PLISKA
STUDIA MATHEMATICA
BULGARICA

ПЛИСКА
БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. Bulgar. **22** (2013), 71–88

PLISKA
STUDIA MATHEMATICA
BULGARICA

A GENERALIZED QUASI-LIKELIHOOD ESTIMATOR FOR NONSTATIONARY STOCHASTIC PROCESSES–ASYMPTOTIC PROPERTIES AND EXAMPLES

Christine Jacob

Abstract. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a real stochastic process on $(\Omega, \mathcal{F}, P_{\boldsymbol{\theta}_0})$, where θ_0 is a unknown p-dimensional parameter. We propose a GQLE (Generalized Quasi-Likelihood Estimator) of θ_0 based on a single trajectory of the process and defined by $\widehat{\theta}_n := \arg\min_\theta \sum_{k=1}^n \Psi_k(Z_k, \boldsymbol{\theta})$, where $\Psi_k(z, \boldsymbol{\theta})$ is $\mathcal{F}_{k-1}$-measurable, $\{\mathcal{F}_n\}_n$ being an increasing sequence of σ-algebras. This class of estimators includes many different types of estimators such as conditional least squares estimators, least absolute deviation estimators and maximum likelihood estimators, and allows missing data, outliers, or infinite conditional variance. We give general conditions leading to the strong consistency and the asymptotic normality of $\widehat{\theta}_n$. The key tool is a uniform strong law of large numbers for martingales. We illustrate the results in the branching processes setting.

1. Introduction. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a real discrete time stochastic process on $(\Omega, \mathcal{F}, P_{\boldsymbol{\theta}_0})$, $\boldsymbol{\theta}_0 \in \mathbb{R}^p$, $p < \infty$, that may depend on an environmental process $\{\mathbf{U}_n\}_{n\in\mathbb{N}}$. The processes $\{Z_n\}_{n\in\mathbb{N}}$ and $\{\mathbf{U}_n\}_{n\in\mathbb{N}}$ are observed. They can correspond to the discrete time observations of some underlying processes in continuous time. As examples, cite (non)linear time series, *ARMAX* models, Markov

2010 *Mathematics Subject Classification:* 62F12, 62M05, 62M09, 62M10, 60G42.

Key words: Quasi-likelihood estimator, minimum contrast estimator, least-squares estimator, least absolute deviation estimator, maximum likelihood estimator, uniform strong law of large numbers for martingales, nonstationary stochastic process, stochastic regression, consistency, asymptotic distribution.

chain, or branching processes. We propose to estimate $\boldsymbol{\theta}_0$ from a single trajectory $\{Z_0, Z_1, \ldots, Z_n, \mathbf{U}_0, \mathbf{U}_1, \mathbf{U}_n\}$ by the following GQLE (Generalized Quasi-Likelihood Estimator) that we could also call "generalized minimum contrast estimator":

$$\widehat{\boldsymbol{\theta}}_n := \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S_n(\boldsymbol{\theta}),\ S_n(\boldsymbol{\theta}) := \sum_{k=1}^{n} \Psi_k(Z_k, \boldsymbol{\theta}),\ \boldsymbol{\Theta} \subset \mathbb{R}^p,\ \boldsymbol{\Theta} \text{ compact},\ \boldsymbol{\theta}_0 \in \overset{\circ}{\boldsymbol{\Theta}},$$

where $\Psi_k(z, \boldsymbol{\theta})$ is $\mathcal{F}_{k-1}$-measurable, $\{\mathcal{F}_k\}_k$ being an increasing sequence of σ-algebras included in $\mathcal{F}$. The quantities $\{\Psi_k(z, \boldsymbol{\theta})\}_k$ may have very general forms and may handle missing data, outliers, and processes with infinite conditional variance if we define $\mathcal{F}_{k-1} = \sigma\text{-}\big(1_{\{Z_k\in I\}}, \{Z_{k-l}\}_{l\geq 1}, \{\mathbf{U}_{k-l}\}_{l\geq 0}\big)$, where I is a finite or infinite subset of $\mathbb{R}$, and if $\Psi_k(z, \boldsymbol{\theta}) = \lambda_k \tilde{\Psi}_k(z, \boldsymbol{\theta})$, where λ_k and $\tilde{\Psi}_k(z, \boldsymbol{\theta})$ are $\mathcal{F}_{k-1}$-measurable and λ_k is the "weight" of the contrast assumed to be null if the elements of $\{Z_k\}_k$ and $\{\mathbf{U}_k\}_k$ involved in $\Psi_k(z, \boldsymbol{\theta})$ do not belong to I. This type of weights allows to keep in the contrast only bounded quantities when it is necessary (otherwise $I = \mathbb{R}$). It is for example the case when the conditional variance of the process is infinite at each time.

We present some examples of estimators, assuming to simplify the presentation that $\lambda_k \overset{a.s.}{=} 1$, for all k.

1. *Conditional Least Squares Estimator (or "Quasi-likelihood estimator")*:

$$\Psi_k(Z_k, \boldsymbol{\theta}) = (Z_k - g_k(\boldsymbol{\theta}))^2,\ g_k(\boldsymbol{\theta}) := E_{\boldsymbol{\theta}}\Big(Z_k | \mathcal{F}_{k-1}\Big). \tag{1}$$

When more generally $\Psi_k(z, \boldsymbol{\theta}) = \lambda_k \tilde{\Psi}_k(z, \boldsymbol{\theta})$, where $\tilde{\Psi}_k(z, \boldsymbol{\theta}) = (Z_k - g_k(\boldsymbol{\theta}))^2$, and if $\widehat{\boldsymbol{\theta}}_n(\{\lambda_k\})$ denotes the corresponding estimator, then according to [5], $\widehat{\boldsymbol{\theta}}_n(\{\lambda_k\})$ is O_F-optimal at time n (fixed sample optimality) among the $\{\widehat{\boldsymbol{\theta}}_n(\{\nu_k\})\}$, if the information criterion

$$E_{\boldsymbol{\theta}_0}\Big(\ddot{S}_n^T(\boldsymbol{\theta}_0)\Big)\Big(E_{\boldsymbol{\theta}_0}\Big(\dot{S}_n(\boldsymbol{\theta}_0)\dot{S}_n^T(\boldsymbol{\theta}_0)\Big)\Big)^{-1} E_{\boldsymbol{\theta}_0}\Big(\ddot{S}_n(\boldsymbol{\theta}_0)\Big) \tag{2}$$

is maximal when $\{\nu_k\} = \{\lambda_k\}$, where $\dot{S}_n(\boldsymbol{\theta}_0) := \big(\partial S_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\big)(\boldsymbol{\theta}_0)$, $\ddot{S}_n(\boldsymbol{\theta}_0) := \big(\partial \dot{S}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\big)(\boldsymbol{\theta}_0)$. The maximality is got according to the partial order of nonnegative definite matrices (Loewner partial order: $\mathbf{A} \geq \mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is positive semidefinite). The information criterion (2) is the natural generalization of the Fisher information.

When $p = 1$, an optimal set $\{\lambda_k\}$ satisfies $\lambda_k = \alpha\big(Var_{\boldsymbol{\theta}_0}\big(Z_k | \mathcal{F}_{k-1}\big)\big)^{-1}$, where α is any nonnull constant ([3], Theorem 2.1 p.14 [5]).

2. *Least Absolute Deviation Estimator (or "L_1-norm estimator")*:

$$\Psi_k(Z_k, \boldsymbol{\theta}) = |Z_k - g_k(\boldsymbol{\theta})|,\ g_k(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\big(Z_k|\mathcal{F}_{k-1}\big). \tag{3}$$

3. *Maximum Likelihood Estimator*:

$$\Psi_k(Z_k, \boldsymbol{\theta}) = -\ln p_{\boldsymbol{\theta}}(Z_k|\mathcal{F}_{k-1})\ \text{(conditional likelihood)}. \tag{4}$$

Remark. Robust estimators reducing the effects of outliers that are based on some percentile of the data, such as Winsorised Estimators or Trimmed Estimators, do not belong to this class because the corresponding quantities $\{\Psi_k(z, \boldsymbol{\theta})\}_k$ depend on the whole set of observations until time n, and therefore are $\mathcal{F}_n$-measurable but not $\mathcal{F}_{k-1}$-measurable.

We give here conditions leading to the asymptotic properties, as $n \to \infty$, of $\widehat{\boldsymbol{\theta}}_n$: its strong consistency ($\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s.}{=} \boldsymbol{\theta}_0$), and its asymptotic distribution (existence in distribution, for some matrix $\boldsymbol{\Upsilon}_n$, of $\lim_n \boldsymbol{\Upsilon}_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$).

Since, except for particular classes of models and contrasts, $\widehat{\boldsymbol{\theta}}_n$ has generally no explicit expression, the proofs will be indirect proofs based on the properties of the contrast $S_n(\boldsymbol{\theta})$. The key tool will be a USLLNM (Uniform Strong Law of Large Numbers for Martingales), direct generalization of Proposition 3.1 in [6]. This USLLNM allows to get asymptotic properties of the estimators in a very general nonlinear, nonindependent and nonstationary setting.

Section 2 is devoted to the strong consistency of $\widehat{\boldsymbol{\theta}}_n$, where two different types of proofs based on the contrast properties are compared. In Section 3, we deal with the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$. Then Section 4 illustrates the results of the previous sections in the frame of a branching process with a long memory. A short conclusion is given in Section 5. We will see that the strong consistency of the estimator is easily got on the nonextinction set of the process, provided that $\boldsymbol{\theta}$ is identifiable in $\{\Psi_k(Z_k, \boldsymbol{\theta})\}_{k=1}^{\infty}$, while the asymptotic distribution of the estimator requires some stronger properties on the behavior of the process.

Since we do not deal here with inverse functions, in order to simplify the notations, we will write everywhere "$f^{-1}(\boldsymbol{\theta})$" for "$\big(f(\boldsymbol{\theta})\big)^{-1}$". We will also write "martingale" for a martingale on $(\Omega, \mathcal{F}, P_{\boldsymbol{\theta}_0})$ adapted to the filtration $\{\mathcal{F}_n\}_n \subset \mathcal{F}$. Finally if $\mathbf{A}$ is a real matrix, then $\|\mathbf{A}\|^2 := \lambda_{\max}(\mathbf{A}^T\mathbf{A})$ (largest eigenvalue of $\mathbf{A}^T\mathbf{A}$), and $\lambda_{\min}(\mathbf{A}^T\mathbf{A})$ is the smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$.

2. Strong consistency. To prove the consistency of $\widehat{\boldsymbol{\theta}}_n$, we use some properties of $S_n(\boldsymbol{\theta})$. The first type of proof will be based on the first order Taylor series expansion of $\dot{S}_n(\boldsymbol{\theta})$, when $\Psi_k(\boldsymbol{\theta})$ is twice differentiable in $\boldsymbol{\theta}$. Thus, from

$\dot{S}_n(\widehat{\boldsymbol{\theta}}_n) = 0 = \dot{S}_n(\boldsymbol{\theta}_0) + \ddot{S}_n(\boldsymbol{\theta}_n)\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)$, where $\boldsymbol{\theta}_n$ "lies" between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, we deduce as usual that

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = -\ddot{S}_n^{-1}(\boldsymbol{\theta}_n)\dot{S}_n(\boldsymbol{\theta}_0). \tag{5}$$

Note that if $\ddot{S}_n(\boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$ and if there exists a matrix $\boldsymbol{\Upsilon}_n$ such that $\boldsymbol{\Upsilon}_n\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)$ converges in distribution to a centered variable, then $\widehat{\boldsymbol{\theta}}_n$ is weakly consistent. However, in a very general setting, $\ddot{S}_n(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$. So the convergence in distribution of $\boldsymbol{\Upsilon}_n\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)$ requires to prove first the consistency of $\widehat{\boldsymbol{\theta}}_n$.

The second type of proof is based on the minimum contrast idea, that is $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta})$ should tend to $S_n(\boldsymbol{\theta}_0)$, as $n \to \infty$.

We study conditions leading to the strong consistency of $\widehat{\boldsymbol{\theta}}_n$ through each type of proof. We will see that the proof based on the first order Taylor's expansion of $\dot{S}_n(\boldsymbol{\theta})$ requires unnecessary strong conditions, contrary to the proof based on the minimum contrast property.

2.1. Strong consistency of $\widehat{\boldsymbol{\theta}}_n$ based on the Taylor's expansion of $\dot{S}_n(\widehat{\boldsymbol{\theta}}_n)$. We may write (5) in the following form:

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = -\Big(\boldsymbol{\Phi}_n^{-1}\ddot{S}_n(\boldsymbol{\theta}_n)\Big)^{-1}\boldsymbol{\Phi}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0),$$

where $\boldsymbol{\Phi}_n$ is a $p \times p$ matrix. We want to define conditions on $\boldsymbol{\Phi}_n$ such that $\lim_n \boldsymbol{\Phi}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0) \overset{a.s.}{=} 0$ and $\overline{\lim}_n \|\Big(\boldsymbol{\Phi}_n^{-1}\ddot{S}_n(\boldsymbol{\theta}_n)\Big)^{-1}\| \overset{a.s.}{<} \infty$. We will use a SLLNM (Strong Law of Large Numbers for Martingales) to prove that $\lim_n \boldsymbol{\Phi}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0) \overset{a.s.}{=} 0$. We point out that, in these theorems, $\boldsymbol{\Phi}_n$ should be $\mathcal{F}_{n-1}$-measurable. Thus $\ddot{S}_n(\boldsymbol{\theta}_n)$ which is $\mathcal{F}_n$-measurable, is not a good candidate for $\boldsymbol{\Phi}_n$.

Proposition 1. *Assume that $S_n(\boldsymbol{\theta})$ is twice differentiable in $\boldsymbol{\theta}$ and that $\dot{S}_n(\boldsymbol{\theta}_0)$ is a martingale. Let $\boldsymbol{\Phi}_n = \sum_{k=1}^n E\big(\dot{\Psi}_k(Z_k, \boldsymbol{\theta}_0)\dot{\Psi}_k^T(Z_k, \boldsymbol{\theta}_0)|\mathcal{F}_{k-1}\big)$. Then*

$$\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \{\overline{\lim}_n \sup_{\boldsymbol{\theta}} \|\ddot{S}_n(\boldsymbol{\theta})^{-1}\boldsymbol{\Phi}_n\| < \infty, \quad \lim_n \lambda_{\min}\big(\boldsymbol{\Phi}_n\big) = \infty,\ \lim_n \big(\ln(\|\boldsymbol{\Phi}_n\|)\big)^{\nu}\big(\lambda_{\min}(\boldsymbol{\Phi}_n)\big)^{-1} = 0\}, \nu > 1. \tag{6}$$

Proof. The property $\lim_n \boldsymbol{\Phi}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0) \overset{a.s.}{=} 0$ is directly deduced from a SLLNM ([4] ($p = 1$), [8] ($1 \leq p < \infty$)).

1. Consider the case $p = 1$. Note first that the last condition of (6) is automatically checked in this case. We use here the following classical SLLNM

[4], applied to $M_n := \dot{S}_n(\boldsymbol{\theta})$, that is $X_k := \dot{\Psi}_k(Z_k, \theta)$: let $\{M_n\}_n$ be a martingale, that is $M_n = \sum_{k=1}^n X_k$, where $X_k := M_k - M_{k-1}$, $E(X_k|\mathcal{F}_{k-1}) = 0$, and let $\{\Phi_n\}$ be a non decreasing sequence ($\Phi_n \leq \Phi_{n+1}$) such that Φ_n is $\mathcal{F}_{n-1}$-measurable. Then

$$\text{(7)} \quad \lim_n \Phi_n^{-1} M_n \overset{a.s.}{=} 0 \text{ on } \{\lim_n \Phi_n{=}\infty,\ \sum_n \Phi_n^{-2} E\big(X_n^2|\mathcal{F}_{n-1}\big){<}\infty\}.$$

Moreover when $\Phi_n = \sum_{k=1}^n E\big(X_k^2|\mathcal{F}_{k-1}\big)$, then (7) is reduced to

$$\lim_n \Phi_n^{-1} M_n \overset{a.s.}{=} 0 \text{ on } \{\lim_n \Phi_n{=}\infty\}$$

because any real sequence $\{x_k^2\}_k$ satisfies (see [4] p.158, [6]):

$$\text{(8)} \qquad \sum_{n=1}^{\infty} \big(\sum_{k=1}^{n} x_k^2\big)^{-2} x_n^2 {<} \infty.$$

Remark. Writing $E\big(X_k^2|\mathcal{F}_{k-1}\big) =: x_k^2$, and $\Phi_n^2 =: \sum_{k=1}^n y_k^2$, then we may have $\sum_n \big(\sum_{k=1}^n y_k^2\big)^{-2} x_n^2 {<} \infty$ with $x_n^2 > y_n^2$. For instance, take $y_n^2 = b^n$ and $x_n^2 = a^n$ with $b^2 > a > b > 1$. So Proposition 1 may be extended to some Φ_n tending to ∞ at a smaller rate than $\sum_{k=1}^n E\big(\dot{\Psi}_k(Z_k, \theta_0)\dot{\Psi}_k^T(Z_k, \theta_0)|\mathcal{F}_{k-1}\big)$.

2. Consider now the general case $1 \leq p < \infty$. We use a SLLNMM (SLLN for Multivariate Martingales) [8]: let $\{\mathbf{M}_n\}_n$ defined by $\mathbf{M}_n := \sum_{k=1}^n \mathbf{X}_k$ be a multivariate martingale, and let $\boldsymbol{\Phi}_n$ be a symmetric $\mathcal{F}_{n-1}$-measurable matrix such that $\boldsymbol{\Phi}_{n+1}^2 - \boldsymbol{\Phi}_n^2$ is nonnegative definite, that is $\mathbf{V}^T\big(\boldsymbol{\Phi}_{n+1}^2 - \boldsymbol{\Phi}_n^2\big)\mathbf{V} \geq 0$, for all $p \times 1$ vector $\mathbf{V}$. Then

$$\text{(9)} \quad \lim_n \boldsymbol{\Phi}_n^{-1}\mathbf{M}_n \overset{a.s.}{=} 0 \quad \text{on } \{\lim_n \lambda_{\min}\big(\boldsymbol{\Phi}_n\big){=}\infty,\ \sum_n E\big(\|\boldsymbol{\Phi}_n^{-1}\mathbf{X}_n\|^2|\mathcal{F}_{n-1}\big){<}\infty\}.$$

Moreover in the particular case $\boldsymbol{\Phi}_n = \sum_{k=1}^n E\big(\mathbf{X}_k\mathbf{X}_k^T|\mathcal{F}_{k-1}\big)$, then

$$\text{(10)} \quad \lim_n \boldsymbol{\Phi}_n^{-1}\mathbf{M}_n \overset{a.s.}{=} 0 \quad \text{on } \{\lim_n \lambda_{\min}\big(\boldsymbol{\Phi}_n\big){=}\infty,\ \lim_n \big(\ln(\|\boldsymbol{\Phi}_n\|)\big)^{\nu}\lambda_{\min}^{-1}(\boldsymbol{\Phi}_n){=}0\},\ \nu > 1.$$

□

Remark. When $p > 1$, the condition $\lim_n \big(\ln(\|\boldsymbol{\Phi}_n\|)\big)^{\nu} \big(\lambda_{\min}(\boldsymbol{\Phi}_n)\big)^{-1} \overset{a.s.}{=} 0$ is not always checked contrary to the case $p = 1$ (take $\lambda_{\min}(\boldsymbol{\Phi}_n) = O(n)$, $\|\boldsymbol{\Phi}_n\| = O(\beta^n)$, $\beta > 1$).

Consider the conditions of Proposition 1 applied to (1), (3), (4):

1. $\Psi_k(Z_k, \boldsymbol{\theta}) = (Z_k - g_k(\boldsymbol{\theta}))^2$. Then $\dot{S}_n(\boldsymbol{\theta}_0) = -2\sum_{k=1}^n e_k(\boldsymbol{\theta}_0)\dot{g}_k(\boldsymbol{\theta}_0)$ and $\ddot{S}_n(\boldsymbol{\theta}) = 2\sum_{k=1}^n \dot{g}_k(\boldsymbol{\theta})\dot{g}_k^T(\boldsymbol{\theta}) - 2\sum_{k=1}^n e_k(\boldsymbol{\theta})\ddot{g}_k(\boldsymbol{\theta})$, where $e_k(\boldsymbol{\theta}) := Z_k - g_k(\boldsymbol{\theta})$. The quantities $\sum_{k=1}^n e_k(\boldsymbol{\theta}_0)\dot{g}_k(\boldsymbol{\theta}_0)$ and $\sum_{k=1}^n e_k(\boldsymbol{\theta}_0)\ddot{g}_k(\boldsymbol{\theta}_0)$ are martingales. We see that unfortunately the first condition of (6) is generally not checked in the nonlinear case.

2. $\Psi_k(Z_k, \boldsymbol{\theta}) = |Z_k - g_k(\boldsymbol{\theta})|$ is not differentiable;

3. $\Psi_k(Z_k, \boldsymbol{\theta}) = -\ln p_k(\boldsymbol{\theta})$, where $p_k(\boldsymbol{\theta}) := p_{\boldsymbol{\theta}}(Z_k|\mathcal{F}_{k-1})$. Then $\dot{S}_n(\boldsymbol{\theta}_0) = -\sum_{k=1}^n \dot{p}_k(\boldsymbol{\theta}_0)p_k^{-1}(\boldsymbol{\theta}_0)$. This implies that $\dot{S}_n(\boldsymbol{\theta}_0)$ is a martingale, under the classical assumption that integral and derivatives may be exchanged. In addition $\ddot{S}_n(\boldsymbol{\theta}) = \sum_{k=1}^n \dot{p}_k(\boldsymbol{\theta})\dot{p}_k^T(\boldsymbol{\theta})p_k^{-2}(\boldsymbol{\theta}) - \sum_{k=1}^n \ddot{p}_k(\boldsymbol{\theta})p_k^{-1}(\boldsymbol{\theta})$, where $\sum_{k=1}^n \ddot{p}_k(\boldsymbol{\theta}_0)p_k^{-1}(\boldsymbol{\theta}_0)$ is a martingale. As in item 1, the first condition of (6) is generally not cheched.

2.2. Strong consistency of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ based on the minimum contrast property.

Proposition 2. *Assume that* $S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) = D_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta})$, *where* $D_n(\boldsymbol{\theta})$ *is* $\mathcal{F}_{n-1}$*-measurable and* $M_n(\boldsymbol{\theta}) =: \sum_{k=1}^n X_k(\boldsymbol{\theta})$ *is a martingale. Then*

$$\text{(11)} \quad \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0$$

$$\text{on } \cap_\delta \Big\{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta}) = \infty,\ \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \sum_{n=1}^{\infty} E_{\boldsymbol{\theta}_0}\Big(X_n^2(\boldsymbol{\theta})|\mathcal{F}_{n-1}\Big)D_n^{-2}(\boldsymbol{\theta}) < \infty\Big\}.$$

P r o o f. From $S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) = D_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta})$, we deduce that

$$\text{(12)} \quad \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) \geq \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta})\Big(1 - \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} |M_n(\boldsymbol{\theta})D_n^{-1}(\boldsymbol{\theta})|\Big).$$

Then we use Wu's Lemma [11]: if $\underline{\lim}_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}(S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0)) \overset{a.s.}{>} 0$, for all $\delta > 0$, then $\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s.}{=} \boldsymbol{\theta}_0$.

So thanks to (12), it is sufficient to prove that

$$\underline{\lim}_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta})\Big(1 - \overline{\lim}_n \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} |M_n(\boldsymbol{\theta})D_n^{-1}(\boldsymbol{\theta})|\Big) \overset{a.s.}{>} 0.$$

For this purpose, we use the following USLLNSM (Uniform SLLNM):

Lemma 1. *Let, for all* $\boldsymbol{\theta} \in \tilde{\boldsymbol{\Theta}}$ *compact,* $M_n(\boldsymbol{\theta}) := \sum_{k=1}^n X_k(\boldsymbol{\theta})$ *be a martingale* ($E_{\boldsymbol{\theta}_0}\big(X_k(\boldsymbol{\theta})|\mathcal{F}_{k-1}\big) = 0$, *for all* k), *and let* $D_n(\boldsymbol{\theta}) := \sum_{k=1}^n d_k^2(\boldsymbol{\theta})$ *be a nonnegative quantity with* $d_k^2(\boldsymbol{\theta})$ $\mathcal{F}_{k-1}$*-measurable, for all* k. *Assume that* $E_{\boldsymbol{\theta}_0}\big(|X_k(\boldsymbol{\theta}_1) - X_k(\boldsymbol{\theta}_2)||\mathcal{F}_{k-1}\big) \le h_X(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|)u_k$, *where* u_k *is* $\mathcal{F}_{k-1}$*-measurable, and* $\lim_{x\to 0} h_X(x) = 0$, *and similarly for* $|d_k^2(\boldsymbol{\theta}_1) - d_k^2(\boldsymbol{\theta}_2)|$. *Then*

$$(13) \quad \lim_n \sup_{\boldsymbol{\theta}\in\tilde{\boldsymbol{\Theta}}} |M_n(\boldsymbol{\theta}) D_n^{-1}(\boldsymbol{\theta})| \overset{a.s.}{=} 0$$

$$\text{on } \{\lim_n \inf_{\boldsymbol{\theta}\in\tilde{\boldsymbol{\Theta}}} D_n(\boldsymbol{\theta}){=}\infty,\ \sup_{\boldsymbol{\theta}\in\tilde{\boldsymbol{\Theta}}} \sum_{n=1}^{\infty} E_{\boldsymbol{\theta}_0}\big(X_n^2(\boldsymbol{\theta})|\mathcal{F}_{n-1}\big) D_n^{-2}(\boldsymbol{\theta}){<}\infty\}.$$

This proposition is the direct generalization of Proposition 3.1 of [6] which concerns the particular class $M_n(\boldsymbol{\theta}) := \sum_{k=1}^n e_k(\boldsymbol{\theta}_0)d_k(\boldsymbol{\theta})$. It can be viewed as a generalization of the SLLNM (Strong Law of Large Numbers for Martingales) [4], and as a generalization of the uniform strong law of large numbers for i.i.d. variables, originally due to Le Cam. □

Corollary 1. *Assume that* $D_n(\boldsymbol{\theta}) - D_{n-1}(\boldsymbol{\theta}) \ge 0$, *and let* $D_n(\boldsymbol{\theta}) - D_{n-1}(\boldsymbol{\theta}) =: d_n^2(\boldsymbol{\theta})$. *Then*

$$(14) \quad \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\ge\delta} D_n(\boldsymbol{\theta}){=}\infty,$$
$$\overline{\lim}_n \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\ge\delta} E_{\boldsymbol{\theta}_0}\big(X_n^2(\boldsymbol{\theta})|\mathcal{F}_{n-1}\big) d_n^{-2}(\boldsymbol{\theta}) 1_{\{d_n(\boldsymbol{\theta})\neq 0\}}{<}\infty\}.$$

P r o o f. (14) implies (11) thanks to (8). □

Consider (11) and (14) applied to (1), (3), (4).

1. $\Psi_k(Z_k, \boldsymbol{\theta}) = (Z_k - g_k(\boldsymbol{\theta}))^2$. Writing $Z_k - g_k(\boldsymbol{\theta}) =: e_k(\boldsymbol{\theta})$ and $g_k(\boldsymbol{\theta}_0) - g_k(\boldsymbol{\theta}) =: d_k(\boldsymbol{\theta})$, we get that

$$\begin{aligned} S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) &= \sum_{k=1}^n (e_k^2(\boldsymbol{\theta}) - e_k^2(\boldsymbol{\theta}_0)) \\ &= \sum_{k=1}^n d_k^2(\boldsymbol{\theta}) + 2\sum_{k=1}^n e_k(\boldsymbol{\theta}_0) d_k(\boldsymbol{\theta}) =: D_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}). \end{aligned}$$

Consequently, defining $\sigma_n^2 := E_{\boldsymbol{\theta}_0}\big(e_n^2(\boldsymbol{\theta}_0)|\mathcal{F}_{n-1}\big)$, (11) is reduced to

$$(15)\quad \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta}){=}\infty, \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \sum_{n=1}^{\infty} \sigma_n^2 d_n^2(\theta)\big(\sum_{k=1}^{n} d_k^2(\boldsymbol{\theta})\big)^{-2}{<}\infty\},$$

and, when $\overline{\lim}_n \sigma_n^2 \overset{a.s.}{<} \infty$, (14) is reduced to

$$(16)\qquad \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta}){=}\infty\}.$$

Note that $\overline{\lim}_n \sigma_n^2 \overset{a.s.}{<} \infty$ under appropriate weights $\{\lambda_k\}_k$: for all k, $\lambda_k \propto \sigma_k^{-2}$ or $\lambda_k \propto \widehat{\sigma}_k^{-2}$, where $\widehat{\sigma}_k^2$ is a consistent estimator of σ_k^2.

Remark. Assume that $g_k(\boldsymbol{\theta})$ is differentiable with $\dot{g}_k(\boldsymbol{\theta})$ continuous in $\boldsymbol{\theta}$. Then we may write $g_k(\boldsymbol{\theta}) - g_k(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \dot{g}_k(\tilde{\boldsymbol{\theta}}_k)$, where $\tilde{\boldsymbol{\theta}}_k$ lies "between" $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. In this case,

$$(17)\qquad \cap_\delta\{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta}){=}\infty\} \overset{a.s.}{=} \{\lim_n \lambda_{\min}(\boldsymbol{\Phi}_n){=}\infty\},$$

where $\boldsymbol{\Phi}_n = \sum_{k=1}^n \dot{g}_k(\boldsymbol{\theta}_0)\dot{g}_k^T(\boldsymbol{\theta}_0)$. This implies that(16) is reduced to

$$(18)\qquad \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \{\lim_n \lambda_{\min}(\boldsymbol{\Phi}_n){=}\infty\}.$$

Comparing (18) to (6), we see that (6) contains unnecessary conditions, while (18) is reduced to a necessary and sufficient condition in the sense that it is easy to find examples with $\overline{\lim}_n \lambda_{\min}(\boldsymbol{\Phi}_n) \overset{a.s.}{<} \infty$ and such that the estimator is not consistent (take $p = 1$, $g_k(\theta) = \theta W_k$ and the $\{e_k(\theta_0)\}_k$ i.i.d.. Then $\widehat{\theta}_n - \theta_0 = \big(\sum_{k=1}^n W_k^2\big)^{-1}\big(\sum_{k=1}^n e_k(\theta_0)W_k\big)$ the variance of which is proportional to $\big(\sum_{k=1}^n W_k^2\big)^{-1} = \boldsymbol{\Phi}_n^{-1}$).

Consequence in the particular linear case

Assume that $g_k(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{W}_k$ with $\overline{\lim}_n \sigma_n^2 \overset{a.s.}{<} \infty$. Then (18) is reduced to

$$(19)\quad \lim_n \left(\sum_{k=1}^n \mathbf{W}_k \mathbf{W}_k^T\right)^{-1} \left(\sum_{k=1}^n e_k(\boldsymbol{\theta}_0)\mathbf{W}_k\right) \overset{a.s.}{=} 0 \quad \text{on } \left\{\lim_n \lambda_{\min}\left(\sum_{k=1}^n \mathbf{W}_k \mathbf{W}_k^T\right) {=}\infty\right\}$$

while, using directly the SLLNMM [8], we get that

(20)
$$\Big(\sum_{k=1}^{n}\mathbf{W}_k\mathbf{W}_k^T\Big)^{-1}\Big(\sum_{k=1}^{n}e_k(\boldsymbol{\theta}_0)\mathbf{W}_k\Big)\overset{a.s.}{=}0 \text{ on } \{\lim_n \lambda_{\min}\Big(\sum_{k=1}^{n}\mathbf{W}_k\mathbf{W}_k^T\Big)=\infty,$$
$$\lim_n\Big(\ln(\|\sum_{k=1}^{n}\mathbf{W}_k\mathbf{W}_k^T\|)\Big)^{\nu}\Big(\lambda_{\min}(\sum_{k=1}^{n}\mathbf{W}_k\mathbf{W}_k^T)\Big)^{-1}=0\},\ \text{for some } \nu>1.$$

The last condition in (20) is equivalent to

$$\lim_n\Big(\ln(\lambda_{\max}(\sum_{k=1}^{n}\mathbf{W}_k\mathbf{W}_k^T))\Big)^{\nu}\Big(\lambda_{\min}(\sum_{k=1}^{n}\mathbf{W}_k\mathbf{W}_k^T)\Big)^{-1}\overset{a.s.}{=}0,\ \text{for some } \nu>1,$$

which is automatically satisfied only when $p=1$. Note that (20) leads to the strong consistency of the estimator in the linear stochastic regression setting and was given in [7].

We then obtain with (19) an improvement of Lin's conditions (10) for the particular class of martingales $\mathbf{M}_n=\sum_{k=1}^{n}e_k(\boldsymbol{\theta}_0)\mathbf{W}_k$, where $\overline{\lim}_n\sigma_n^2\overset{a.s.}{<}\infty$, and $\mathbf{W}_k$ is $\mathcal{F}_{k-1}$-measurable.

2. $\Psi_k(Z_k,\boldsymbol{\theta})=|Z_k-g_k(\boldsymbol{\theta})|=:|e_k(\boldsymbol{\theta})|$. This implies that

$$S_n(\boldsymbol{\theta})-S_n(\boldsymbol{\theta}_0)=\sum_{k=1}^{n}\big(|e_k(\boldsymbol{\theta})|-|e_k(\boldsymbol{\theta}_0)|\big)=\sum_{k=1}^{n}\frac{e_k^2(\boldsymbol{\theta})-e_k^2(\boldsymbol{\theta}_0)}{|e_k(\boldsymbol{\theta})|+|e_k(\boldsymbol{\theta}_0)|}.$$

Therefore

$$S_n(\boldsymbol{\theta})-S_n(\boldsymbol{\theta}_0)\geq A_n^{-1}(\boldsymbol{\theta})\big(D_n(\boldsymbol{\theta})+M_n(\boldsymbol{\theta})\big),$$

where $A_n(\boldsymbol{\theta}):=\sup_{1\leq k\leq n}\big(|d_k(\boldsymbol{\theta})|+2|e_k(\boldsymbol{\theta}_0)|\big)$, $D_n(\boldsymbol{\theta}):=\sum_{k=1}^{n}(g_k(\boldsymbol{\theta}_0)-g_k(\boldsymbol{\theta}))^2$, $M_n(\boldsymbol{\theta}):=\sum_{k=1}^{n}e_k(\boldsymbol{\theta}_0)(g_k(\boldsymbol{\theta}_0)-g_k(\boldsymbol{\theta}))$, $e_k(\boldsymbol{\theta}_0):=Z_k-g_k(\boldsymbol{\theta}_0)$, $d_k(\boldsymbol{\theta}_0):=g_k(\boldsymbol{\theta}_0)-g_k(\boldsymbol{\theta})$. Then using the same steps as in Proposition 2 and defining $\sigma_n^2:=E_{\boldsymbol{\theta}_0}\big(e_k^2(\boldsymbol{\theta}_0)|\mathcal{F}_{n-1}\big)$, we obtain that

(21)
$$\lim_n\widehat{\boldsymbol{\theta}}_n\overset{a.s}{=}\boldsymbol{\theta}_0 \text{ on } \cap_\delta\{\underline{\lim}_n\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}\frac{D_n(\boldsymbol{\theta})}{A_n(\boldsymbol{\theta})}>0,$$
$$\lim_n\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}D_n(\boldsymbol{\theta})=\infty,\quad\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}\sum_n\frac{\sigma_n^2d_n^2(\theta)}{\big(\sum_{k=1}^{n}d_k^2(\boldsymbol{\theta})\big)^2}<\infty\},$$

or, if $\overline{\lim}_n \sigma_n^2 \overset{a.s.}{<} \infty$, according to (14),

$$\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\underline{\lim}_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \frac{D_n(\boldsymbol{\theta})}{A_n(\boldsymbol{\theta})}>0, \lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta})=\infty\}.$$

Note that the subset above is reduced to $\{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta})=\infty\}$ if each $\Psi_k(Z_k, \boldsymbol{\theta})$ is bounded that is, if $\Psi_k(Z_k, \boldsymbol{\theta}) = \tilde{\Psi}_k(z, \boldsymbol{\theta}) 1_{\{\mathbf{Z}_k \in \mathbf{I}\}}$, where $\mathbf{Z}_k = (Z_k, Z_{k-1}, \ldots, Z_{k-d})$, d being the memory of the process, and $\mathbf{I}$ being any $d+1$-dimensional vector of large finite intervals.

3. $\Psi_k(Z_k, \boldsymbol{\theta}) = -\ln p_k(\boldsymbol{\theta})$, where $p_k(\boldsymbol{\theta}) := p_{\boldsymbol{\theta}}(Z_k|\mathcal{F}_{k-1})$ (conditional likelihood of Z_k). Then $S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) = \sum_{k=1}^n \ln(p_k(\boldsymbol{\theta}_0) p_k^{-1}(\boldsymbol{\theta}))$, implying in turn that

$$\begin{aligned} & S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) \\ = \; & \sum_{k=1}^n E_{\boldsymbol{\theta}_0}\Big(\ln \frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\Big) + \sum_{k=1}^n \Big(\ln \frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})} - E_{\boldsymbol{\theta}_0}\Big(\ln \frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\Big)\Big) \\ =: \; & D_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}), \end{aligned}$$

where by construction $M_n(\boldsymbol{\theta})$ is a martingale, and $D_n(\boldsymbol{\theta}) \geq 0$ because

$$E_{\boldsymbol{\theta}_0}\Big(\ln \frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\Big) \geq 0 \text{ (with equality if and only if } p_k(\boldsymbol{\theta}) = p_k(\boldsymbol{\theta}_0)).$$

This quantity is the Kullback-Leibler divergence between the true density $p_k(\boldsymbol{\theta}_0)$ and $p_k(\boldsymbol{\theta})$. Then (14) becomes

$$\text{(22)} \quad \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta})=\infty,$$

$$\overline{\lim}_n \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \frac{Var_{\boldsymbol{\theta}_0}\Big(\ln \frac{p_{\boldsymbol{\theta}_0}(Z_n|\mathcal{F}_{n-1})}{p_{\boldsymbol{\theta}}(Z_n|\mathcal{F}_{n-1})}|\mathcal{F}_{n-1}\Big)}{E_{\boldsymbol{\theta}_0}\Big(\ln \frac{p_{\boldsymbol{\theta}_0}(Z_n|\mathcal{F}_{n-1})}{p_{\boldsymbol{\theta}}(Z_n|\mathcal{F}_{n-1})}|\mathcal{F}_{n-1}\Big)} 1_{\{E_{\boldsymbol{\theta}_0}\Big(\ln \frac{p_{\boldsymbol{\theta}_0}(Z_n|\mathcal{F}_{n-1})}{p_{\boldsymbol{\theta}}(Z_n|\mathcal{F}_{n-1})}\Big)\neq 1\}} < \infty\}.$$

3. Asymptotic distribution We use the classical approach based on the first order Taylor series of $\dot{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$ (see (5)), and we write

$$\boldsymbol{\Upsilon}_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\Big(\tilde{\boldsymbol{\Upsilon}}_n^{-1} \ddot{\mathbf{S}}_n(\boldsymbol{\theta}_n) \boldsymbol{\Upsilon}_n^{-1}\Big)^{-1} \Big(\tilde{\boldsymbol{\Upsilon}}_n^{-1} \dot{\mathbf{S}}_n(\boldsymbol{\theta}_0)\Big),$$

where $\boldsymbol{\Upsilon}_n$ and $\tilde{\boldsymbol{\Upsilon}}_n$ are $p \times p$ invertible matrices.

Proposition 3. *Assume that $S_n(\boldsymbol{\theta})$ is twice differentiable in $\boldsymbol{\theta}$ and that $\dot{\mathbf{S}}_n(\boldsymbol{\theta}_0)$ is a martingale. Let $\boldsymbol{\Upsilon}_n$ and $\tilde{\boldsymbol{\Upsilon}}_n$ be $p \times p$ invertible matrices such that $\tilde{\boldsymbol{\Upsilon}}_n$ is deterministic and assume that*

$$\lim_n \tilde{\boldsymbol{\Upsilon}}_n^{-1} \ddot{\mathbf{S}}_n(\boldsymbol{\theta}_n) \boldsymbol{\Upsilon}_n^{-1} \text{ exists in probability and is deterministic} \tag{23}$$

$$\lim_n \tilde{\boldsymbol{\Upsilon}}_n^{-1} \dot{\mathbf{S}}_n(\boldsymbol{\theta}_0) \text{ exists in distribution.} \tag{24}$$

Then $\lim_n \boldsymbol{\Upsilon}_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ exists in distribution.

The proof is directly deduced from Slutsky's convergence theorem.

Remark. According to Subsection 2.1, when $\Psi_k(Z_k, \boldsymbol{\theta}) = (Z_k - g_k(\boldsymbol{\theta}))^2$ or when $\Psi_k(Z_k, \boldsymbol{\theta}) = -\ln p_k(\boldsymbol{\theta})$, then $\dot{\mathbf{S}}_n(\boldsymbol{\theta}_0)$ is a martingale.

Remark. Condition (23) should be easily checked thanks to the strong consistency of $\widehat{\boldsymbol{\theta}}_n$ under continuity conditions in $\boldsymbol{\theta}$ on the $\{\ddot{\Psi}_k(z, \boldsymbol{\theta})\}_k$, and using in addition the USLLNM, when $\ddot{\mathbf{S}}_n(\boldsymbol{\theta})$ contains a term that is a martingale. Condition (24) should be easily checked thanks to an appropriate CLT (Central Limit Theorem). Denoting $\tilde{\boldsymbol{\Upsilon}}_n^{-1} \dot{\mathbf{S}}_n(\boldsymbol{\theta}_0) =: \sum_{k=1}^n \mathbf{X}_{k,n} =: \mathbf{M}_k^n$, and since $\{\mathbf{M}_k^n\}_{k \le n}$ is a martingale, then, if $\lim_n \sum_{k=1}^n E\Big(\mathbf{X}_{k,n} \mathbf{X}_{k,n}^T | \mathcal{F}_{k-1}^n\Big)$ exists in probability and is a deterministic matrix $\boldsymbol{\Gamma}$, we may use a CLT for martingales to get the limit distribution of $\mathbf{M}_n^n$. However, in the branching processes setting, $\boldsymbol{\Gamma}$ is the most often random, preventing us to use some CLT for martingales. Since in this setting, $\mathbf{M}_n^n$ may be usually written as a random sum of independent variables, then we may rather use a CLT for random sums (see for example *e.g.* [1], [2]). Recall first the following CLT for multivariate martingale arrays ([10] (continuous time), [2]) which is the direct generalization of the CLT for univariate martingale arrays [4]: Let $\mathbf{M}_k^n =: \sum_{l=1}^k \mathbf{X}_{l,n}$ be a multidimensional $\{\mathcal{F}_{k-1}^n\}_k$-martingale triangular array. Then $\lim_n \mathbf{M}_n^n \stackrel{D}{=} \mathcal{N}(0, \boldsymbol{\Gamma})$, under the following assumptions:

$$\lim_n \sum_{k=1}^n E\Big(\mathbf{X}_{k,n} \mathbf{X}_{k,n}^T | \mathcal{F}_{k-1}^n\Big) \stackrel{P}{=} \boldsymbol{\Gamma} \text{ (semi-definite deterministic matrix)},$$

$$\lim_n \sum_{k=1}^n E\Big(\|\mathbf{X}_{k,n}\|^2 1_{\{\|\mathbf{X}_{k,n}\|^2 \ge \epsilon\}} | \mathcal{F}_{k-1}^n\Big) \stackrel{P}{=} 0, \ \forall \epsilon > 0.$$

4. Example: a branching process with memory. Let the following BGW (Bienaymé-Galton-Watson) with memory d:

$$(25)\quad Z_k = \sum_{l=1}^{d} \sum_{i=1}^{Z_{k-l}} Y_{k-l,k,i},\ Y_{k-l,k,i}|\mathcal{F}_{k-1} \overset{D}{=} Poisson\big(m_l(\boldsymbol{\theta}_0)\big),\ m_l(\boldsymbol{\theta}_0) > 0, \forall l,$$

$$\{Y_{k-l,k,i}\}_{i,l}|\mathcal{F}_{k-1} \text{ independent},\ \mathcal{F}_{k-1} = \sigma\text{-}\big(\{Z_{k-l}\}_{l\geq 1}\big),$$

where $Y_{k-l,k,i}$ represents the offspring at time k generated by the individual i of the population at time $k-l$. The quantity l represents the maturation time to produce the offspring $Y_{k-l,k,i}$. We assume that $m_l(\boldsymbol{\theta}_0)$ is independent of the process $\{Z_k\}_k$ and depends on a unknown p-dimensional parameter $\boldsymbol{\theta}_0$, where $1 \leq p \leq d$. When $d = 1$, this process is a single-type BGW process.

As example, $Y_{k-l,k,i}$ may be the number of newborns at time k of i born at time $k-l$. Then, if 1 is the age of i at his birth time $k-l$, $l+1$ represents the age of i at the birth time k of $Y_{k-l,k,i}$. Therefore $\{Z_k\}_k$ is the process of newborns, and the older populations are then directly deduced from this newborns process using survival. In [9], Z_k represents the incidence of infectives at time k. An infective is assumed to be removed at the following time, and $Y_{k-l,k,i}$ is the number of secondary infectives generated at time k, by an infective of time $k-l$. Here l represents the latent (non infectious) period.

According to the assumptions,

$$Z_k|\mathcal{F}_{k-1} \overset{D}{=} Poisson\big(g_k(\boldsymbol{\theta}_0)\big),\ g_k(\boldsymbol{\theta}) = \sum_{l=1}^{d} Z_{k-l} m_l(\boldsymbol{\theta}).$$

This process may be represented as a multitype BGW process $\{\mathbf{Z}_k\}_k$, where $\mathbf{Z}_k := (Z_k, Z_{k-1}, \ldots, Z_{k-(d-1)})$. The behavior of $\{Z_k\}_k$ is then deduced from the behavior of $\{\mathbf{Z}_k\}_k$ [9]. We obtain for example that $\lim_n Z_n \rho^{-n} \overset{a.s.}{=} W$, where ρ is solution of $\sum_{l=1}^{d} \rho^{-l} m_l(\boldsymbol{\theta}_0) = 1$ and W is a nonnegative integrable random variable. Moreover the process becomes a.s. extinct if $\rho \leq 1$ which is equivalent to $\sum_{l=1}^{d} m_l(\boldsymbol{\theta}_0) \leq 1$, while its probability of nonextinction is $P(W > 0)$ which is strictly positive when $\rho > 1$, i.e. when $\sum_{l=1}^{d} m_l(\boldsymbol{\theta}_0) > 1$.

We consider here the estimators of the parameter $\boldsymbol{\theta}_0$, given by (1), (3), (4), and we give conditions leading to their strong consistency, and then to their asymptotic distribution using the results of Sections 2 and 3.

4.1. Strong consistency of $\widehat{\boldsymbol{\theta}}_n$.

- $\Psi_k(Z_k,\boldsymbol{\theta}) = -\ln p_k(\boldsymbol{\theta})$, where $p_k(\boldsymbol{\theta}) = \exp(-g_k(\boldsymbol{\theta}))(Z_k!)^{-1}\big(g_k(\boldsymbol{\theta})\big)^{Z_k}$. This implies that $\ln\big(p_k(\boldsymbol{\theta}_0)p_k^{-1}(\boldsymbol{\theta})\big) = g_k(\boldsymbol{\theta}) - g_k(\boldsymbol{\theta}_0) - Z_k\ln\big(g_k(\boldsymbol{\theta})g_k^{-1}(\boldsymbol{\theta}_0)\big)$. Defining $x_k(\boldsymbol{\theta}) := g_k(\boldsymbol{\theta})g_k^{-1}(\boldsymbol{\theta}_0)$, we notice that

$$(26)\quad x_k(\boldsymbol{\theta}) \in \big[x_{\min}, x_{\max}\big],\ x_{\min} := \min_{l,\boldsymbol{\theta}} m_l(\boldsymbol{\theta})\big(\max_l m_l(\boldsymbol{\theta}_0)\big)^{-1},$$
$$x_{\max} := \max_{l,\boldsymbol{\theta}} m_l(\boldsymbol{\theta})\big(\min_l m_l(\boldsymbol{\theta}_0)\big)^{-1}.$$

Moreover assume the following condition C:

$$(27)\quad \forall\delta>0,\, \forall\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta,\, \exists\xi_\delta : x_k(\boldsymbol{\theta}) \notin \big[1-\xi_\delta, 1+\xi_\delta\big].$$

This assumption is checked in particular when $p = 1$ and the $m_l(\theta)$ are, for all $l = 1,\dots,d$, strictly increasing (resp. decreasing) functions of θ. This is the case in [9], where $m_l(\theta)$ is affine in θ. We have

$$(28)\qquad \frac{Var_{\boldsymbol{\theta}_0}\big(\ln\frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\big)}{E_{\boldsymbol{\theta}_0}\big(\ln\frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\big)} = \frac{(\ln x_k(\boldsymbol{\theta}))^2}{x_k(\boldsymbol{\theta})-1-\ln x_k(\boldsymbol{\theta})}.$$

Thanks to (26) and assuming C, there exists $\beta_\delta < \infty$ such that

$$(29)\qquad (\ln x_k(\boldsymbol{\theta}))^2\big(x_k(\boldsymbol{\theta})-1-\ln x_k(\boldsymbol{\theta})\big)^{-1} \leq \beta_\delta.$$

Then, defining $D_n(\boldsymbol{\theta}) = \sum_{k=1}^n E_{\boldsymbol{\theta}_0}\big(\ln\frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\big)$, (22) becomes

$$\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta}){=}\infty\}.$$

Note that assuming C, thanks to (29), then, for $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta$,

$$\begin{aligned}\sum_{k=1}^n E_{\boldsymbol{\theta}_0}\big(\ln\frac{p_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta})}|\mathcal{F}_{k-1}\big) &= \sum_{k=1}^n (x_k(\boldsymbol{\theta})-1-\ln x_k(\boldsymbol{\theta}))g_k(\boldsymbol{\theta}_0)\\ &\geq \beta_\delta^{-1}\sum_{k=1}^n g_k(\boldsymbol{\theta}_0)\big(\ln x_k(\boldsymbol{\theta})\big)^2\\ &\geq \beta_\delta^{-1}\min_l m_l(\boldsymbol{\theta}_0)\big(\ln(1+\xi_\delta)\big)^2\sum_{k=1}^n\sum_{l=1}^d Z_{k-l}.\end{aligned}$$

Then, we obtain the following result.

Proposition 4. *Assume C. Then* $\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0$ *on the nonextinction set, that is on* $\{W > 0\}$.

- $\Psi_k(Z_k, \boldsymbol{\theta}) = \lambda_k (Z_k - g_k(\boldsymbol{\theta}))^2$, where $\lambda_k = (\sum_{l=1}^d Z_{k-l})^{-1}$. Thus $S_n(\boldsymbol{\theta})$ is the direct generalization of the contrast leading to the CLSE in the BGW setting ($d = 1$). Let $\sigma_k^2 := E_{\boldsymbol{\theta}_0}\big(\lambda_k (Z_k - g_k(\boldsymbol{\theta}_0))^2 | \mathcal{F}_{k-1}\big)$. Then $\sigma_k^2 = \big(\sum_{l=1}^d Z_{k-l}\big)^{-1} \sum_{l=1}^d Z_{k-l} m_l(\boldsymbol{\theta}_0) \in \big(\min_l m_l(\boldsymbol{\theta}_0), \max_l m_l(\boldsymbol{\theta}_0)\big)$. Consequently, according to (16),

$$\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0 \text{ on } \cap_\delta \{\lim_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} D_n(\boldsymbol{\theta}) = \infty\},$$

$$\text{where } D_n(\boldsymbol{\theta}) := \sum_{k=1}^n \big(g_k(\boldsymbol{\theta}_0) - g_k(\boldsymbol{\theta})\big)^2 \Big(\sum_{l=1}^d Z_{k-l}\Big)^{-1}.$$

Under C and (29), for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \delta$, we obtain that

$$\begin{aligned} D_n(\boldsymbol{\theta}) &\geq \sum_{k=1}^n g_k^2(\boldsymbol{\theta}_0) \big(\ln x_k(\boldsymbol{\theta}) + \beta_\delta^{-1} \big(\ln x_k(\boldsymbol{\theta})\big)^2\big)^2 \Big(\sum_{l=1}^d Z_{k-l}\Big)^{-1} \\ &\geq \min_l m_l^2(\boldsymbol{\theta}_0) \big(\ln(1 + \xi_\delta)\big)^2 \sum_{k=1}^n \sum_{l=1}^d Z_{k-l}, \end{aligned}$$

which implies the same result as for the MLE.

Proposition 5. *Assume C. Then* $\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0$ *on the nonextinction set, that is on* $\{W > 0\}$.

Remark. In the BGW setting ($d = 1$), the strong consistency of the CLSE $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0 := m_1(\boldsymbol{\theta}_0)$ is classically proved directly using the explicit expression of $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ and the asymptotic behavior of the martingale $W_n = Z_n m_0^{-n}$, contrary to our indirect proof based only on the nonextinction set:

$$\widehat{m}_n - m_0 = \frac{\sum_{k=1}^n (Z_k - m_0 Z_{k-1})}{\sum_{k=1}^n Z_{k-1}} = \frac{\sum_{k=1}^n (W_k - W_{k-1}) m_0^k}{\sum_{k=1}^n W_{k-1} m_0^{k-1}}.$$

- $\Psi_k(Z_k, \boldsymbol{\theta}) = \lambda_k^{1/2} 1_{\{Z_k \in I\}} |Z_k - g_k(\boldsymbol{\theta})|$, where $\lambda_k = (\sum_{l=1}^d Z_{k-l})^{-1}$ and I is any very large finite interval. Then according to (21) and to the results in the CLSE setting (previous item), we obtain the following result.

Proposition 6. *Assume C. Then* $\lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s}{=} \boldsymbol{\theta}_0$ *on the nonextinction set, that is on* $\{W > 0\}$.

4.2. Asymptotic distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$.

- $\Psi_k(Z_k,\boldsymbol{\theta}) = \big(Z_k - g_k(\boldsymbol{\theta})\big)^2\lambda_k$, where $g_k(\boldsymbol{\theta}) = \sum_{l=1}^d Z_{k-l}m_l(\boldsymbol{\theta})$, $\lambda_k = \sum_{l=1}^d Z_{k-l}$. Then

$$\text{(30)}\qquad \dot{S}_n(\boldsymbol{\theta}_0) = -2\sum_{k=1}^n e_k(\boldsymbol{\theta}_0)\dot{g}_k(\boldsymbol{\theta}_0)\lambda_k,$$

$$\text{(31)}\qquad \ddot{S}_n(\boldsymbol{\theta}) = 2\sum_{k=1}^n \dot{g}_k(\boldsymbol{\theta})\dot{g}_k^T(\boldsymbol{\theta})\lambda_k - 2\sum_{k=1}^n e_k(\boldsymbol{\theta})\ddot{g}_k(\boldsymbol{\theta})\ \lambda_k,$$

where $e_k(\boldsymbol{\theta}) := Z_k - g_k(\boldsymbol{\theta})$. The quantity $\dot{S}_n(\boldsymbol{\theta}_0)$ is a martingale and using $Z_k \sim W\rho^k$ and $Var_{\boldsymbol{\theta}_0}(e_k(\boldsymbol{\theta}_0)) = g_k(\boldsymbol{\theta}_0)$, we have, for any large n,

$$\begin{aligned}
&\sum_{k=1}^n E\Big(e_k^2(\boldsymbol{\theta}_0)\dot{g}_k(\boldsymbol{\theta}_0)\dot{g}_k^T(\boldsymbol{\theta}_0)\lambda_k^2|\mathcal{F}_{k-1}\Big)\\
= &\sum_{k=1}^n \big(\sum_{l=1}^d Z_{k-l}m_l(\boldsymbol{\theta}_0)\big)\big(\sum_{l=1}^d Z_{k-l}\dot{m}_l(\boldsymbol{\theta}_0)\big)\big(\sum_{l=1}^d Z_{k-l}\dot{m}_l^T(\boldsymbol{\theta}_0)\big)\big(\sum_{l=1}^d Z_{k-l}\big)^{-2}\\
\sim\ &W\sum_{k=1}^n \rho^k \frac{\sum_{l=1}^d \rho^{-l}m_l(\boldsymbol{\theta}_0)\sum_{l=1}^d \rho^{-l}\dot{m}_l(\boldsymbol{\theta}_0)\sum_{l=1}^d \rho^{-l}\dot{m}_l^T(\boldsymbol{\theta}_0)}{\big(\sum_{l=1}^d \rho^{-l}\big)^2}.
\end{aligned}$$

Choosing $\tilde{\boldsymbol{\Upsilon}}_n = (\sum_{k=1}^n \rho^k)^{1/2}\mathbf{I}$ and denoting $\tilde{\boldsymbol{\Upsilon}}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0) =: \sum_{k=1}^n \mathbf{X}_{k,n}$, then $\sum_{k=1}^n E\big(\mathbf{X}_{k,n}\mathbf{X}_{k,n}^T|\mathcal{F}_{k-1}\big)$ converges a.s. to a random variable, preventing us to use a CLT for martingales. So using (25), we rather write

$$\tilde{\boldsymbol{\Upsilon}}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0) \sim -\frac{2}{\big(\sum_{k=1}^n \rho^k\big)^{1/2}}\sum_{l=1}^d\sum_{j=1}^{S_{1-l,n-l}}\big(Y_{l,j} - m_l(\boldsymbol{\theta}_0)\big)\frac{\sum_{l=1}^d \rho^{-l}\dot{m}_l(\boldsymbol{\theta}_0)}{\sum_{l=1}^d \rho^{-l}},$$

where $S_{1-l,n-l} = \sum_{k=1}^n Z_{k-l}$ and $Y_{l,j} \overset{D}{=} Y_{k-l,k,i}$. Then we use a CLT for random sums for each l, and since $\lim_n S_{1-l,n-l}\big(\sum_{k=1}^n \rho^k\big)^{-1} \overset{a.s.}{=} \rho^{-l}W$ (Toeplitz Lemma), we get that, for $l = 1,\dots,d$,

$$\lim_n \tilde{\boldsymbol{\Upsilon}}_n^{-1}\dot{S}_n(\boldsymbol{\theta}_0) \overset{D}{=} -2W^{1/2}\mathbf{V}(\boldsymbol{\theta}_0)\sum_{l=1}^d U_l\rho^{-l/2} \text{ on } \{W>0\},$$

where $U_l \overset{D}{\sim} \mathcal{N}\big(0, m_l(\boldsymbol{\theta}_0)\big)$, $\mathbf{V}(\boldsymbol{\theta}_0) := \big(\sum_{l=1}^d \rho^{-l}\dot{m}_l(\boldsymbol{\theta}_0)\big)\big(\sum_{l=1}^d \rho^{-l}\big)^{-1}$, and the $\{U_l\}_l$ are independent and independent of W. Then using $\sum_{l=1}^d \rho^{-l}m_l(\boldsymbol{\theta}_0) =$

1, we get that

$$\lim_n \tilde{\boldsymbol{\Upsilon}}_n^{-1} \dot{S}_n(\boldsymbol{\theta}_0) \stackrel{D}{=} -2W^{1/2}\mathbf{V}(\boldsymbol{\theta}_0)U \text{ on } \{W>0\},\ U \stackrel{D}{\sim} \mathcal{N}\big(0,1\big).$$

Consider now $\tilde{\boldsymbol{\Upsilon}}_n^{-1}\ddot{S}_n(\boldsymbol{\theta}_n)\boldsymbol{\Upsilon}_n^{-1}$ with $\boldsymbol{\Upsilon}_n = \tilde{\boldsymbol{\Upsilon}}_n = \big(\sum_{k=1}^n \rho^k\big)^{1/2}\mathbf{I}$. Using again $Z_{k-l} \sim W\rho^{k-l}$, we have, for any large n,

$$\begin{aligned}
&\tilde{\boldsymbol{\Upsilon}}_n^{-1}\ddot{S}_n(\boldsymbol{\theta}_n)\boldsymbol{\Upsilon}_n^{-1}\\
\sim\ & \frac{2W}{\sum_{k=1}^n \rho^k}\sum_{k=1}^n \rho^k \frac{\sum_{l=1}^d \rho^{-l}\dot{m}_l(\boldsymbol{\theta}_n)\sum_{l=1}^d \rho^{-l}\dot{m}_l^T(\boldsymbol{\theta}_n)}{\sum_{l=1}^d \rho^{-l}} - 2\tilde{\mathbf{M}}_n(\boldsymbol{\theta}_n)\\
=:\ & 2W\mathbf{B}(\boldsymbol{\theta}_n) - 2\tilde{\mathbf{M}}_n(\boldsymbol{\theta}_n),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\mathbf{M}}_n(\boldsymbol{\theta}_n) &= \sum_{k=1}^n e_k(\boldsymbol{\theta}_0)\big(\sum_{l=1}^d \rho^{-l}\ddot{m}_l(\boldsymbol{\theta}_n)\big)\big(\sum_{l=1}^d \rho^{-l}\big)^{-1}\\
&+\sum_{k=1}^n d_k(\boldsymbol{\theta}_n)\big(\sum_{l=1}^d \rho^{-l}\ddot{m}_l(\boldsymbol{\theta}_n)\big)\big(\sum_{l=1}^d \rho^{-l}\big)^{-1}.
\end{aligned}$$

The first term of $\tilde{\mathbf{M}}_n(\boldsymbol{\theta}_n)$ converges a.s. to 0 thanks to the USLLNM, and the second term converges a.s. to 0 thanks to the continuity of $m_l(\boldsymbol{\theta})$ and of $\ddot{m}_l(\boldsymbol{\theta})$ and the strong consistency of $\widehat{\boldsymbol{\theta}}_n$. So finally the conditions of Proposition 3 are checked leading to the following result.

Proposition 7. *Assume C and define*

$$\mathbf{V}_*(\boldsymbol{\theta}_0) := \big(\sum_{l=1}^d \rho^{-l}\dot{m}_l(\boldsymbol{\theta}_0)\sum_{l=1}^d \rho^{-l}\dot{m}_l^T(\boldsymbol{\theta}_0)\big)^{-1}\big(\sum_{l=1}^d \rho^{-l}\dot{m}_l(\boldsymbol{\theta}_0)\big).$$

Then on $\{W>0\}$,

$$\lim_n \big(\sum_{k=1}^n \rho^k\big)^{1/2}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \stackrel{D}{=} W^{-1/2}\mathbf{V}_*(\boldsymbol{\theta}_0)U,\ U \stackrel{D}{\sim} \mathcal{N}(0,1), \tag{32}$$

where U and W are independent.

- $\Psi_k(Z_k, \boldsymbol{\theta} = -\ln p_k(\boldsymbol{\theta})$. Then, writing $Z_k - g_k(\boldsymbol{\theta}) =: e_k(\boldsymbol{\theta})$, we have

$$\begin{aligned}
\dot{S}_n(\boldsymbol{\theta}_0) &= -\sum_{k=1}^{n} \frac{\dot{p}_k(\boldsymbol{\theta}_0)}{p_k(\boldsymbol{\theta}_0)} = -\sum_{k=1}^{n} e_k(\boldsymbol{\theta}_0)\frac{\dot{g}_k(\boldsymbol{\theta}_0)}{g_k(\boldsymbol{\theta}_0)}, \\
\ddot{S}_n(\boldsymbol{\theta}) &= \sum_{k=1}^{n} \frac{\dot{p}_k(\boldsymbol{\theta})\dot{p}_k^T(\boldsymbol{\theta})}{p_k^2(\boldsymbol{\theta})} - \sum_{k=1}^{n} \frac{\ddot{p}_k(\boldsymbol{\theta})}{p_k(\boldsymbol{\theta})} \\
&= \sum_{k=1}^{n} \frac{\dot{g}_k(\boldsymbol{\theta})\dot{g}_k^T(\boldsymbol{\theta})}{g_k(\boldsymbol{\theta})}\big(1 + \frac{e_k(\boldsymbol{\theta})}{g_k(\boldsymbol{\theta})}\big) - \sum_{k=1}^{n} e_k(\boldsymbol{\theta})\frac{\ddot{g}_k(\boldsymbol{\theta})}{g_k(\boldsymbol{\theta})}.
\end{aligned}$$

We see that these expressions are similar to those corresponding to the CLSE, except that λ_k is here replaced by $g_k^{-1}(\boldsymbol{\theta})$, and that there is an additional term in $\ddot{S}_n(\boldsymbol{\theta})$ depending on the $\{e_k(\boldsymbol{\theta})g_k^{-1}(\boldsymbol{\theta})\}_k$. Consequently we finally get the same limit as for the CLSE.

Proposition 8. *Assume C. Then on* $\{W > 0\}$,

$$\lim_n \big(\sum_{k=1}^{n} \rho^k\big)^{1/2}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \overset{D}{=} W^{-1/2}\mathbf{V}_*(\boldsymbol{\theta}_0)U,\ U \overset{D}{\sim} \mathcal{N}(0,1), \tag{33}$$

where U *and* W *are independent, and* $\mathbf{V}_*(\boldsymbol{\theta}_0)$ *is defined in Proposition 7.*

5. Conclusion. We saw that the indirect way of proof does not require any explicit expression of $\widehat{\boldsymbol{\theta}}_n$. Moreover, thanks to the USLLNM and to Wu's lemma, the condition for the strong consistency of $\widehat{\boldsymbol{\theta}}_n$ is reduced to the strong identifiability of $\boldsymbol{\theta}$ in $D_n(\boldsymbol{\theta})$, when the $\{\Psi_k\}_k$ are not "explosive" (condition on the variance of the martingale). Note also that, thanks to (5), when $\ddot{S}_n(\boldsymbol{\theta}_n)$ is independent of $\boldsymbol{\theta}_n$, then $\widehat{\boldsymbol{\theta}}_n$ has an explicit expression given by (5). Finally the USLLNM with an appropriate CLT lead to general conditions for an asymptotic distribution that are easily checked.

REFERENCES

[1] J. R. Blum, D. L. Hanson, J. I. Rosenblatt. On the central limit theorem for the sum of a random number of independent random variables. *Z. Wahrsch. Verw. Gebiete*, **1** (1963), 389–393.

[2] D. Dacunha-Castelle, M. Duflo. Probabilités et statistiques. 2. Problèmes à temps mobile. Paris, Masson, 1993.

[3] V. P. Godambe, C. C. Heyde. Quasi-Likelihood and Optimal Estimation. *International Statistical Review*, **55** (1987), No 3, 231–244.

[4] P. Hall, C. C. Heyde. Martingale limit theory and its application. New-York, Academic Press, 1980.

[5] C. C. Heyde. Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation. New-York, Springer, 1997.

[6] C. Jacob. Conditional Least Squares Estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.*, **38** (2010), No 1, 566–597.

[7] T. L. Lai, C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** (1982), 154–166.

[8] Y.-X. Lin. On the strong law of large numbers of multivariate martingales with random norming. *Stoch. Proc. Appl.* **54** (1994), No 2, 355–360.

[9] S. Pénisson, C. Jacob. Stochastic methodology for the study of an epidemic decay phase, based on a branching model. Special Issue on Applications of Stochastic Processes in Biology and Medicine. *Int. J. Stoch. An.*, Vol. 2012 (2012), article ID 598701, 32 pp.

[10] R. Rebolledo. Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** (1980), No 3, 269–286.

[11] C. F. Wu. Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** (1981), 501–513.

Christine Jacob
MIA unit, UR341, INRA
78352 Jouy-en-Josas Cedex, France
e-mail: `christine.jacob@jouy.inra.fr`