorithms

2.1 Least Squares Algorithm.
For this algorithm the estimators $\hat{\theta}_t, t \geq 0$ of θ are given by the following two equations.

$$d\hat{\theta}_t = R_t\phi_t\,(dn_t - \phi_t^T\hat{\theta}_t dt), \;\; \hat{\theta}_0 \tag{2.1}$$

$$dR_t = -R_t\phi_t\phi_t^T R_t dt, \;\; R_0 \tag{2.2}$$

Here R_0 is taken to be a symmetric positive definite matrix. Observe that (2.2) guarantees that R_t stays symmetric and positive definite for all t. If we would take $R_0^{-1} = 0$ (which is not positive), then

$$\hat{\theta}_t = [\int_0^t \phi_s\phi_s^T ds]^{-1} \int_0^t \phi_s dn_s \tag{2.3}$$

satifies (2.1), (2.2) with $\hat{\theta}_0 = 0$. It is easily seen that $\hat{\theta}_t$ given by (2.3) minimizes

$$\int_0^t (\phi_s^T\theta)^2 ds - 2\int_0^t \phi_s^T\theta dn_s$$

as a function of θ, which accounts for the name of the algorithm [11]. In [13] it has been proved that $\hat{\theta}_t$ as given above converges with probability one to θ_0. We quote the precise result.

Theorem 2.1: *Let $\{\hat{\theta}_t\}$ be given by (2.1), (2.2) and let*

$$\psi_t = \phi_t^T\phi_t, \;\; \Psi_t = \int_0^t \psi_s ds + tr(R_0^{-1})$$

Assume that the following three conditions are satisfied

(i) $as - \lim_{t\to\infty} \Psi_t = \infty$

(ii) $\int_0^\infty \Psi_t^{-2}\psi_t\phi_t dt < \infty$ a.s.

(iii) $as - \lim_{t\to\infty} \Psi_t^{-1} \int_0^t \phi_s\phi_s^T ds = C > 0$

Then

i) $as - \lim_{t\to\infty} \hat{\theta}_t = \theta_0$

ii) $as - \lim_{t\to\infty} \Psi_t^{-1} \int_0^t (\phi_s^T(\hat{\theta}_s - \theta_0))^2 ds = 0$

The algorithm (2.1), (2.2) is invariant under non-singular linear transformations in the following sense. Let $S \in \mathbb{R}^{dxd}$ be a non-singular matrix. Write $\eta = S\theta$, $\hat{\eta}_t = S\hat{\theta}_t$, $\xi_t = S^{-T}\phi_t$ and $T_t = SR_tS^T$. Then (2.1), (2.2) transform into

$$d\hat{\eta}_t = T_t\xi_t(dn_t - \xi_t^T\hat{\eta}_t dt), \;\; \hat{\eta}_0$$

$$dT_t = -T_t\xi_t\xi_t^T T_t\, dt, \; T_0$$

which is exactly the least squares algorithm that corresponds to $dn_t = \xi_t^T\eta dt + dm_t$, but this is nothing else then (1.3) because $\xi_t^T\eta = \phi_t^T\theta$.

We apply theorem 2.1 to some examples.

EXAMPLE 2.1: Let $\phi:[0,\infty)\to\mathbb{R}^2_+, \phi_t = [1, 1+\sin t]$. Then $\Psi_t = \frac{5}{2}t - 2cost - \sin 2t + tr(R_0^{-1})$. Clearly assumptions 2.1.i and 2.1.ii are satisfied and

$$\lim_{t\to\infty} \Psi_t^{-1} \int_0^t \phi_s\phi_s^T\, ds = \lim_{t\to\infty} \frac{\frac{2}{5}}{t} \int_0^t \begin{bmatrix} 1 & 1+\sin s \\ 1+\sin s & \frac{3}{2}+2\sin s - \frac{1}{2}\cos 2s \end{bmatrix} ds$$

$$= \frac{1}{5}\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

EXAMPLE 2.2: [see 13]: let

$$\phi[0,\infty)\times\Omega\to\mathbb{R}^2_+, \phi_t = [1, 1+(-1)^{n_{t^-}}]^T, \theta = [\theta_1\ \theta_2]^T$$

Then $\Psi_t = (3+2X_t)t$, where

$$X_t = \frac{1}{t}\int_0^t (-1)^{n_s} ds.$$

From [13] we know that $\lim_{t\to\infty} X_t = \dfrac{-\theta_2}{\theta_1+\theta_2}$ a.s. Assumptions 2.1.i,ii are easily seen to be satisfied and

$$\lim_{t\to\infty} \Psi_t^{-1} \int_0^t \phi_s\phi_s^T ds = \frac{1}{3\theta_1+\theta_2}\begin{bmatrix} \theta_1+\theta_2 & \theta_1 \\ \theta_1 & 2\theta_1 \end{bmatrix} > 0,$$

EXAMPLE 2.3: let X be a Markov process which takes its values in $\{0,1\}$. Assume that the holding times in 0 and 1 are exponentially distributed with means μ_0 and μ_1 respectively. Assume that n_t has intensity $\theta_1 X_{t-} + \theta_0(1-X_{t-})$ which is left continuous, thus predictable. So $\phi_t = [X_{t-} 1 - X_{t-}]^T$. Now $\Psi_t = t + tr(R_0^{-1})$. Again assumptions 2.1.,ii are easy to verify and

$$\lim_{t\to\infty} \frac{1}{\Psi t} \int_0^t \phi_s\phi_s^T ds = \lim_{t\to\infty} \frac{1}{t}\int_0^t \begin{bmatrix} X_s & 0 \\ 0 & 1-X_s \end{bmatrix} ds = \frac{1}{\mu_1+\mu_0}\begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_0 \end{bmatrix}$$

2.2. *Approximate Maximum Likelihood Estimation*

Before stating the estimation algorithm, we prefer to formulate a preliminary version of it and provide a heuristic derivation.

The preliminary algorithm is

$$d\hat{\theta}_t = \frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dn_t - \phi_t^T\hat{\theta}_t dt),\ \hat{\theta}_0 \tag{2.4}$$

$$dQ_t = -\frac{Q_t\phi_t\phi_t^T Q_t}{\phi_t^T\hat{\theta}_t} dt,\ Q_0 \tag{2.5}$$

We will give three approaches that suggest, at least heuristically, the form of this preliminary algorithm. The first one is based on a "implicit-function theorem" type argument (2.2.1). The second approach is based on an associated filtering problem (2.2.2) while the last one uses an asymptotic expression of the likelihood functional (2.2.3). Before presenting the three approaches we give the formula of the likelihood functional which is the Radon-Nikodym derive dP_t / dQ_t, where P_t is the measure on the trajectory space of counting processes defined on $[0,t]$ that is induced by (2.3) and Q_t the measure on the same space induced by a standard Poisson process. In order to express the dependence of dP_t/dQ_t on θ we write $L_t(\theta)=dP_t/dQ_t$. Then the following expression holds [1,p. 171]

$$L_t(\theta) = \exp[\int_0^\infty \log\phi_s^T\theta dn_s - \int_0^t(\phi_s^T\theta-1)ds] \tag{2.6}$$

2.2.1 The maximum likelihood estimator $\hat{\theta}_t$ by definition maximizes (2.6).

Equivalently, $\hat{\theta}_t$ minimizes

$$J_t(\theta) = \int_0^t \phi_s^T\theta ds - \int_0^t \log\phi_s^T\theta dn_s \tag{2.7}$$

If differentation with respect to θ under the integral sign is allowed we look for zero's of

$$l_t(\theta) = \nabla_\theta J_t(\theta) = \int_0^t \phi_s ds - \int_0^t \frac{\phi_s}{\phi_s^T\theta} dn_s \tag{2.8}$$

If $J_t(\theta) = J(t,\theta)$ happens to be a smooth function of both θ and t, it follows from the implicit function theorem that $\hat{\theta}_t$ satisfies the equation

$$\frac{d}{dt}\hat{\theta}_t = -[\nabla_\theta l_t(\hat{\theta}_t)]^{-1}\frac{\partial}{\partial t}l_t(\hat{\theta}_t)$$

A similar expression in the present situation where $l_t(\theta)$ is not smooth, but has jumps, is

$$d\hat{\theta}_t = -[\nabla_\theta l_t(\hat{\theta}_{t-})]^{-1}\partial_t l_t(\hat{\theta}_{t-}) \tag{2.9}$$

where ∂_t is the forward partial differential operator with respect to t. Since we have

$$\partial_t l_t(\hat{\theta}_{t-}) = \phi_t dt - \frac{\phi_t}{\phi_t^T\hat{\theta}_{t-}}dn_t$$

and

$$\nabla_\theta l_t(\theta) = \int_0^t \frac{\phi_s\phi_s^T}{(\phi_s^T\theta)^2}dn_s$$

equation (2.9) becomes after writing $Q_t=[\nabla_\theta l_t(\hat{\theta}_t)]^{-1}$

$$d\hat{\theta}_t = \frac{Q_{t-}\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dn_t - \phi_t^T\hat{\theta}_t dt) \tag{2.10}$$

The next problem is to find an evolution equation for Q. Recall that one of the objectives is that the algorithm gives us strongly consistent estimators. Therefore we should have for large t $\hat{\theta}_t \approx \theta_0$. Hence for large t

$$Q_t^{-1} \approx \int_0^t \frac{\phi_s\phi_s^T}{(\phi_s^T\theta_0)^2}dn_s = \int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds + \int_0^t \frac{\phi_s\phi_s^T}{(\phi_s^T\theta_0)^2}dm_s \tag{2.11}$$

The last term of the right hand side of (2.11) is a zero mean martingale. We get a new approximation of Q_t^{-1} by deleting it:

$$Q_t^{-1} \approx \int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds$$

Finally we replace θ_0 by $\hat{\theta}_s$ and we arrive at

$$dQ_t = -\frac{Q_t\phi_t\phi_t^T Q_t}{\phi_t^T\hat{\theta}_t}dt \tag{2.5}$$

Observe from (2.5) that Q_t is continiuous. Consequently (2.10) is indeed (2.4).

2.2.2. Another way of arriving at (2.4), (2.5) is the following. Consider the following filtering problem. We have an observation equation

$$dn_t = \phi_t^T\theta dt + dm_t, n_0 = 0$$

Here ϕ is a $\mathcal{F}_t^n$-predictable random process where $\mathcal{F}_t^n = \sigma\{n_s, 0 \leq s \leq t\}$ and θ is an unobserved random parameter, that is $\sigma(\theta) \not\subset \mathcal{F}_t^n$ for all t. It is known [1] that the optimal (in mean squared error sense) estimator of θ given the observations $\mathcal{F}_t^n$ is $\hat{\theta}_t := E[\theta|\mathcal{F}_t^n]$, and that it satisfies the following equation

$$d\hat{\theta}_t = \frac{P_{t-}\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dn_t - \phi_t^T\hat{\theta}_t dt),\ \hat{\theta}_0 = E\theta.$$

Here P_t is the conditional covariance matrix $E[(\theta-\hat{\theta}_t)(\theta-\hat{\theta}_t)^T|\mathcal{F}_t^n]$ and satisfies

$$dP_t = -\frac{P_t\phi_t\phi_t^T P_t}{\phi_t^T\hat{\theta}_t}dt + [E[(\theta-\hat{\theta}_t)(\theta-\hat{\theta}_t)^T(\theta-\hat{\theta}_t)^T\phi_t|\mathcal{F}_t^n]$$
$$-\frac{P_t\phi_t\phi_t^T P_t}{\phi_t^T\hat{\theta}_t}]_{t=t-}\frac{1}{\phi_t^T\hat{\theta}_{t-}}(dn_t - \phi_t^T\hat{\theta}_t dt)$$

In this setting the innovations process $n_t - \int_0^t \phi_s^T\hat{\theta}_s ds$ is a martingale with zero mean. We can approximate this equation by setting the martingale term zero. Denoting the approximation of P_t by Q_t we find as a truncated second order filter

$$d\hat{\theta}_t = \frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dn_t - \phi_t^T\hat{\theta}_t dt) \tag{2.12}$$

$$dQ_t = -Q_t\frac{\phi_t\phi_t^T}{\phi_t^T\hat{\theta}_t}Q_t dt \tag{2.13}$$

It can be argued that the effect of the prior distribution of θ decays with time. Hence we will eventually get estimators $\hat{\theta}_t$ of θ that are hardly depending on the prior distribution. Consequently the $\hat{\theta}_t$'s for large t will not change much if we would take θ as a deterministic parameter. This suggests the use of the same formulas (2.12), (2.13) for our original estimation problem.

2.2.3. A third way to obtain the recursive scheme (2.4), (2.5) is to make use of an asymptotic expression of the logarithm of the likelihood functional. This expression expression, that reveals the so called local asymptotic normality property (LAN), is a key result in proving consistency and asymptotic normality of the (off-line) maximum likelihood estimator [5,6,8]. Of course the same properties are desired for our recursive estimator. We will exploit the LAN property for the case where ϕ is a deterministic function. Similar considerations can be found in [2]. Denote by P_θ the probability measure on the space of counting process trajectories induced by a counting process with intensity $\phi_t^T\theta$. Define

$$\bar{Q}_t = [\int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds]^{-1}$$

Let $u \in \mathbb{R}^d$ and denote by $Z_t(u)$ the restriction of $dP_{\theta_0+\bar{Q}_t^{1/2}u}/dP_{\theta_0}$ to the space of trajectories defined on $[0,t]$. Under certain conditions (the precise form of those are not important at this point) we have the following result [5, 6]

$$\log Z_t(u) = u^T \overline{Q}_t^{-\frac{1}{2}} \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dm_s - \frac{1}{2} u^T u + p_t \tag{2.14}$$

where p is a stochastic process that converges to zero in probability for $t \to \infty$ and $\overline{Q}_t^{-\frac{1}{2}} \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dm_s$ converges in law to a gaussian $(0,I)$ random variable.
If we set $p_t = 0$, then the value of u that maximizes (2.14) is

$$\hat{u}_t = \overline{Q}_t^{-\frac{1}{2}} \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dm_s$$

Hence an approximate maximum likelihood estimator of θ_0 is

$$\overline{\theta}_t = \theta_0 + \overline{Q}_t \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dn_s$$

Or

$$\overline{\theta}_t = \overline{Q}_t \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dn_s \tag{2.15}$$

Observe that $\overline{Q}_t^{-\frac{1}{2}} (\overline{\theta}_t - \theta_0)$ converges in law to a gaussian (O,I) random variable.
Of course $\overline{\theta}_t$ is useless as an estimator of θ_0, since it depends on θ_0. We just use it at an intermediate step in obtaining our algorithm (2.4), (2.5). A simple calculation shows that $\overline{\theta}_t$ and $\overline{Q}_t$ satisfy

$$d\overline{\theta}_t = \frac{\overline{Q}_t \phi_t}{\phi_t^T \theta_0} (dn_t - \phi_t^T \overline{\theta}_t dt) \tag{2.16}$$

$$d\overline{Q}_t = -\frac{\overline{Q}_t \phi_t \phi_t^T \overline{Q}_t}{\phi_t^T \theta_0} dt \tag{2.17}$$

As before since one is looking for $\hat{\theta}_t$'s that are close to θ_0 (and thus close to $\overline{\theta}_t$) we replace θ_0 and $\overline{\theta}_t$ in (2.16), (2.17) by $\hat{\theta}_t$ and write Q_t instead of $\overline{Q}_t$, thus arriving again at (2.4), (2.5).

Having finished the explanation of the form of the preliminary version of our algorithm, we will now present it in its final form. The change that has been made is just for technical convenience and makes the proof work. The reasons for the change will be apparent from the proof of theorem 3.2. We give a little discussion that tells us that this change is not dramatic. Suppose that $\hat{\theta}_t$ given by (2.4), (2.5) converges almost surely to θ_0. Then eventually $\hat{\theta}_t$ will be in any neighbourhood of θ_0. Hence if $\epsilon \in \mathbb{R}_+^d$ is such that all its components are smaller than the corresponding components of θ_0 we have $\phi_t^T \hat{\theta}_t > \phi_t^T \epsilon$ eventually. This is exactly the property that we need in the analysis. However (2.4), (2.5) do not guarantee us, that this inequality holds. Obviously the modification (2.18)-(2.21) below has the desired property. Define the indicator process I_t as follows

$$I_t = I_{\{\phi_t^T x_t > \phi_t^T \epsilon\}} \tag{2.18}$$

where $\epsilon \in \mathbb{R}_+^d$ is such that $0 < \epsilon_i < \theta_{0i}, i = 1, \ldots, d$. We are now in the position to state our

APPROXIMATE MAXIMUM LIKELIHOOD (AML) ALGORITHM

$$dx_t = \frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dn_t - \phi_t^T x_t dt), \; x_0 \tag{2.19}$$

$$dQ_t = -\frac{Q_t\phi_t\phi_t^T Q_t}{\phi_t^T\hat{\theta}_t}dt, \; Q_0 \tag{2.20}$$

$$\hat{\theta}_t = x_t I_t + \epsilon(1-I_t) \tag{2.21}$$

Here x_0 is taken such that $\hat{\theta}_0 = x_0$, and Q_0 is a symmetric positive definite matrix

Apparently one should be able to establish lower bounds for the components θ_0 in order to compute $\hat{\theta}_t$ according to (2.18)-(2.21). In practical situations there are often physical considerations that enable us to do so. As for the least squares algorithm we can also prove invariance of (2.18) - (2.21) under non singular linear transformations. Contrary to (2.4), (2.5) we even have invariance of (2.18) - (2.21) under time transformations. Let $\tau = f(t)$ be a (possibly random) time transformation with inverse $t = g(\tau)$. Assume that g has a derivative g' almost everywhere and $g' \geqslant 0$. Write $\tilde{y}_t = y_{g(\tau)}$ for the time transformed process y. Then we have

$$d\tilde{n}_\tau = \tilde{\phi}_t^T \theta_0 g'(\tau)d\tau + d\tilde{m}_\tau \tag{2.22}$$

The algorithm corresponding to (2.22) is

$$d\tilde{x}_t = \frac{\tilde{Q}_\tau\tilde{\phi}_{\tau-}}{\tilde{\phi}_{\tau-}\tilde{\theta}_{\tau-}}(d\tilde{n}_\tau - \tilde{\phi}_\tau^T\tilde{x}_\tau \; g'(\tau)d\tau) \tag{2.23}$$

$$d\tilde{Q}_\tau = -\frac{\tilde{Q}_\tau\tilde{\phi}_\tau\tilde{\phi}_\tau^T\tilde{Q}_\tau}{\tilde{\phi}_\tau^T\tilde{\theta}_\tau}g'(\tau)\, d\tau \tag{2.24}$$

$$\tilde{\theta}_\tau = \tilde{x}_\tau\tilde{I}_\tau + \epsilon(1-\tilde{I}_\tau) \tag{2.25}$$

which is indeed the same as the time transformed version of (2.18) - (2.21)

3. CONSISTENCY OF THE AML ALGORITHM

In the proof of theorem 3.2 below, where strong consistency of $\hat{\theta}_t$ coming from (2.18) - (2.21) is proved we use the following lemma, which is a simplified version of a more general result in [12].

LEMMA 3.1: *Let x,a,b be nonnegative stochastic processes: a and b increasing with $a_0 = 0$ and m a local martingale. Assume that the following relation holds for all t :$x_t = a_t - b_t + m_t$.*
Assume
i) $\exists c \in \mathbb{R}_+ : \Delta a_t \leqslant c, \forall t \geqslant 0$.
ii) *as* $- \lim_{t\to\infty} a_t < \infty$
Then
i) *as* $- \lim_{t\to\infty} x_t$ *exists* and *is finite*
ii) *as* $- \lim_{t\to\infty} b_t < \infty$.

The principal result of this section is

THEOREM 3.2: *Let $\theta_0 \in \mathbb{R}_+^d$ and let $\epsilon \in \mathbb{R}_+^d$ be such that $\theta_0 - \epsilon \in \mathbb{R}_+^d$. Let $\Phi_t = \int_0^t \phi_s ds$ and assume*

i) $\Phi_t^T\theta_0 \to \infty$ *a.s* $(t \to \infty)$
ii) $\liminf_{t\to\infty} \frac{1}{\Phi^T\theta_0}\int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds = C > 0.$

Then

i) $a.s.\lim_{t\to\infty}\hat{\theta}_t = \theta_0$

ii) $a.s.\lim_{t\to\infty}\frac{1}{\theta_0^T\Phi_t}\int_0^t\frac{(\phi_s^T(\hat{\theta}_s-\theta_0))^2}{\phi_s^T\theta_0}ds = 0$

Before proving the theorem we notice that conditions 3.2.i, 3.2ii are equivalent with $(\mathbf{1}=(1,\ldots,1)^T)$

i') $\Phi_t^T\mathbf{1}\to\infty a.s.$

ii') $\liminf\frac{1}{\Phi_t^T\mathbf{1}}\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\mathbf{1}}ds>0$

The equivalence of i) and i') can easily be seen by noting that $\underline{\theta}\Phi_t^T\mathbf{1}\leq\theta_0^T\Phi_t\leq\overline{\theta}\Phi_t^T\mathbf{1}$, where $\underline{\theta}=\min\{\theta_{0i}, i=1,\ldots,d\}, \overline{\theta}=\max\{\theta_{0i}, i=1,\ldots,d\}$. The equivalence of ii) and ii') follows similary.

PROOF OF THEOREM 3.2: i) Let $\tilde{x}_t=x_t-\theta_0$. Then

$$d\tilde{x}_t=\frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dm_t - \phi_t^T\tilde{x}_t dt),$$

$$d(\tilde{x}_t^T Q_t^{-1}\tilde{x}_t)=\frac{2\phi_t^T\tilde{x}_{t-}}{\phi_t^T\hat{\theta}_{t-}}dm_t-\frac{(\tilde{x}_t^T\phi_t)^2}{\phi_t^T\hat{\theta}_t}dt+\frac{\phi_t^T Q_t\phi_t}{(\phi_t^T\hat{\theta}_{t-})^2}dn_t \tag{3.1}$$

$$=dm_{1t}-\frac{(\tilde{x}_t^T\phi_t)^2}{(\phi_t^T\hat{\theta}_t)^2}dt+\frac{\phi_t^T Q_t\phi_t}{(\phi_t^T\hat{\theta}_t)^2}\phi_t^T\theta_0 dt \tag{3.2}$$

where we have summarized the martingale term of (3.1) as dm_{1t}. Define

$$r_t=tr(Q_0^{-1})+\int_0^t\frac{\phi_s^T\phi_s}{\phi_s^T\epsilon}ds,$$

Then

$$r_t\geq tr(Q_0^{-1})+\int_0^t\frac{\phi_s^T\phi_s}{\phi_s^T\hat{\theta}_s}ds=tr(Q_t^{-1}).$$

Define

$$u_t=r_t^{-1}[\tilde{x}_t^T Q_t^{-1}\tilde{x}_t + \int_0^t\frac{(\phi_s^T\tilde{x}_s)^2}{\phi_s^T\hat{\theta}_s}ds],$$

then

$$du_t=-r_t^{-1}\frac{\phi_t^T\phi_t}{\phi_t^T\epsilon}u_t dt+r_t^{-1}dm_{1t}+r_t^{-1}\frac{\phi_t^T Q_t\phi_t}{(\phi_t^T\hat{\theta}_t)^2}\phi_t^T\theta_0 dt \tag{3.3}$$

We are able to apply lemma 3.1 as soon as we have verified assumption ii) which leads us to the calculation of

$$0\leq\int_0^\infty r_t^{-1}\frac{\phi_t^T Q_t\phi_t}{(\phi_t^T\hat{\theta}_t)^2}\phi_t^T\theta_0 dt\leq tr\int_0^\infty r_t^{-1}tr(Q_t^{-1})\frac{Q_t\phi_t\phi_t^T Q_t}{\phi_t^T\hat{\theta}_t}\frac{\phi_t^T\theta_0}{\phi_t^T\hat{\theta}_t}dt$$

$$\leq tr\int_0^\infty\frac{Q_t\phi_t\phi_t^T Q_t}{\phi_t^T\hat{\theta}_t}\frac{\phi_t^T\theta_0}{\phi_t^T\hat{\theta}_t}dt\leq tr\int_0^\infty(-dQ_t)\frac{\overline{\theta}\phi_t^T\mathbf{1}}{\underline{\epsilon}\phi_t^T\mathbf{1}}dt$$

$$=\frac{\overline{\theta}}{\underline{\epsilon}}tr\int_0^\infty(-dQ_t)\leq\frac{\overline{\theta}}{\underline{\epsilon}}tr(Q_0)<\infty.$$

Having verified assumption 3.1ii we conclude that as-lim u_t exists and is finite a.s. We also get from the same lemma and eq (3.3) that

$$\text{a.s.}\lim_{t\to\infty}\int_0^\infty r_t^{-1}\frac{\phi_t^T\phi_t}{\phi_t^T\epsilon}u_t dt<\infty. \tag{3.4}$$

Now

$$r_t\geq\frac{1}{\overline{\epsilon}}\int_0^t\frac{\phi_s^T\phi_s}{\phi_s^T\mathbf{1}}ds+tr(Q_0^{-1})\geq\frac{1}{d\overline{\epsilon}}\int_0^t\phi_s^T\mathbf{1}ds+tr(Q_0^{-1}).$$

where we used in the last inequality that $\phi_t^T\phi_t\geq\frac{1}{d}(\phi_t^T\mathbf{1})^2$. Hence from assumption 3.2i. $r_t\to\infty$ a.s. Suppose now that on a set $\Omega_1\subset\Omega$ of positive probability we have lim $u_t\geq\delta$ for some $\delta>0$. Then there is τ such that $t\geq\tau$ implies $u_t\geq\frac{1}{2}\delta$. But then

$$\int_\tau^\infty r_t^{-1}\frac{\phi_t^T\phi}{\phi_t^T\epsilon}u_t dt\geq\frac{1}{2}\delta\int_\tau^\infty d\log r_t=\infty$$

which contradicts (3.4). Hence $\text{as-}\lim_{t\to\infty}u_t=0$. Since u is the sum of two positive processes we have in particular

$$\text{as}-\lim_{t\to\infty}\tilde{x}_t^T\frac{Q_t^{-1}}{r_t}\tilde{x}_t=0. \tag{3.5}$$

Define now $\overline{\theta}_t=\sup\{\hat{\theta}_{is},s\in[0,t],i=1,\ldots,d\}$ and write $\lambda_{\min}(A)=\min\sigma(A)$ for the minimal eigenvalue of a matrix A. Then

$$Q_t^{-1}-Q_0^{-1}=\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\hat{\theta}_s}ds\geq\frac{1}{\overline{\theta}_t}\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\mathbf{1}}ds.$$

Hence

$$0\leq\frac{\tilde{x}_t^T\tilde{x}_t}{\overline{\theta}_t}\lambda_{\min}\left[\frac{1}{\Phi_t^T\mathbf{1}+tr(Q_0^{-1})}\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\mathbf{1}}ds\right]\leq$$

$$\leq\tilde{x}_t^T\tilde{x}_t\lambda_{\min}\left[\frac{1}{\Phi_t^T\mathbf{1}+tr(Q_0^{-1})}\left[Q_t^{-1}-Q_0^{-1}\right]\right]$$

$$\leq\frac{1}{\Phi_t^T\mathbf{1}+tr(Q_0^{-1})}\tilde{x}_t^T(Q_t^{-1}-Q_0^{-1})\tilde{x}_t\leq\frac{1}{\underline{\epsilon}}r_t^{-1}\tilde{x}_t^T(Q_t^{-1}-Q_0^{-1})\tilde{x}_t$$

which tends to zero by (3.5). Consequently from assumption 3.2.ii

$$\frac{\tilde{x}_t^T\tilde{x}_t}{\overline{\theta}_t}\to 0\text{ a.s.} \tag{3.6}$$

Now it is easy to prove that $\overline{\theta}_t$ is bounded. For suppose not, then there is $\hat{\theta}_{it}$ such that limsup $\hat{\theta}_{it}=\infty$. But then also limsup $\tilde{x}_{it}=\infty$ and we get immediately from (3.6) that this cannot happen. Hence $\overline{\theta}_t\leq K$ for some $K>0$. But then from

$$\tilde{x}_t^T\tilde{x}_t\leq K\frac{\tilde{x}_t^T\tilde{x}_t}{\overline{\theta}_t}.$$

we see that $\tilde{x}_t \to 0$ and so eventually

$$\phi_t^T x_t = \phi_t^T \tilde{x}_t + \phi_t^T \theta_0 > \phi_t^T \epsilon.$$

Then $I_t \to 1$ and consequently

$$\hat{\theta}_t = (\tilde{x}_t + \theta_0) I_t + \epsilon(1 - I_t) \to \theta_0$$

ii) $\tilde{\theta}_t = \tilde{x}_t I_t + (1 - I_t)(\epsilon - \theta_0)$. Let τ be such that $t \geqslant \tau$ implies $I_t = 1$ then for $t \geqslant \tau$ $\tilde{\theta}_t = \tilde{x}_t$.
Hence

$$\int_0^t \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_s} ds = \int_0^\tau \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_s} ds + \int_\tau^t \frac{(\phi_s^T \tilde{x}_s)^2}{\phi_s^T \hat{\theta}_s} ds$$

From the fact that $u_t \to 0$ we have

$$\frac{1}{r_t} \int_\tau^t \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_s} ds \to 0.$$

But then it is easy to deduce from the fact that $\hat{\theta}_t \to \theta_0$ a.s. that we also have

$$\frac{1}{r_t} \int_\tau^t \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \theta_0} ds \to 0 \text{ a.s.}$$

and

$$\frac{1}{\theta_0^T \Phi_s} \int_0^t \frac{(\phi_s^T \tilde{\theta}_s)^2)}{\phi_s^T \theta_0} ds \to 0 \text{ a.s. Q.E.D.}$$

Before giving a few examples to which this theorem can be applied let us remark that a necessary condition for assumption 3.2.ii' is

$$\liminf_{t \to \infty} \frac{1}{\Phi^T \mathbf{1}} \int_0^t \phi_{is} ds > 0.$$

Clearly this condition is not sufficient. $\phi_t \equiv \mathbf{1}$ is a counterexample.

EXAMPLE 3.1. Let $\phi : [0, \infty) \to \mathbb{R}_+^2, \phi_t = [1, 1 + \sin t]^T$. The following result will be used. For $a > b \geqslant 0$.

$$\int_0^{2\pi} \frac{1}{a + b\sin x} dx = \frac{2\pi}{\sqrt{a^2 - b^2}},$$

hence

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \frac{1}{a + b\sin x} dx = \frac{1}{\sqrt{a^2 - b^2}},$$

Then

$$\lim_{t \to \infty} \frac{1}{\Phi_t^T \mathbf{1}} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds = \lim_{t \to \infty} \frac{1}{2t - \cos t + 1} \int_0^t \begin{bmatrix} 1 & 1 + \sin x \\ 1 + \sin x & 1 + 2\sin x + \sin^2 x \end{bmatrix} \frac{dx}{2 + \sin x}$$

$$= \frac{1}{2} \begin{bmatrix} \frac{1}{3}\sqrt{3} & 1 - \frac{1}{3}\sqrt{3} \\ 1 - \frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} \end{bmatrix}, \text{which is positive definite.}$$

EXAMPLE 3.2. Let $\phi:[0,\infty)\times\Omega\rightarrow\mathbb{R}^2_+, \phi_t = [1,1+(-1)^{n_{t-}}]^T, \theta = [\theta_1\ \theta_2]^T$
Introduce

$$X_t = \frac{1}{t}\int_0^t (-1)^{n_s} ds$$

Then

$$\frac{1}{\Phi_t^T\mathbb{1}}\int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\mathbb{1}}ds = \frac{1}{t(2+X_t)}\int_0^t \begin{bmatrix} 1 & 1+(-1)^{n_s} \\ 1+(-1)^{n_s} & 2+2(-1)^{n_s}\end{bmatrix}\frac{ds}{2+(-1)^{n_s}}$$

$$= \frac{\frac{1}{3}}{t(2+X_t)}\int_0^t \begin{bmatrix} 2-(1)^{n_s} & 1+(-1)^{n_s} \\ 1+(-1)^{n_s} & 2+2(-1)^{n_s}\end{bmatrix}ds$$

$$= \frac{\frac{1}{3}}{2+X_t}\begin{bmatrix} 2-X_t & 1+X_t \\ 1+X_t & 2+2X_t\end{bmatrix}$$

After some calculations [13] we find that

$$as-\lim_{t\rightarrow\infty} X_t = -\frac{\theta_2}{\theta_1+\theta_2}.$$

So

$$\lim_{t\rightarrow\infty}\frac{1}{\Phi_t^T\mathbb{1}}\int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\mathbb{1}}ds = \frac{\frac{1}{3}}{2\theta_1+\theta_2}\begin{bmatrix} 2\theta_1+3\theta_2 & \theta_1 \\ \theta_1 & 2\theta_1\end{bmatrix} >0.$$

EXAMPLE 3.3.: Let X be a Markov process that takes its values in $\{0,1\}$. Assume that the holding times in 0 and 1 are exponentially distributed with means μ_0 and μ_1 respectively. Assume that n_t has the intensity $\theta_1 X_{t-}+\theta_0(1-X_{t-})$, which corresponds to $\phi_t=[X_{t\_},1-X_{t\_}]^T$ and $\theta=(\theta_1\ \theta_0)^T$. Then

$$\lim_{t\rightarrow\infty}\frac{1}{\Phi_t^T\mathbb{1}}\int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\mathbb{1}}ds = \lim_{t\rightarrow\infty}\frac{1}{t}\int_0^t \begin{bmatrix} X_s & 0 \\ 0 & 1-X_s\end{bmatrix}ds = \frac{1}{\mu_1+\mu_0}\begin{bmatrix}\mu_1 & 0 \\ 0 & \mu_0\end{bmatrix}.$$

4. SOME CENTRAL LIMIT THEOREMS

This section provides the background for analyzing the asymptotic distribution of the $\hat{\theta}_t$' s generated by the least squares or asymptotic maximum likelihood algorithm. That analysis will be carried out in section 5. The tools we will use in section 5 are certain central limit theorems for martingales. Two types of results are available in this direction. On the one hand we have central limit theorems for sequences of martingales or stochastic integrals that are obtained for instance by REBOLLEDO [10]. On the other hand there are results for the asymptotic distribution of a stochastic integral with respect to a local martingale measure as t goes to infinity that can be found in e.g. LIN'KOV [7]. It will be shown in the sequel that the latter type is contained in the first.

4.1. We start by presenting Rebolledo's results and follow his notation here in. Since we are only interested in martingales associated with point processes, we do not quote his results in full generality but only for quasi left continuous locally square integrable martingales.

Suppose we have on $(\Omega,\mathcal{F},P)$ a sequence of right continuous complete filtrations $\mathbb{F}^n=\{\mathcal{F}_t^n\}$. Let $\mathfrak{M}^{2,loc}(\mathbb{F}^n)$ be the set of locally square integrable $\mathbb{F}^n$-adapted martingales starting in zero. We will consider sequences $\{M^n\}$ where for each n $M^n\in\mathfrak{M}^{2,loc}(\mathbb{F}^n)$ and is assumed to be quasi left continuous. Define $\sigma^\epsilon[M^n]:\Omega\times[0,\infty)\to\mathbb{R}$ by

$$\sigma^\epsilon[M^n]_t = \sum_{s\leq t}|\Delta M_s^n|^2 I_{\{|\Delta M_s^n|>\epsilon\}}.$$

Denote by $\tilde{\sigma}_\epsilon[M^n]$ its dual predictable projection.
We say that $\{M^n\}$ satisfies the strong asymptotic rarefaction of jumps property of the second kind ($SARJ2$) if

$$\tilde{\sigma}^\epsilon[M^n]_t \overset{P}{\to} 0, \ as\ n\to\infty, \forall t\geq 0 \tag{4.1}$$

The sequence $\{M^n\}$ is said to satisfy the Lindeberg condition if

$$E\tilde{\sigma}^\epsilon[M^n]_t\to 0, \quad as\ n\to\infty, \forall t\geq 0 \tag{4.2}$$

In connection with sections 4.2, 4.3 we introduce the following definition. We say that the sequence $\{M^n\}$ satisfies the Lyapunov condition if there exists a $\delta>0$ such that for all n $M^n\in\mathcal{M}^{2+\delta,loc}(\mathbb{F}^n)$ and if

$$E\sum_{s\leq t}|\Delta M_s^n|^{2+\delta}\to 0, \tag{4.3}$$

In [10] it has been shown that the Lindeberg condition implies *SARJ 2*. Also the following implication holds

PROPOSITION 4.1: *If the sequence $\{M^n\}$ satifies Lyapunov's condition then it also satisfies Lindeberg's condition.*

PROOF: A corresponding result is known in Central limit theory for sequences of random variables. The proof that we will give is similar to the proof in the random variables case.
Let us introduce the jump times τ_k^n defined by

$$\tau_0^n = 0, \tau_k^n = inf\{t>\tau_{k-1}^n:\Delta M_t \neq 0\}.$$

Then

$$E\sum_{s\leq t}|\Delta M_s^n|^2 I_{\{\Delta M_s^n|>\epsilon\}} = E\sum_{k=1}^{\infty}|\Delta M_{\tau_k^n}^n|^2 I_{\{|\Delta M_{\tau_k^n}^n|>\epsilon\}\cap\{\tau_k^n\leq t\}}$$

which is by application of Holder's inequality ($p=1+\frac{\delta}{2}, q=1+\frac{2}{\delta}$) less than

$$\sum_{k=1}^{\infty}[E|\Delta M_{\tau_k^n}^n|^{2+\delta}I_{\{\tau_k^n\leq t\}}]^{\frac{2}{2+\delta}}[EI_{\{|\Delta M_{\tau_k^n}^n|>\epsilon\}}I_{\{\tau_k^n\leq t\}}]^{\frac{\delta}{2+\delta}}$$

which is by application of Chebychev's inequality (see below) less than

$$\sum_{k=1}^{\infty}[E|\Delta M_{\tau_k^n}|^{2+\delta}I_{\{\tau_k^n\leq t\}}]^{\frac{2}{2+\delta}}[E|\Delta M_{\tau_k^n}^n|^{2+\delta}I_{\{\tau_k^n\leq t\}}]^{\frac{\delta}{2+\delta}}\epsilon^{-\delta}$$

$$= \epsilon^{-\delta}\sum_{k=1}^{\infty}E|\Delta M_{\tau_k^n}^n|^{2+\delta}I_{\{\tau_k^n\leq t\}} = \epsilon^{-\delta}E\sum_{s\leq t}|\Delta M_s^n|^{2+\delta}\to 0 \quad \square$$

In the proof we used the following form of Chebychev's inequality.

$$EI_{\{|x|>\epsilon\}\cap A}\leq \frac{E|x|^p I_A}{\epsilon^p}, \forall A\in\mathcal{F}, \forall p.$$

Notice that an even stronger condition than Lyapunov's is

$$\sup\{|\Delta M_t^n|:t\geqslant 0\}\leqslant c_n$$

where $c_n\in\mathbb{R}$ and $c_n\downarrow 0$.
Now we formulate Rebolledo's main result.

THEOREM 4.2: *Let M be a continuous martingale and assume that $A=<M>$ is a determinstic function. Let $M^n\in\mathfrak{M}^{2,loc}(\mathbb{F}^n),\forall_n$ such that*
i) $\{M^n\}$ satisfies *SARJ 2*
ii) $[M^n]_t \xrightarrow{P} A_t$ *as* $n\to\infty,\forall t\geqslant 0$.
Then M^n weakly converges to M. Notation $M_n \xrightarrow{\mathfrak{L}} M$.

4.2. In this section we summarize LIN'KOV'S result [7]. As usual we have a complete right continuous filtration $\{\mathfrak{F}_t\}$ on $(\Omega,\mathfrak{F},P)$ Let μ be a local martingale measure on $[0,\infty)\times Z$, where $(Z,\mathfrak{Z})$ is a Blackwell space. $\mathfrak{Z}$ is a σ-algebra on Z. Assume that μ is such that $\mu(\{t\},E)\in\{0,1\}$ for all $E\in\mathfrak{Z}$ and suppose that the characteristic $<\mu>=\nu$ of μ is continuous. Notice that $\mu+\nu$ is on an integer valued random measure. Denote by $L_2^\nu[0,T]$ the set of $\mathfrak{F}_t$ adapted proceses $f=f(\omega,t,x)$ which are measureable as a function of (ω,t,x) such that there exists a sequence of simple functions $\{f_n\}$ such that.

$$P-\lim_{n\to\infty}\int_0^T\int_Z|f(t,x)-f_n(t,x)|^2\nu(dt,dx) = 0.$$

For such f the following stochastic integral is well-defined:

$$\zeta_t = \int_0^t\int_Z f(s,x)\mu(ds,dx),\quad t\in[0,T]$$

In [7] one can find the following

PROPOSITION 4.3.: *Assume that $f\in L_2^\nu[0,T],\forall T\geqslant 0$ and $\exists$ function $g:[0,\infty)\to[0,\infty)$ such that $g(t)\to\infty$, as $t\to\infty$ with*
i) $P-\lim\limits_{T\to\infty} g(T)^{-2}\int_0^T\int_Z f^2(t,x)\ \nu\ (dt,dx)=1$
ii) $\lim\limits_{T\to\infty} g(T)^{-(2+\delta)}E\int_0^T\int_Z|f(t,x)|^{2+\delta}\nu(dt,dx)$ for some $\delta>0$.

Then ζ_T is asymptotically normal with parameters $(0,g(T)^2)$ for $T\to\infty$.

4.3. In this section we will show how Lin'kov's result can be deduced from Rebolledo's (theorem 4.2.) which provides thus an alternative proof for the one that can be found in [7].

PROOF OF PROPOSITION 4.3.
We have to show that for all sequences b_n with $b_n\to\infty\zeta_{b_n}$ has a distribution which is asymptotically normal with parameter $(0,g(b_n)^2)$. Define $a_n=g(b_n)^2$. Without loss of generality we can assume that g is strictly increasing. Hence its inverse h is well defined. Let $t\in[0,1]$ and define

$$M_t^n = \frac{1}{\sqrt{a_n}}\int_0^{h(\sqrt{a_n}t)}\int_Z f(s,x)\mu(ds,dx). \tag{4.4}$$

Let $\mathfrak{F}_t^n=\mathfrak{F}_{h(\sqrt{a_n}t)}$, then M^n is $\mathbb{F}^n$-adapted. We will now show that M^n as defined in (4.4) satisfies the Lyapunov condition (4.3).

$$\Delta M_t^n = \frac{1}{\sqrt{a_n}}\int_Z f(h(\sqrt{a_n}t),x)\mu(\{h(\sqrt{a_n}t)\},dx)$$

Because $\mu(\{h(\sqrt{a_n t})\})$ is in fact a dirac measure for each ω on Z, concentrated on some point $z=z(a_n t,\omega)$ [3], we have

$$|\Delta M_t^n|^{2+\delta} = a_n^{-1-\delta/2} \int_Z |f(h(\sqrt{a_n t}),x)|^{2+\delta}\mu(\{h(\sqrt{a_n t})\},dx)$$

Hence

$$\sum_{s\leq t}|\Delta M_s^n|^{2+\delta} = a_n^{-1-\delta/2} \int_0^{h(\sqrt{a_n t})} \int_Z |f(s,x)|^{2+\delta}(\mu+\nu)(ds,dx)$$

Since μ is a local martingale measure

$$E\sum_{s\leq t}|\Delta M_s^n|^{2+\delta}=a_n^{-1-\delta/2}E \int_0^{h(\sqrt{a_n t})} \int_Z |f(s,x)|^{2+\delta}\nu(ds,dx) =$$

$$\leq a_n^{-1-\delta/2}\, E \int_0^{h(\sqrt{a_n})} \int_Z |f(s,x)|^{2+\delta}\nu(ds,dx) =$$

$$=(g(h(\sqrt{a_n})))^{-(2+\delta)}\, E \int_0^{h(\sqrt{a_n})} \int_Z |f(s,x)|^{2+\delta}\nu(ds,dx) \to 0$$

by assumption 4.3.ii. So a forteriori the sequence $\{M^n\}$ satisfies the *SARJ 2* condition by proposition 4.1 and the remark preceeding it.
We proceed to investigate the process $<M^n>$. A simple calculation gives

$$<M^n>_t = \frac{1}{a_n} \int_0^{h(\sqrt{a_n t})} \int_Z |f(s,x)|^2\nu(ds,dx) =$$

$$= t\, \frac{1}{(g(h(\sqrt{a_n t})))^2} \int_0^{h(\sqrt{a_n t})} \int_Z |f(s,x)|^2\nu(ds,dx) \to t$$

in probability by assumption 4.3.i.

We are now in the position to apply theorem 4.2 and we conclude that $<M^n> \xrightarrow{\mathcal{L}} W$, where W is a standard brownian motion. In particular

$$M_1^n \xrightarrow{\mathcal{L}} N(0,1)$$

or

$$\frac{1}{\sqrt{a_n}} \int_0^{h(\sqrt{a_n})} \int_Z f(s,x)\mu(ds,dx) \xrightarrow{\mathcal{L}} N(0,1)$$

which gives us the desired result by definition of a_n.

Along the same lines as the proof of proposition 4.3 we can show a stronger result which is formulated in the same way as theorem 4.2.

PROPOSITION 4.4: *Let μ and f as in section 4.2. Assume that there exists a function $g:[0,\infty)\to[0,\infty)$ with $g(t)\to\infty$ as $t\to\infty$ such that*

i) $P-\lim_{T\to\infty} g(T)^{-2} \int_0^T \int_Z |fs,x|^2\nu(ds,dx)=1$

ii) $P-\lim_{T\to\infty} g(T)^{-2} \int_0^T \int_Z |f(s,x)|^2 I_{\{|\int_Z f(s,x)\mu(\{s\},dx)|>\epsilon g(T)\}}\nu(ds,dx)=0 \forall\epsilon>0.$

Then

$$\frac{1}{g(T)}\int_0^T\int_Z f(s,x)\mu(ds,dx)\xrightarrow{\mathcal{L}} N(0,1) \text{ as } T\to\infty.$$

A special case occurs when the measure μ is the difference of a counting process n and its absolutely continuous compensator $\int_0^{\circ}\lambda_s ds$. Then we can take $Z=\{1\}$ and proposition 4.4 reads

PROPOSITION 4.5: *If there exists a function* $g:[0,\infty)\infty[0,\infty)$ *with* $g(t)\to\infty$ *as* $t\to\infty$ *such that*

i) $P-\lim_{T\to\infty} g(T)^{-2}\int_0^T f_s^2\lambda_s ds = 1$

ii) $P-\lim_{T\to\infty} g(T)^{-2}\int_o^T f_s^2 I_{\{|f_s|>\epsilon g(T)\}}\lambda_s ds = 0, \ \forall\epsilon>0.$

Then

$$\frac{1}{g(T)}\int_0^T f_s(dn_s-\lambda_s ds)\xrightarrow{\mathcal{L}} N(0,1).$$

REMARK: Assumption 4.5.ii is certainly satisfied if $g(t)^{-1}\sup\{|f_s|:s\in[0,t]\}\leqslant c(t)$, where $c(t)\downarrow 0$ for $t\to\infty$.

5. Asymptotic Distributions of Recursive Estimators

5.1. *Least squares algorithm*

In this section we will show that the algorithm (2.1), (2.2) provides us with estimators $\hat{\theta}_t$ that are asymptotically normally distributed if we impose some additional requirements on the process ϕ. It immediately follows from (2.1), (2.2) that

$$\hat{\theta}_t = R_t[R_0^{-1}\hat{\theta}_0 + \int_0^t \phi_s dn_s] \tag{5.1}$$

and

$$\tilde{\theta}_t = \hat{\theta}_t-\theta_0 = R_t[R_0^{-1}(\hat{\theta}_0-\theta_0) + \int_0^t \phi_s dm_s] \tag{5.2}$$

Introduce the vector valued martingale

$$M_t = \int_0^t \phi_s dm_s \tag{5.3}$$

then

$$<M>_t = \int_0^t \phi_s\phi_s^T\phi_s^T\theta_0 ds$$

Clearly the distributions of $\hat{\theta}_t$ and $\tilde{\theta}_t$ are governed by the one of M_t. For the latter we have the following result

THEOREM 5.1: *Let M be as defined in (5.3). Assume that there exists a function* μ*:* $[0,\infty)\to[0,\infty)$ *with*

$\mu(t)\to\infty$ as $t\to\infty$ such that

i) $P-\lim_{t\to\infty}\mu(t)^{-1}<M>_t = D$, where $D\in\mathbb{R}^{d\times d}$ *is* a positive non random matrix

ii) $P-\lim_{t\to\infty}\mu(t)^{-1}\int_0^t \phi_s^T\phi_s I_{\{\phi_s^T\phi_s>\epsilon\mu(t)\}}\phi_s^T\theta_0 ds = 0, \forall\epsilon>0.$

Then

$$<M>_t^{-\frac{1}{2}}M_t \xrightarrow{\mathfrak{L}} N(0,I).$$

PROOF: i) Let $\lambda\in\mathbb{R}^d$ and define $M_t^\lambda=\lambda^T D^{-\frac{1}{2}}M_t=\int_0^t \lambda^T D^{-\frac{1}{2}}\phi_s dm_s$. Then

$$<M^\lambda>_t = \lambda^T D^{-\frac{1}{2}}<M>_t D^{-\frac{1}{2}}\lambda$$

Hence

$$(\lambda^T\lambda\mu(t))^{-1}<M^\lambda>_t = (\lambda^T\lambda)^{-1}\lambda^T D^{-\frac{1}{2}}\mu(t)^{-1}<M>_t D^{-\frac{1}{2}}\lambda\to 1$$

in probability. Hence condition i in proposition 4.5 is satisfied with $g(t)^2=\lambda^T\lambda\mu(t)$.
In order to establish *SARJ 2* (condition 4.5.ii) we compute

$$(\lambda^T\lambda\mu(t))^{-1}\int_0^t \lambda^T D^{-\frac{1}{2}}\phi_s\phi_s^T D^{-\frac{1}{2}}\lambda I_{\{|\lambda^T D^{-\frac{1}{2}}\phi_s|>\epsilon(\lambda^T\lambda\mu(t))^{\frac{1}{2}}\}}\phi_s^T\theta_0 ds =$$

$$(\lambda^T\lambda\mu(t))^{-1}\lambda^T D^{-\frac{1}{2}}\int_0^t \phi_s\phi_s^T I_{\{\lambda^T D^{-\frac{1}{2}}\phi_s\phi_s^T D^{-\frac{1}{2}}\lambda>\epsilon^2\lambda^T\lambda\mu(t)\}}\phi_s^T\theta_0 ds D^{-\frac{1}{2}}\lambda\leqslant$$

$$(\lambda^T\lambda\mu(t))^{-1}\lambda^T D^{-1}\lambda\int_0^t \phi_s^T\phi_s I_{\{\lambda^T D^{-1}\lambda\phi_s^T\phi_s>\epsilon^2\lambda^T\lambda\mu(t)\}}\phi_s^T\theta_0 ds$$

which tends to zero in probability according to assumption ii since we can replace ϵ by $\epsilon^2\lambda^T\lambda(\lambda^T D^{-1}\lambda)^{-1}$. Now we have proved

$$(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}M_t^\lambda \xrightarrow{\mathfrak{L}} N(0,1)$$

ii) According to the Cramer-World device

$$\forall\lambda\in\mathbb{R}^d:(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}M_t^\lambda \xrightarrow{\mathfrak{L}} N(0,1)$$

if and only if

$$\mu(t)D^{-\frac{1}{2}}M_t \xrightarrow{\mathfrak{L}} N(0,1).$$

Since

$$\mu(t)^{-\frac{1}{2}}D^{-\frac{1}{2}}M_t = (D^{-\frac{1}{2}}\mu(t)^{-\frac{1}{2}}<M>_t^{\frac{1}{2}})<M>_t^{-\frac{1}{2}}M_t$$

and

$$D^{-\frac{1}{2}}\mu(t)^{-\frac{1}{2}}<M>_t^{\frac{1}{2}}\to I$$

in probability, we have finished the proof.

REMARK : Stronger conditions than 5.1.ii) are the corresponding Lindeberg or Lyapunov conditions

$$\forall\epsilon>0:\mu(t)^{-1}E\int_0^t \phi_s^T\phi_s I_{\{\phi_s^T\phi_s>\epsilon\mu(t)\}}\phi_s^T\theta_0 ds\to 0$$

$$\exists\delta>0:\mu(t)^{-1-\delta/2}E\int_0^t\|\phi_s\|^{3+\delta}ds\to 0,$$

where $\|\cdot\|$ denotes the (Euclidean) norm on $\mathbb{R}^d$.

COROLLARY 5.2: *Under the conditions of theorem 5.1 we have*

$$<M>_t^{-\frac{1}{2}}R_t^{-1}(\hat{\theta}_t-\theta_0)\xrightarrow{\mathcal{L}}N(0,I)$$

PROOF : $<M>_t^{-\frac{1}{2}}R_t^{-1}(\hat{\theta}_t-\theta_0) = <M>_t^{-\frac{1}{2}}[R_0^{-1}(\hat{\theta}_0-\theta_0) + M_t]$. The fact that $<M>_t^{-\frac{1}{2}}R_0^{-1}(\hat{\theta}_0-\theta_0)\to 0$ in propability (this follows from 5.1.i) gives us the desired result

REMARK: $<M>_t$ depends on the unknown parameter θ_0. As usual we can estimate $<M>_t$ by substituting $\hat{\theta}_t$, which is strongly consistent, for θ_0.

The examples given below are continuations of examples 2.1-2.3.

EXAMPLE 5.1: $\phi(t)=[1,1+sint]^T,\theta_0=(\theta_1,\theta_2)$. Take $\mu(t)=t$. Then we can calculate

$$\lim_{t\to\infty}\mu(t)^{-1}<M>_t = \begin{bmatrix}\theta_1+\theta_2 & \theta_1+\frac{3}{2}\theta_2\\ \theta_1+\frac{3}{2}\theta_2 & \frac{3}{2}\theta_1+\frac{5}{2}\theta_2\end{bmatrix}$$

which is a positive definite matrix. So assumption 5.1.i is satisfied. To establish that assumption 5.1.ii holds it is sufficient to remark thast $\phi_s^T\phi_s\leq 5$, Hence for $t>\frac{5}{\epsilon}$ we have

$$I_{\{\phi_s^T\phi_s>\epsilon t\}} = 0.$$

Another calculation shows that we have asymptotically

$$(\hat{\theta}_t-\theta_0) \approx N(0,\frac{1}{t}\begin{bmatrix}3\theta_1+\theta_2 & -2\theta_1-\theta_2\\ -2\theta_1-\theta_2 & 2\theta_1+2\theta_2\end{bmatrix})$$

EXAMPLE 5.2 : $\phi_t=[1,1+(-1)^{n_{t-}}],\theta_0=(\theta_1,\theta_2)$. Take $\mu(t)=t$. Then a simple calculation yields:

$$as-\lim_{t\to\infty}\mu(t)^{-1}<M>_t = \frac{\theta_1^2+2\theta_1\theta_2}{\theta_1+\theta_2}\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}$$

which is positive definite. As in example 5.1 $\phi_s^T\phi_s$ is bounded, so again assumption 5.1.ii trivially holds. Combined with an expression for R_t we can calculate that

$$(\hat{\theta}_t-\theta_0) \approx N(0,\frac{1}{t}\frac{\theta_1+\theta_2}{\theta_1^2+2\theta_1\theta_2}\begin{bmatrix}2\theta_1^2 & -\theta_1^2\\ -\theta_1^2 & (\theta_1+\theta_2)^2+\theta_2^2\end{bmatrix})$$

EXAMPLE 5.3: $\phi_t=[X_{t-}\ \ 1-X_{t-}]$. Again take $\mu(t)=t$. Then

$$as-\lim_{t\to\infty}\mu(t)^{-1}<M>_t = \frac{1}{\mu_1+\mu_0}\begin{bmatrix}\theta_1\mu_1 & 0\\ 0 & \theta_0\mu_0\end{bmatrix}$$

Since $\phi_t^T\phi_t=1$, again assumption 5.1.ii is trivially satisfied. Asymptotically we have

$$(\hat{\theta}_t-\theta_0) \approx N(0,\frac{\mu_1+\mu_0}{t}\begin{bmatrix}\theta_1/\mu_1 & 0\\ 0 & \theta_0/\mu_0\end{bmatrix})$$

5.2. Asymptotic normality of the AML algorithm

The purpose of this section is to show that the $\hat{\theta}_t$'s generated by (2.18) - (2.21) have a limiting distribution which is approximately normal. After some definitions we state a useful lemma.
Define the following matrix valued stochastic processes

$$\overline{Q}_t^{-1} = Q_0^{-1} + \int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds \tag{5.4}$$

$$V_t = \int_0^t \frac{\phi_s\phi_s^T}{(\phi_s^T\hat{\theta}_s)^2}\phi_s^T\theta_0 ds \tag{5.5}$$

Lemma 5.3: *Let $\hat{\theta}_t,Q_t$ as defined by (2.18) - (2.21)) and let the assumptions of theorem 3.2 be in force. Then*

i) $$as-\lim_{t\to\infty} \overline{Q}_t^{\frac{1}{2}}Q_t^{-1}\overline{Q}_t^{\frac{1}{2}} = I \tag{5.6}$$

ii) $$as-\lim_{t\to\infty} \overline{Q}_t^{\frac{1}{2}}V_t\overline{Q}_t^{\frac{1}{2}} = I. \tag{5.7}$$

Proof:

i) Let $\delta>0$ and fix ω, taken from the set with probability one where $\hat{\theta}_t(\omega)\to\theta_0$. Then there is $\tau=\tau(\omega)$ such that $\forall t\geq\tau$ we have $|\hat{\theta}_{it}-\theta_{0i}|\leq\delta$ for all components i.

Consequently $(1-\delta)\phi_t^T\theta_0\leq\phi_t^T\hat{\theta}_t\leq(1+\delta)\phi_t^T\theta_0$ for $t\geq\tau$. In the ordering of positive matrices we then have

$$\frac{1}{1+\delta}\int_\tau^t\frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}\leq Q_t^{-1}-Q_\tau^{-1}\leq\frac{1}{1-\delta}\int_\tau^t\frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds$$

or

$$\frac{1}{1+\delta}(\overline{Q}_t^{-1}-\overline{Q}_\tau^{-1})\leq Q_t^{-1}-Q_\tau^{-1}\leq\frac{1}{1-\delta}(\overline{Q}_t^{-1}-\overline{Q}_\tau^{-1})$$

which yields

$$\frac{1}{1+\delta}(I-\overline{Q}_t^{\frac{1}{2}}\overline{Q}_\tau^{-1}Q_t^{\frac{1}{2}})\leq Q_t^{\frac{1}{2}}(Q_t^{-1}-\overline{Q}_\tau^{-1})Q_t^{\frac{1}{2}}\leq\frac{1}{1-\delta}(I-\overline{Q}_t^{\frac{1}{2}}\overline{Q}_\tau^{-1}\overline{Q}_t^{\frac{1}{2}})$$

Now take limits for $t\to\infty$ and use that $\overline{Q}_t\to 0$ to get

$$\frac{1}{1+\delta}I\leq \liminf_{t\to\infty}\overline{Q}_t^{\frac{1}{2}}Q_t^{-1}\overline{Q}_t^{\frac{1}{2}}\leq\limsup_{t\to\infty}\overline{Q}_t^{\frac{1}{2}}Q_t^{-1}\overline{Q}_t^{\frac{1}{2}}\leq\frac{1}{1-\delta}I \tag{5.8}$$

Since (5.8) holds for all $\delta>0$ the proof of (5.6) is complete.

ii) The proof of (5.7) is analogous.

The following vector valued martingale is important. Define

$$M_t = \int_0^t\frac{\phi_s}{\phi_s^T\hat{\theta}_{s-}}dm_s \tag{5.9}$$

Notice that we have $<M>_t = V_t$.

THEOREM 5.4: *Assume that there exists a function* $\mu:[0,\infty)\rightarrow[0,\infty)$ *such that*

$$P-\lim_{t\to\infty}\ \mu(t)^{-1}\theta_0^T\Phi_t = 1 \tag{5.10}$$

Then

$$\overline{Q}_t^{-\frac{1}{2}} M_t \xrightarrow{\mathcal{L}} N(0,I).$$

PROOF: Let C be as in assumption 3.2.ii,

$$C = as-\lim_{t\to\infty}\frac{1}{\theta_0^T\Phi_t}\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds = as-\lim_{t\to\infty}\frac{1}{\theta_0^T\Phi_t}\overline{Q}_t^{-1}$$

Then we also have

$$P-\lim_{t\to\infty}\frac{1}{\mu(t)}\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds = C. \tag{5.11}$$

Define

$$M_t^\lambda = \lambda^T C^{-\frac{1}{2}} M_t = \int_0^t\frac{\lambda^T C^{-\frac{1}{2}}\phi_s}{\phi_s^T\hat{\theta}_{s-}}\ dm_s,$$

then

$$(\lambda^T\lambda\mu(t))^{-1}<M>_t =$$

$$= (\lambda^T\lambda)^{-1}\lambda^T C^{-\frac{1}{2}}\overline{Q}_t^{-\frac{1}{2}}\mu(t)^{-\frac{1}{2}}\overline{Q}_t^{\frac{1}{2}}V_t\overline{Q}_t^{\frac{1}{2}}\mu(t)^{-\frac{1}{2}}\overline{Q}_t^{-\frac{1}{2}}C^{-\frac{1}{2}}\lambda\xrightarrow{P}1$$

by (5.10), (5.11). Hence assumption 4.5.i is satisfied.
As in the proof of lemma 5.3, let $\tau(\omega)$ be such that $t\geqslant\tau(\omega)$ implies

$$|\phi_t^T\hat{\theta}_t-\phi_t^T\theta_0|\leqslant\phi_t^T\theta_0\delta.$$

Consider

$$(\lambda^T\lambda\mu(t))^{-1}\ \int_0^t\frac{(\lambda^T C^{-\frac{1}{2}}\phi_s)^2}{(\phi_s^T\hat{\theta}_s)^2}\ I_{\{|\frac{\lambda^T C^{-\frac{1}{2}}\phi_s}{\phi_s^T\hat{\theta}_s}|\geqslant\delta(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}\}}\phi_s^T\theta_0 ds \tag{5.12}$$

Let us split the integral in two pieces, one with integration bounds 0 and $t\wedge\tau$ and the second with bounds $t\wedge\tau$ and t. Then clearly $(\lambda^T\lambda\mu(t))^{-1}$ times the former integral tends to zero almost surely. Hence we continue our investigation of the second integral which is after multiplication with $(\lambda^T\lambda\mu(t))^{-1}$ less than

$$(1-\delta)^{-2}(\lambda^T\lambda\mu(t))^{-1}\int_{\tau\wedge t}^t\frac{(\lambda^T C^{-\frac{1}{2}}\phi_s)^2}{(\phi_s^T\theta_0)^2}I_{\{\frac{\lambda^T C^{-\frac{1}{2}}\phi_s}{\phi_s^T\theta_0(1-\delta)}\geqslant\delta(\lambda^T\lambda\mu(t))^{\frac{1}{2}}\}}\phi_s^T\theta_0 ds\leqslant$$

$$\leqslant(1-\delta)^{-2}(\lambda^T\lambda\mu(t))^{-1}\int_{t\wedge\tau}^t\frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s)^2}{(\phi_s^T\theta_0)^2}I_{\{\frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s}{(\phi_s^T\theta_0)^2(1-\delta)^2}\geqslant\delta^2\lambda^T\lambda\mu(t)\}}\phi_s^T\theta_0 ds \tag{5.13}$$

Now let t be such that

$$\mu(t)\geqslant\frac{\lambda^T C^{-1}\lambda}{\delta^2(1-\delta)^2\lambda^T\lambda\underline{\theta}^2},\ \text{ where } \underline{\theta} = \min\{\theta_{0i}, i=1,\dots,d\}$$

Then

$$\delta^2\lambda^T\lambda\mu(t) \geq \frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s}{\underline{\theta}^2(1-\delta)^2(\phi_s^T 1)^2} \geq \frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s}{(1-\delta)^2(\phi_s^T\theta_0)^2}$$

Consequently for large t the indicator appearing in the integral in (5.13) will be zero. As a result (5.12) converges to zero almost surely and a forteriori in probability, which gives us assumption 4.5.ii. Conclusion

$$(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}M_t^\lambda \xrightarrow{\mathcal{L}} N(0,I).$$

As in the proof of theorem 5.1 the Cramer - Wold device gives us

$$(\mu(t)C)^{-\frac{1}{2}}M_t \xrightarrow{\mathcal{L}} N(0,I)$$

if and only if

$$(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}M_t^\lambda \xrightarrow{\mathcal{L}} N(0,I),$$

which has just been proved.

Finally

$$\overline{Q}_t^{\frac{1}{2}}M_t = \mu(t)^{\frac{1}{2}}C^{\frac{1}{2}}\mu(t)^{-\frac{1}{2}}C^{-\frac{1}{2}}M_t.$$

We know from (5.11) that $\mu(t)^{\frac{1}{2}}\overline{Q}^{\frac{1}{2}}C^{\frac{1}{2}} \to I$ in probability, which completes the proof.

COROLLARY 5.5: *Under the assumptions of theorem 5.4*

i) $Q_t^{-\frac{1}{2}}\tilde{\theta}_t \xrightarrow{\mathcal{L}} N(0,I)$

ii) $\overline{Q}_t^{-\frac{1}{2}}\tilde{\theta}_t \xrightarrow{\mathcal{L}} N(0,I)$

PROOF: i) By writing out the stochastic differential equation for $Q_t^{-1}\tilde{x}_t$ one can show that the following relation holds

$$\tilde{x}_t = Q_t[\int_0^t \frac{\phi_s}{\phi_s^T\hat{\theta}_{s-}}dm_s + Q_0^{-1}(x_0-\theta_0)] \tag{5.14}$$

And consequently

$$Q_t^{-\frac{1}{2}}\tilde{\theta}_t = I_tQ_t^{-\frac{1}{2}}M_t + I_tQ_t^{-\frac{1}{2}}Q_0(x_0-\theta_0) + Q_t^{-\frac{1}{2}}(1-I_t)(\epsilon-\theta_0) \tag{5.15}$$

Since $I_t \to 1$ a.s. and $Q_t^{\frac{1}{2}} \to 0$ a.s. as $t \to \infty$ we see from (5.15) that the asymptotic distribution of $Q_t^{-\frac{1}{2}}\tilde{\theta}_t$ will be same as that of $Q_t^{\frac{1}{2}}M_t$. From lemma 5.3 we know that we can replace Q_t by $\overline{Q}_t$ and the conclusion follows from theorem 5.4.

ii) This is an immediate consequence of i)

The examples below are examples 3.1 - 3.3 continued.

EXAMPLE 5.4: $\phi s(t)=[1,1+\sin t]^T$ Take $\mu(t)=t$. Then one finally gets after some tedious calculations: Aproximately

$$\tilde{\theta}_t \approx N(0, \frac{\theta_1+\theta_2+\sqrt{\theta_1^2+2\theta_1\theta_2}}{t} V)$$

with

$$V = \begin{bmatrix} (\frac{1}{\theta_1}-\frac{\theta_1}{\theta_2})\sqrt{\theta_1^2+2\theta_1\theta_2}+\frac{\theta_1^2}{\theta_2^2} & -\frac{1}{\theta_2}(\sqrt{\theta_1^2+2\theta_1\theta_2}-\theta_1) \\ -\frac{1}{\theta_2}(\sqrt{\theta_1^2+2\theta_1\theta_2}-\theta_1) & 1 \end{bmatrix}$$

EXAMPLE 5.5: $\phi_t=[1,1+(-1)^{n_{t-}}]$. One gets

$$as-\lim_{t\to\infty}\frac{1}{t}\overline{Q}_t^{-1} = \frac{\theta_1+\theta_2}{\theta_1+2\theta_2}\begin{bmatrix}\theta_1+\theta_2 & \theta_1 \\ \theta_1 & 2\theta_1\end{bmatrix}$$

and the asymptotic distribution

$$\tilde{\theta}_t \approx N(0,\frac{1}{(\theta_1^2+\theta_1\theta_2)t}\begin{bmatrix}2\theta_1 & -\theta_1 \\ -\theta_1 & \theta_1+\theta_2\end{bmatrix})$$

EXAMPLE 5.6: $\phi_t=[X_{t-},1-X_{t-}]$. Here

$$as-\lim_{t\to\infty}\frac{1}{t}\overline{Q}_t^{-1} = \frac{1}{\mu_1+\mu_0}\begin{bmatrix}\mu_1/\theta_1 & 0 \\ 0 & \mu_0/\theta_0\end{bmatrix}$$

and asymptotically

$$\tilde{\theta}_t \approx N(0,\frac{\mu_1+\mu_0}{t}\begin{bmatrix}\theta_1/\mu_1 & 0 \\ 0 & \theta_0/\mu_0\end{bmatrix}).$$

We see that in this case the asymptotic variance of $\tilde{\theta}_t$ is the same as in example 5.3.

6. Some Remarks

6.1. Other possible limit Distributions

The basic assumption in getting a limiting distribution for $\hat{\theta}_t$ or $\tilde{\theta}_t$ which is gaussian is 5.1.i or (5.10) depending on the algorithm. This assumption more or less tells us that the covaration process of the martingale M becomes deterministic as t grows. If this assumption is dropped one can still derive results for the asymptotic distribution of $\hat{\theta}_t$. The idea then is to performe some random time transformation $\tau=f(t)$ after which the transformed version of $\langle M \rangle$ becomes deterministic. For the transformed asgorithm (which looks the same in the AML case (2.22), (2.23)) we can then infer asymptotic normality as τ tends to infinity. In the AML case a useful transformation is $\tau=\Phi_t^T\theta_0$. This idea has also been carried out in [11] for the off-line maximum likelihood estimation problem.

6.2. Asymptotic Efficiency

From the examples 5.1 - 5.6 it becomes clear that the asymptotic distributions of $\tilde{\theta}_t$, generated by (2.1), (2.2) or (2.18)-(2.21) will differ in general. Thus they cannot both give us efficient estimators. In general we have the following Cramer-Rao inequality. An unbaised estimator of θ based on the observations in $[0,t]$ has a covariance matrix which is at least

$$C_t(\theta) = \{E_\theta[\frac{\partial}{\partial\theta}\log L_t(\theta)][\frac{\partial}{\partial\theta}\log L_t(\theta)]^T\}^{-1} \tag{6.1}$$

where the likelihood ratio $L_t(\theta)$ is as in (2.6). Calculation of (6.1) gives us

$$C_t(\theta) = \left[E_\theta\int_0^t\frac{\phi_s\phi_s^T}{\phi_s^T\theta}ds\right]^{-1} \tag{6.2}$$

This means that $\hat{\theta}_t$ is an asymptotically efficient estimator if we have

$$C_t(\theta_0)^{1/2}(\hat{\theta}_t-\theta_0)\xrightarrow{\mathcal{L}}N(0,I) \tag{6.3}$$

Clearly by comparing corollary 5.2 and (6.3) we see that the LS estimator of θ will not be asymptotically efficient in general except for some specific choices of ϕ (see examples 5.3, 5.6) On the other hand the AML estimator given by (2.18) -(2.21) is a good candidate for being an asymptotically efficient estimator by corollary 5.5.ii. We will indeed have this property as soon as $C_t(\theta_0)\overline{Q}_t^{-1}\to I$ in probability. However assumption (5.10) in theorem 5.4 does not seem to be sufficient for guaranteeing this. But if we impose as an additional requirement that $\mu(t)^{-1}C_t(\theta_0)^{-1}\to C$ then indeed from (5.11)

$$C_t(\theta_0)\overline{Q}_t^{-1} = C_t(\theta_0)\mu(t)\mu(t)^{-1}\overline{Q}_t^{-1}\xrightarrow{P}C^{-1}C = I.$$

In fact under the assumption (5.10) requiring $\mu(t)^{-1}C_t(\theta_0)$ to converge to C is nothing else than demanding the collection $\{C_t(\theta_0)\overline{Q}_t^{-1}\}_{t\geq 0}$ to be uniformly integrable.
Let us summarize the discussion of the preceeding paragraph in

PROPOSITION 6.1: *Assume that there exists a function* $\mu:[0,\infty)\to[0,\infty)$ *such that*

$$P-\lim_{t\to\infty}\mu(t)^{-1}\Phi_t^T\theta_0 = 1$$

$$\lim_{t\to\infty}\mu(t)^{-1}C_t(\theta_0)^{-1} = C$$

where C *is as in assumption 3.2.ii. Then the AML estimator* $\hat{\theta}_t$ *generated by (2.18) - (2.21) is asymptotically efficient.*

One easily checks that one can take in the preceeding examples $\mu(t)=t$.

6.3. Relation with Hellinger process
The Hellinger process is a convenient tool to describe the relation between two probability measures on the whole trajectory space of a certain stochastic process. In the counting process case there is an explicit expression available in terms of the compensator of the counting process. For the model (1.3) the Hellinger process of order ½ is

$$h(\theta,\theta_0)_t = \int_0^t(\sqrt{\phi_s^T\theta}-\sqrt{\phi_s^T\theta_0})^2ds = \int_0^t\phi_s^T\theta_0\left[\sqrt{\frac{\phi_s^T\theta_s}{\phi_s^T\theta_0}}-1\right]^2ds$$

The following theorem can be found in [3, p.253] and tells us whether or not we can distinguish between θ and θ_0.

THEOREM 6.2: *Let* P_θ,P_{θ_0} *be two measures on the whole counting process trajectory space such that under* P_θ (P_{θ_0}) n_t *admits the intensity process* $\phi_t^T\theta$ ($\phi_t^T\theta_0$). *Then* P_θ *and* P_{θ_0} *are mutually singular if and only if* $\lim_{t\to\infty}h(\theta,\theta_0)_t=\infty$ *with probability one.*

Since $\dfrac{\phi_s^T\theta}{\phi_s^T\theta_0}$ is a bounded function of $s\in[0,\infty)$ if all components of θ_0 are positive the statement $\lim_{t\to\infty}h(\theta,\theta_0)_t=\infty$ is equivalent with

$$\lim\int_0^t\frac{(\phi_s^T(\theta-\theta_0))^2}{\phi_s^T\theta_0}ds = \infty \tag{6.5}$$

by

$$(\sqrt{x}-1)^2 \leqslant (x-1)^2 \leqslant c(\sqrt{x}-1)^2 \quad \text{for } x \leqslant (\sqrt{c}-1)^2, c \geqslant 1.$$

which says that

$$\liminf \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \theta_0} ds = \infty.$$

This is clearly implied by our indentifiability condition in theorem 3.2. One might hope that the converse would also be true. This is not the case, despite the specific form (1.3) of the intensity process. In a more general situation this has already noticed in [11]. There the notion of θ_0-distinctness has been introduced which appears to contain the identifiability criterion of theorem 3.2. At this point it is not clear whether we can relax our indentifiability condition to (6.5) or that it is the price we have to pay to get recursive estimators.

References

1. P. Brémaud, (1981). *Point Processes and Queues,* Springer.
2. R.B. Davies, (1973). Asymptotic inference in stationary gaussian time-series, *Adv. Appl. Prob. 5,* pp 469-497.
3. J. Jacod, (1979). *Calcul Stochastique et Problèmes de Martingales, LNM 714,* Springer
4. F. Konecny, (1984). Parameter estimation for point processes with partial observations, A filtering approach, *System and Control letters 4,* pp 281-286.
5. Yu.A. Kutoyants, (1979). Intensity parameters estimation of an inhomogeneous Poisson process, *Problems of Control and Information Theory, Vol. 8, no. 2,* pp 1-13.
6. Yu.A. Kutoyants, (1979). Locally asymptotic normality for processes of poisson type, *Izv. Ak. Nauk Armyanskoi SSSR. Matematika Vol. 14, no 1,* pp 3-20.
7. Yu.N. Lin'kov, (1981). The asymptotic normality of stochastic integrals and the statistics of discontinuous processes, *Theory of Probability and Mathematical Statistics no. 23,* pp 97-107.
8. Yu.N. Lin'kov, (1982), Estimates of parameters of counting processes, *Problems of Information Transmission 18,* pp 63-75.
9. Y. Ogata, (1978), The asymptotic behaviour of maximum likelihood estimators for stationary point processes, *Ann. Inst. Statist. Math. 30, Part A,* pp 243-261.
10. R. Rebolledo, (1980). Central limit theorems for local martingales, *Z. Wahrscheinlichkeitstheorie verw. Geb. 51,* pp 269-286.
11. B. Sagalovsky, (1983). Maximum likelihood and related estimation methods in point processes and point process systems, *Ph.D. Dissertation* University of California, Berkely.
12. J.H. van Schuppen, (1983), Convergence results for continuous time adaptive stochastic filtering algorithms, *J. Math. Anal. Appl. 96,* pp 209-225.
13. P.J.C. Spreij, (1985). An on-line parameter estimation algorithm for counting process observations, *IEEE Trans. Information Theory. to appear.*