

*On the Almost Sure Central Limit Theorem for
Vector Martingales : Convergence of Moments and
Statistical Applications*

Bernard Bercu — Peggy Cénac — Guy Fayolle

N° 6780

December 2008

Thème NUM

 *Rapport
de recherche*

On the Almost Sure Central Limit Theorem for Vector Martingales : Convergence of Moments and Statistical Applications

Bernard Bercu ^{*}, Peggy Cénac [†], Guy Fayolle [‡]

Thème NUM — Systèmes numériques

Équipe-Projet Imara

Rapport de recherche n° 6780 — December 2008 — 23 pages

Abstract: We investigate the almost sure asymptotic properties of vector martingale transforms. Assuming some appropriate regularity conditions both on the increasing process and on the moments of the martingale, we prove that normalized moments of any even order converge in the almost sure central limit theorem for martingales. A conjecture about almost sure upper bounds under wider hypotheses is formulated. The theoretical results are supported by examples borrowed from statistical applications, including linear autoregressive models and branching processes with immigration, for which new asymptotic properties are established on estimation and prediction errors.

Key-words: Almost sure central limit theorem, vector martingale, moment, stochastic regression.

^{*} Université Bordeaux 1, Institut de Mathématiques de Bordeaux UMR 5251, and INRIA Bordeaux Sud-Ouest, 351 cours de la Libération, 33405 Talence Cedex, France. Bernard.Bercu@math.u-bordeaux1.fr

[†] Université de Bourgogne, Institut de Mathématiques de Bourgogne, UMR 5584, 9 rue Alain Savary, BP 47870, 21078 Dijon Cedex, France. Peggy.Cenac@u-bourgogne.fr

[‡] INRIA CR Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France. Guy.Fayolle@inria.fr

Théorème de la limite centrale pour les martingales vectorielles : convergence des moments et applications statistiques

Résumé : On étudie dans ce rapport des propriétés de convergence presque sûre de transformées de martingales vectorielles. Sous certaines conditions d'existence de moments et de régularité du processus croissant, on montre en particulier la convergence des moments normalisés de tout ordre pair dans le théorème central limite presque sûr pour les martingales vectorielles. On formule une conjecture de borne presque sûre, sous des hypothèses moins restrictives et couvrant des familles plus vastes de processus. Enfin, on applique ces résultats des exemples issus d'applications statistiques, notamment les modèles autorégressifs linéaires et certains processus de branchement avec immigration, ce qui permet d'établir de nouvelles propriétés asymptotiques sur les erreurs d'estimation et de prédiction.

Mots-clés : Théorème central limite presque sûr, martingale vectorielle, moment, regression stochastique.

1 Introduction

Let (X_n) be a sequence of real independent identically distributed random variables with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = \sigma^2$ and denote

$$S_n = \sum_{k=1}^n X_k.$$

It follows from the ordinary central limit theorem (CLT) that

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

which implies, for any bounded continuous real function h

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[h \left(\frac{S_n}{\sqrt{n}} \right) \right] = \int_{\mathbb{R}} h(x) dG(x)$$

where G stands for the Gaussian measure $\mathcal{N}(0, \sigma^2)$. Moreover, by the celebrated almost sure central limit theorem (ASCLT), the empirical measure

$$G_n = \frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{\frac{S_k}{\sqrt{k}}}$$

satisfies

$$G_n \implies G \quad \text{a.s.}$$

In other words, for any bounded continuous real function h

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} h \left(\frac{S_k}{\sqrt{k}} \right) = \int_{\mathbb{R}} h(x) dG(x) \quad \text{a.s.}$$

The ASCLT was simultaneously proved by Brosamler [3] and Schatte [15] and, in its present form, by Lacey and Phillip [10]. In contrast with the wide literature on the ASCLT for independent random variables, very few references are available on the ASCLT for martingales apart the recent work of Bercu and Fort [1, 2] and the important contribution of Chaabane and Maaouia [5, 4] and Lifshits [13, 14]. The ASCLT for martingales is as follows. Let (ε_n) be a martingale difference sequence adapted to a filtration $\mathbb{F} = (\mathcal{F}_n)$ with $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s. Let (Φ_n) be a sequence of random variables adapted to \mathbb{F} and denote by (M_n) the real martingale transform

$$M_n = \sum_{k=1}^n \Phi_{k-1} \varepsilon_k.$$

We also need to introduce the explosion coefficient f_n associated with (Φ_n)

$$f_n = \frac{\Phi_n^2}{s_n} \quad \text{where} \quad s_n = \sum_{k=0}^n \Phi_k^2.$$

As soon as (f_n) goes to zero a.s. and under reasonable assumption on the conditional moments of (ε_n) , the ASCLT for martingales asserts that the empirical measure

$$G_n = \frac{1}{\log s_n} \sum_{k=1}^n f_k \delta_{\frac{M_k}{\sqrt{s_{k-1}}}} \implies G \quad \text{a.s.} \quad (1.1)$$

In other words, for any bounded continuous real function h

$$\lim_{n \rightarrow \infty} \frac{1}{\log s_n} \sum_{k=1}^n f_k h\left(\frac{M_k}{\sqrt{s_{k-1}}}\right) = \int_{\mathbb{R}} h(x) dG(x) \quad \text{a.s.} \quad (1.2)$$

It is quite natural to overcome the case of unbounded functions h . To be more precisely, one might wonder if convergence (1.2) remains true for unbounded functions h . It has been recently shown in [1, 2] that, whenever (ε_n) has a finite conditional moment of order $> 2p$, then the convergence (1.2) still holds for any continuous real function h such that $|h(x)| \leq x^{2p}$.

Theorem 1.1 (Convergence of moments in the scalar case). *Assume that (ε_n) is a martingale difference sequence such that $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s. and satisfying for some integer $p \geq 1$ and for some real number $a > 2p$,*

$$\sup_{n \geq 0} \mathbb{E}[|\varepsilon_{n+1}|^a | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

If the explosion coefficient (f_n) tends to zero a.s., then

$$\lim_{n \rightarrow \infty} \frac{1}{\log s_n} \sum_{k=1}^n f_k \left(\frac{M_k^2}{s_{k-1}}\right)^p = \frac{\sigma^{2p} (2p)!}{2^p p!} \quad \text{a.s.} \quad (1.3)$$

The limit (1.3) is exactly the moment of order $2p$ of the Gaussian distribution $\mathcal{N}(0, \sigma^2)$. The purpose of the present paper is to extend the results of [1] to vector martingale transforms, which is strongly needed in various applications arising in statistics and signal processing.

Let (M_n) be a square integrable vector martingale with values in \mathbb{R}^d , adapted to a filtration \mathbb{F} . Its increasing process consists of the sequence $(\langle M \rangle_n)$ of symmetric, positive semi-definite square matrices of order d given by

$$\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}[(M_k - M_{k-1})(M_k - M_{k-1})^t | \mathcal{F}_{k-1}].$$

A first version of the ASCLT for discrete vector martingales was proposed in [5, 6], under fairly restrictive assumptions on the increasing process $(\langle M \rangle_n)$. Hereafter, our goal is to establish the convergence of moments of even order in the ASCLT for (M_n) under suitable assumptions on the behaviour of $(\langle M \rangle_n)$. We shall work in the general setting of vector martingales transforms (M_n) , which can be written as

$$M_n = M_0 + \sum_{k=1}^n \Phi_{k-1} \varepsilon_k$$

where M_0 can be taken arbitrary and (Φ_n) denotes a sequence of random vectors of dimension d , adapted to \mathbb{F} . We also introduce

$$S_n = \sum_{k=0}^n \Phi_k \Phi_k^t + S \quad (1.4)$$

where S is a fixed deterministic matrix, symmetric and positive definite. One can obviously see that if $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s., then the increasing process of (M_n) takes the form $\langle M \rangle_n = \sigma^2 S_{n-1}$. The explosion coefficient associated with (Φ_n) is now given by

$$f_n = \Phi_n^t S_n^{-1} \Phi_n = \frac{d_n - d_{n-1}}{d_n} \quad (1.5)$$

where $d_n = \det(S_n)$.

After this short survey, the paper will be organized as follows. The main theoretical result for vector martingale transforms is given in Section 2, at the end of which a quite plausible interesting conjecture is formulated, involving minimal assumptions. The final Section 3 proposes some statistical applications to estimation and prediction errors in linear autoregressive models and in branching processes with immigration.

2 On the convergence of moments

As mentioned above, our main result is given in theorem 2.1 and extends theorem 1.1 to the vector case. By the way, in mathematics, the difficulty of the problem is almost always a strictly increasing function of the dimension of some underlying space: it is also the case here !

Theorem 2.1. *Let (ε_n) be a martingale difference sequence satisfying the homogeneity condition $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s. and such that, for some integer $p \geq 1$ and some real number $a > 2p$,*

$$(H_p) \quad \sup_{n \geq 0} \mathbb{E}[|\varepsilon_{n+1}|^a | \mathcal{F}_n] < \infty \quad a.s.$$

In addition, assume that the explosion coefficient f_n tends to zero a.s. and that there exists a positive random sequence (α_n) increasing to infinity and an invertible symmetric matrix L , such that

$$\lim_{n \rightarrow \infty} \frac{1}{\alpha_n} S_n = L \quad \text{a.s.} \quad (2.1)$$

Then, the following limits hold almost surely

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k (M_k^t S_{k-1}^{-1} M_k)^p = \ell(p) = d \sigma^{2p} \prod_{j=1}^{p-1} (d + 2j), \quad (2.2)$$

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (M_k^t S_{k-1}^{-1} M_k)^p - (M_k^t S_k^{-1} M_k)^p = \lambda(p) = \frac{p}{d} \ell(p). \quad (2.3)$$

Remark 1. The limit $\ell(p)$ corresponds exactly to the moment of order $2p$ of the norm of a gaussian vector $\mathcal{N}(0, \sigma^2 I_d)$, so that theorem 2.1 shows indeed the convergence of moments of order $2p$ in the ASCLT for vector martingales. Here the deterministic normalization given in [5] has been replaced by the natural random normalization given by the increasing process.

Remark 2. The convergence hypothesis (2.1) clearly implies $f_n \rightarrow 0$ a.s., if and only if $\alpha_n \sim \alpha_{n-1}$ a.s. As a matter of fact, we deduce from (2.1) that

$$\lim_{n \rightarrow \infty} \frac{d_n}{\alpha_n^d} = \det L > 0 \quad \text{a.s.}$$

Proof. For the sake of shortness, we shall define the following variables.

$$\begin{aligned} V_n &= M_n^t S_{n-1}^{-1} M_n, \\ \varphi_n &= \alpha_n^{-1} \Phi_n^t L^{-1} \Phi_n, \\ v_n &= \alpha_{n-1}^{-1} M_n^t L^{-1} M_n. \end{aligned} \quad (2.4)$$

First of all, by using the symmetry of L , the convergence (2.1) ensures that

$$f_n = \varphi_n + o(\varphi_n) \quad \text{a.s.} \quad (2.5)$$

$$V_n = v_n + o(v_n) \quad \text{a.s.} \quad (2.6)$$

For we have

$$f_n = \varphi_n + \alpha_n^{-1} \Phi_n^t L^{-1/2} (\alpha_n L^{1/2} S_n^{-1} L^{1/2} - I) L^{-1/2} \Phi_n.$$

The matrix $R_n = \alpha_n L^{1/2} S_n^{-1} L^{1/2} - I$ is symmetric and denoting by ρ_n its spectral radius, we can write

$$\left| \alpha_n^{-1} \Phi_n^t L^{-1/2} R_n L^{-1/2} \Phi_n \right| \leq \rho_n \varphi_n,$$

and ρ_n converges to 0 almost surely, which leads to (2.5). The equation (2.6) is proved in the same way from the decomposition

$$V_n = v_n + M_n^t (S_{n-1}^{-1} - \alpha_{n-1}^{-1} L^{-1}) M_n.$$

Hence, by (2.6), $V_n^p = v_n^p + o(v_n^p)$ a.s. In order to find the limit (2.2), it suffices, by Toeplitz lemma, to study the convergence

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k V_k^p = \lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n \varphi_k v_k^p. \quad (2.7)$$

Theorem 2.1 will be proved by induction with respect to $p \geq 1$, as in [1] in the scalar case. The first step consists in writing a recursive relation for $M_n^t L^{-1} M_n$.

Let

$$\begin{aligned} \beta_n &= \text{tr}(L^{-1/2} S_n L^{-1/2}), & \gamma_n &= \frac{\beta_n - \beta_{n-1}}{\beta_n}, \\ \delta_n &= \frac{M_n^t L^{-1} \Phi_n}{\beta_n}, & m_n &= \beta_{n-1}^{-1} M_n^t L^{-1} M_n. \end{aligned} \quad (2.8)$$

According to the definition of (M_n) , the following decomposition holds

$$M_{n+1}^t L^{-1} M_{n+1} = M_n^t L^{-1} M_n + 2\varepsilon_{n+1} \Phi_n^t L^{-1} M_n + \varepsilon_{n+1}^2 \Phi_n^t L^{-1} \Phi_n,$$

so that

$$m_{n+1} = (1 - \gamma_n) m_n + 2\delta_n \varepsilon_{n+1} + \gamma_n \varepsilon_{n+1}^2. \quad (2.9)$$

The theorem relies essentially on the following lemma.

Lemma 2.2. *Under the assumptions of theorem 2.1, we have*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n \gamma_k m_k^p = \frac{\ell(p)}{d^{p+1}} \quad a.s. \quad (2.10)$$

In addition, if $g_n = M_n^t S_{n-1}^{-1} \Phi_n$, we also have

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (1 - f_k) g_k^2 m_k^{p-1} = \frac{\lambda(p)}{pd^{p-1}} \quad a.s. \quad (2.11)$$

Proof. Raising equality (2.9) to the power p implies

$$m_{n+1}^p = \sum_{k=0}^p \sum_{\ell=0}^k 2^{k-\ell} C_p^k C_k^\ell \gamma_n^\ell \delta_n^{k-\ell} ((1 - \gamma_n) m_n)^{p-k} \varepsilon_{n+1}^{k+\ell}. \quad (2.12)$$

After some straightforward simplifications, we obtain the relation

$$m_{n+1}^p + \mathcal{A}_n(p) = m_1^p + \mathcal{B}_{n+1}(p) + \mathcal{W}_{n+1}(p), \quad (2.13)$$

where

$$\mathcal{A}_n(p) = \sum_{k=1}^n \beta_k^{-p} (\beta_k^p - \beta_{k-1}^p) m_k^p, \quad \mathcal{B}_{n+1}(p) = \sum_{\ell=1}^{2p-1} \sum_{k=1}^n b_k(\ell) \varepsilon_{k+1}^\ell,$$

$$\mathcal{W}_{n+1}(p) = \sum_{k=1}^n \gamma_k^p \varepsilon_{k+1}^{2p}.$$

For $1 \leq \ell \leq p-1$, we have

$$b_k(\ell) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} 2^{\ell-2j} C_p^{\ell-j} C_{\ell-j}^j \gamma_k^j \delta_k^{\ell-2j} ((1-\gamma_k)m_k)^{p-\ell+j},$$

while, for $p \leq \ell \leq 2p-1$,

$$b_k(\ell) = \sum_{j=\ell-(p-1)}^{\lfloor \ell/2 \rfloor} 2^{\ell-2j} C_p^{\ell-j} C_{\ell-j}^j \gamma_k^j \delta_k^{\ell-2j} ((1-\gamma_k)m_k)^{p-\ell+j} \\ + C_p^{\ell-p} 2^{2p-\ell} - \delta_k^{2p-\ell} \gamma_k^{\ell-p}.$$

In order to take out useful information about $\mathcal{A}_n(p)$, it is necessary to study the asymptotic behaviour $\mathcal{W}_{n+1}(p)$, $\mathcal{B}_{n+1}(p)$ and m_n^p .

The case $p = 1$. Remarking that $\log \beta_n \sim \sum_{k=1}^n \gamma_k$, Chow's lemma [7, page 22] implies

$$\lim_{n \rightarrow \infty} \frac{1}{\log \beta_n} \mathcal{W}_{n+1}(1) = \sigma^2 \quad \text{a.s.}$$

Applying the strong law of large numbers for martingales and the inequality $\delta_n^2 \leq \gamma_n m_n$, we get $\mathcal{B}_{n+1}(1) = o(\mathcal{A}_n(1))$ a.s. In addition, from relation (2.30) in [17], it follows that $m_{n+1} = o(\log \beta_n)$ a.s. Consequently, by (2.13),

$$\lim_{n \rightarrow \infty} \frac{1}{\log \beta_n} \mathcal{A}_n(1) = \sigma^2 \quad \text{a.s.}$$

But the basic convergence assumption (2.1) implies immediately

$$\lim_{n \rightarrow \infty} \frac{\beta_n}{\gamma_n} = d \quad \text{a.s.},$$

so that $\log d_n \sim d \log \beta_n$ a.s. Hence

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \mathcal{A}_n(1) = \frac{\sigma^2}{d} \quad \text{a.s.},$$

which establishes (2.10).

As for the proof of (2.11), one can proceed in the same way, starting from the decomposition

$$\begin{aligned} V_{n+1} &= M_n^t S_n^{-1} M_n + 2\varepsilon_{n+1} \Phi_n^t S_n^{-1} M_n + \varepsilon_{n+1}^2 f_n \\ &= h_n + 2\varepsilon_{n+1} g_n + \varepsilon_{n+1}^2 f_n. \end{aligned}$$

which, for all $n \geq 1$, leads to

$$V_{n+1} + A_n = V_1 + B_{n+1} + W_{n+1}, \quad (2.14)$$

where

$$A_n \stackrel{\text{def}}{=} \sum_{k=1}^n a_k(1), \quad B_{n+1} \stackrel{\text{def}}{=} 2 \sum_{k=1}^n \varepsilon_{k+1} g_k, \quad W_{n+1} \stackrel{\text{def}}{=} \sum_{k=1}^n \varepsilon_{k+1}^2 f_k.$$

By Riccati's formula,

$$g_n^2 = (1 - f_n) a_n(1), \quad (2.15)$$

and, hence, coupling (2.15) with the strong law of large numbers for martingales,

$$B_{n+1} = o(A_n) \quad \text{a.s.}$$

On the other hand, $m_n = o(\log \beta_n)$, which, with (2.6), gives the almost sure estimate $V_n = o(\log d_n)$.

Now, since $\sum_{k=1}^n f_k \sim \log d_n$, Chow's lemma implies

$$\lim_{n \rightarrow \infty} (\log d_n)^{-1} W_{n+1} = \sigma^2 \quad \text{a.s.},$$

and to conclude the proof of Lemma 2.2 for $p = 1$, it suffices to divide (2.14) by $\log d_n$, letting $n \rightarrow \infty$.

The case $p \geq 2$. First, using again Chow's lemma, we can write

$$W_{n+1}(p) = o(\log d_n) \quad \text{a.s.}$$

Also, we shall show that

$$B_{n+1}(p) = \frac{p}{dp+1} \ell(p) \log d_n + o(\log d_n) + o(\mathcal{A}_n(p)) \quad \text{a.s.} \quad (2.16)$$

Setting, for $1 \leq \ell \leq 2p - 1$,

$$\varepsilon_{n+1}^\ell \stackrel{\text{def}}{=} e_{n+1}(\ell) + \mathbb{E}[\varepsilon_{n+1}^\ell | \mathcal{F}_n] = e_{n+1}(\ell) + \sigma_n(\ell), \quad (2.17)$$

we write $\sum_{k=1}^n b_k(\ell) \varepsilon_{k+1}^\ell = \mathcal{C}_{n+1}(\ell) + \mathcal{D}_n(\ell)$, with

$$\mathcal{C}_{n+1}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^n b_k(\ell) e_{k+1}(\ell), \quad \text{and} \quad \mathcal{D}_n(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^n b_k(\ell) \sigma_k(\ell).$$

First, for any ℓ such that $1 \leq \ell \leq p-1$, using again the strong law of large numbers and equation (2.30) of [17], we have

$$\mathcal{C}_{n+1}(\ell) = o(\log d_n) \quad \text{a.s.}$$

Suppose $3 \leq \ell \leq p-1$. From Hlder's inequality and the assumptions on the moments of ε_n , it follows that, for all $1 \leq j \leq 2p-1$, $|\sigma_n(j)|$ is bounded, which implies

$$|\mathcal{D}_n(\ell)| = \mathcal{O}\left(\sum_{k=1}^n \gamma_k^{\ell/2} m_k^{p-\ell/2}\right).$$

For even ℓ , the induction assumption leads to $\mathcal{D}_n(\ell) = o(\log d_n)$ a.s. When ℓ is odd, Cauchy Schwarz inequality and the induction assumption yield

$$|\mathcal{D}_n(\ell)| = \mathcal{O}\left(\left(\sum_{k=1}^n \gamma_k m_k^{p-1}\right)^{1/2} \left(\sum_{k=1}^n \gamma_k^{\ell} m_k^{p-\ell}\right)^{1/2}\right) = o(\log d_n) \quad \text{a.s.}$$

Suppose now $p \leq \ell \leq 2p-1$. It is easy to obtain

$$|\mathcal{D}_n(\ell)| = \mathcal{O}\left(\sum_{k=1}^n \gamma_k^{\ell/2} m_k^{p-\ell/2}\right) \quad \text{a.s.}$$

Now, from the induction assumption, we get, for any integer $\ell \neq 2$,

$$\mathcal{D}_n(\ell) = o(\log d_n) \quad \text{a.s.}$$

It remains to study $\mathcal{C}_{n+1}(\ell)$. By Chow's lemma, we have the almost sure equality

$$\mathcal{C}_{n+1}(\ell) = o(\nu_n(\ell)), \quad \text{with} \quad \nu_n(\ell) = \sum_{k=1}^n |b_k(\ell)|^{2p/\ell} = \mathcal{O}\left(\sum_{k=1}^n \gamma_k^p m_k^{(2p-\ell)p/\ell}\right).$$

For $\ell > p$, we apply Hlder's inequality with exponents ℓ/p and $\ell/(\ell-p)$. Then, $\nu_n(\ell) = o(\log d_n)$ a.s. In the particular case $p = \ell$, we get by the strong law of large numbers

$$|\mathcal{C}_{n+1}(\ell)|^2 = \mathcal{O}(\tau_n(p) \log \tau_n(p)) \quad \text{with} \quad \tau_n(p) = \sum_{k=1}^n b_k(p)^2 = \mathcal{O}\left(\sum_{k=1}^n \gamma_k^p m_k^p\right),$$

which leads to $|\mathcal{C}_{n+1}(\ell)| = o(\log d_n)$ a.s., since, from equation (2.30) of [17], $m_k^p = o((\log d_n)^\delta)$, for $0 < \delta < 1$. As for the last term $\mathcal{D}_n(2)$, one needs to study, for $p \geq 3$, the quantity

$$\begin{aligned} \frac{\mathcal{D}_n(2)}{\log d_n} &= \frac{\sigma^2}{\log d_n} \sum_{k=1}^n \sum_{j=0}^1 2^{2-2j} C_p^{2-j} C_{2-j}^j \gamma_k^j \delta_k^{2-2j} m_k^{p-2+j}, \\ &= \frac{2p(p-1)\sigma^2}{\log d_n} \sum_{k=1}^n \delta_k^2 m_k^{p-2} + \frac{p\sigma^2}{\log d_n} \sum_{k=1}^n \gamma_k m_k^{p-1}. \end{aligned}$$

It is easy to establish

$$g_n^2 = (\Phi_n^t S_{n-1}^{-1} M_n)^2 = d^2 \delta_n^2 + o(\gamma_n m_n). \quad (2.18)$$

Then, the induction assumption and Toeplitz lemma imply

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n \delta_k^2 m_k^{p-2} = \lim_{n \rightarrow \infty} \frac{1}{d^2 \log d_n} \sum_{k=1}^n a_k(1) m_k^{p-2} = \frac{\lambda(p-1)}{(p-1)d^p} \quad \text{a.s.}$$

Thus, we obtain

$$\lim_{n \rightarrow \infty} \frac{\mathcal{D}_n(2)}{\log d_n} = \ell(p-1) \left(\frac{2p(p-1)\sigma^2}{d^{p+1}} + \frac{p\sigma^2}{d^p} \right) = \frac{p}{d^{p+1}} \ell(p) \quad \text{a.s.},$$

[this formula being also valid for $p = 2$] which proves (2.16).

On the other hand, still applying equation (2.30) of [17], we derive the estimate $m_n^p = o(\log d_n)$. Thus,

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \mathcal{A}_n(p) = \frac{p}{d^{p+1}} \ell(p) \quad \text{a.s.}$$

Since

$$\frac{\beta_n^p - \beta_{n-1}^p}{\beta_n^p} = \gamma_n \sum_{q=0}^{p-1} \left(\frac{\beta_{n-1}}{\beta_n} \right)^{p-1-q} \sim p\gamma_n \quad \text{a.s.}, \quad (2.19)$$

we get finally

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n \gamma_k m_k^p = \lim_{n \rightarrow \infty} \frac{1}{p \log d_n} \mathcal{A}_n(p) = \frac{\ell(p)}{d^{p+1}},$$

and the proof of (2.10) is terminated.

As for the second part of Lemma 2.2, i.e. equation (2.11), we could proceed along similar lines, via the equality

$$V_{n+1}^p = \sum_{k=0}^p \sum_{\ell=0}^k 2^{k-\ell} C_p^k C_k^\ell f_n^\ell g_n^{k-\ell} h_n^{p-k} \varepsilon_{n+1}^{k+\ell},$$

with $g_n = \Phi_n^t S_{n-1}^{-1} M_n$ and $h_n = M_n^t S_n^{-1} M_n$. ■

The proof of theorem 2.1 is completed as relations (2.10) and (2.11) are clearly direct consequences of (2.2) and (2.3), respectively. Indeed, since $\beta_n \sim d\alpha_n$, we have almost surely

$$V_n \sim dm_n, \quad \text{and} \quad f_n \sim d\gamma_n,$$

hence convergence (2.10) immediately yields (2.2). Moreover in the particular case $p = 1$, the second convergence (2.11) is exactly (2.3). Now, for $p \geq 2$, the elementary expansion

$$x^p - y^p = (x - y)x^{p-1} \sum_{q=0}^{p-1} \left(\frac{y}{x}\right)^{p-1-q} \quad (2.20)$$

leads to

$$\begin{aligned} a_n(p) &= (M_n^t S_{n-1}^{-1} M_n)^p - (M_n^t S_n^{-1} M_n)^p \\ &= a_n(1) V_n^{p-1} \sum_{q=0}^{p-1} \left(\frac{V_n - a_n(1)}{V_n}\right)^{p-1-q}. \end{aligned} \quad (2.21)$$

Riccati's formula yields

$$\begin{aligned} a_n(1) &= (1 - f_n) M_n^t S_{n-1}^{-1} \Phi_n \Phi_n^t S_{n-1}^{-1} M_n \\ &\leq (1 - f_n) \operatorname{tr}(S_{n-1}^{-1/2} \Phi_n \Phi_n^t S_{n-1}^{-1/2}) V_n \\ &\leq f_n V_n, \end{aligned}$$

hence $a_n(1) = o(V_n)$ and, by (2.21),

$$a_n(p) \sim p a_n(1) V_n^{p-1} \quad \text{a.s.},$$

so that finally

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n a_k(p) = \lim_{n \rightarrow \infty} \frac{p d^{p-1}}{\log d_n} \sum_{k=1}^n a_k(1) m_k^{p-1} = \lambda(p) \quad \text{a.s.} \quad (2.22)$$

■

In most of statistical applications encountered so far, the convergence assumption (2.1) is satisfied. However, this technical hypothesis somehow circumvents the vector problem, which in its full generality is not yet solved. Indeed, (2.1) entails that all eigenvalues of the matrix S_n grow to infinity at the same speed α_n . Thus, our method of proof as some features in common with the scalar case. Hopefully, we should be able to establish the following result, stated for the moment as a conjecture, without assuming (2.1).

Conjecture 2.3. *Let (ε_n) be a martingale difference sequence satisfying the homogeneity condition $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s. and assumption (H_p) introduced in theorem 2.1, for some integer $p \geq 1$. Then, we have*

$$\begin{aligned} \sum_{k=1}^n f_k (M_k^t S_{k-1}^{-1} M_k)^p &= \mathcal{O}(\log d_n) \quad \text{a.s.} \\ \sum_{k=1}^n (M_k^t S_{k-1}^{-1} M_k)^p - (M_k^t S_k^{-1} M_k)^p &= \mathcal{O}(\log d_n) \quad \text{a.s.} \end{aligned}$$

3 Applications

3.1 Linear regression models

Theorem 2.1 is the keystone to understand the asymptotic behavior of cumulative prediction and estimation errors associated to the stochastic regression process given by, for all $n \geq 1$,

$$X_{n+1} = \theta^t \Phi_n + \varepsilon_{n+1}, \quad (3.1)$$

where $\theta \in \mathbb{R}^d$ is the unknown parameter. The random variables $X_n, \Phi_n, \varepsilon_n$ are the scalar observation, the regression vector and the scalar driven noise, respectively. We propose here two applications. The first one concerns stable autoregressive processes while the second one deals with branching processes with immigration.

For a reasonable sequence $(\hat{\theta}_n)$ of estimators of θ , we shall investigate the asymptotic performance of $\hat{\theta}_n^t \Phi_n$, as a predictor of X_{n+1} . More precisely, we shall focus on the prediction error $X_{n+1} - \hat{\theta}_n^t \Phi_n$ and on the estimation error $\hat{\theta}_n - \theta$. In fact, it is more relevant and efficient [9] to consider the cumulative prediction and estimation errors defined, for some $p \geq 1$, as

$$C_n(p) = \sum_{k=0}^{n-1} (X_{k+1} - \hat{\theta}_k^t \Phi_k)^{2p} \quad (3.2)$$

and

$$G_n(p) = \sum_{k=1}^n k^{p-1} \|\hat{\theta}_k - \theta\|^{2p}. \quad (3.3)$$

In the scalar case $d = 1$, under suitable moment conditions, asymptotic estimates on $(C_n(p))$ and $(G_n(p))$ were established in [1] by means of the standard least squares (LS) estimator

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^n \Phi_{k-1} X_k. \quad (3.4)$$

It turns out that theorem 2.1 allows us to improve the results of [1, 2]. Up to our knowledge, only partial results in the particular case $p = 1$ have been obtained, namely in [7, 17], where the authors derived the asymptotics of $(C_n(p))$ and $(G_n(p))$. For the proofs of the strong consistency of the LS estimator for general linear autoregressive model, we refer the reader to [11, 16, 8]. Also, one can find in [8, 11, 12, 16] various results on the asymptotic behavior of the empirical estimator of the covariance associated with process (3.1).

One might wonder how the convergence of the moments in the ASCLT helps us to deduce the almost sure asymptotic properties of the sequences $(C_n(p))$ and $(G_n(p))$. It follows from (3.1) and (3.4) that

$$\widehat{\theta}_n - \theta = S_{n-1}^{-1} M_n \quad (3.5)$$

where

$$M_n = M_0 + \sum_{k=1}^n \Phi_{k-1} \varepsilon_k$$

with $M_0 = -S\theta$. If

$$\pi_n = (\theta - \widehat{\theta}_n)^t \Phi_n = X_{n+1} - \widehat{\theta}_n^t \Phi_n - \varepsilon_{n+1}, \quad (3.6)$$

relations (3.5) and (3.6) yield

$$\pi_n^2 = M_n^t S_{n-1}^{-1} \Phi_n \Phi_n^t S_{n-1}^{-1} M_n.$$

Applying now Riccati's formula given e.g. in [7] pages 96 and 99, it comes

$$S_{n-1}^{-1} = S_n^{-1} + (1 - f_n) S_{n-1}^{-1} \Phi_n \Phi_n^t S_{n-1}^{-1}.$$

Hence, one can write

$$a_n(1) = M_n^t S_{n-1}^{-1} M_n - M_n^t S_n^{-1} M_n = (1 - f_n) \pi_n^2. \quad (3.7)$$

It is often difficult to obtain asymptotic information about the explosion coefficient f_n . Nevertheless, in our framework, it is possible to show that (f_n) converges almost surely to zero. So, the asymptotic behavior of $(G_n(p))$ and $(C_n(p))$ can be derived from the properties of $(a_n(1))^p$, under some suitable moment conditions on the driven noise (ε_n) . By the same token, the moments of order $2p$ can be estimated and controlled.

Corollary 3.1. *Under the assumptions of theorem 2.1, one has almost surely*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (a_k(1))^p = \begin{cases} 0 & \text{if } p > 1, \\ \sigma^2 & \text{if } p = 1. \end{cases} \quad (3.8)$$

Proof. When $p = 1$, the convergence (3.8) corresponds precisely to (2.3). Suppose now that $p > 1$. Since

$$a_n(1) \leq f_n V_n$$

and $f_n \rightarrow 0$ almost surely, we get at once, applying Lemma 2.2 and Kronecker's lemma,

$$0 \leq \lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (a_k(1))^p \leq \lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (f_k V_k)^p = 0 \quad \text{a.s.}$$

■

3.2 Moment estimation, prediction and estimation errors

Assume (ε_n) forms a martingale difference sequence with $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s., for all $n \geq 1$, and let

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n \varepsilon_k^2.$$

If ε_n has a conditional moment of order $a > 2$, then the strong law of large numbers for martingales implies the almost sure convergence of Δ_n to σ^2 . Under the assumptions of theorem 2.1 with $p = 1$, the convergence (3.8) leads to the strong consistency of the estimator of σ^2

$$\Gamma_n = \frac{1}{n} \sum_{k=0}^{n-1} (X_{k+1} - \hat{\theta}_k^t \Phi_k)^2,$$

since

$$\lim_{n \rightarrow \infty} \frac{n}{\log d_n} (\Gamma_n - \Delta_n) = \sigma^2 \quad \text{a.s.} \quad (3.9)$$

Hence, a natural estimator of higher moments of (ε_n) can be proposed. For any integer $p \geq 1$, let

$$\Gamma_n(2q) = \frac{1}{n} \sum_{k=0}^{n-1} (X_{k+1} - \hat{\theta}_k^t \Phi_k)^{2p}. \quad (3.10)$$

One can readily observe that $n\Gamma_n(2p) = C_n(p)$. For any integer $p \geq 1$, let

$$\Delta_n(2p) = \frac{1}{n} \sum_{k=0}^{n-1} \varepsilon_k^{2p}. \quad (3.11)$$

Almost sure asymptotic properties of $\Gamma_n(2p)$ are given in the next corollary.

Corollary 3.2. *Assume that (ε_n) satisfies (H_p) with $p \geq 1$. In addition, suppose that for some integer $1 \leq q \leq p$, $\mathbb{E}[\varepsilon_{n+1}^{2q} | \mathcal{F}_n] = \sigma(2q)$ a.s. Then, $\Gamma_n(2q)$ is a strongly consistent estimator of $\sigma(2q)$ with*

$$\left(\Gamma_n(2q) - \Delta_n(2q) \right)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad (3.12)$$

Proof. We already saw via (3.9) that Corollary 3.2 holds for $q = 1$. Assume now $q \geq 2$. By expanding the expression of $\Gamma_n(2q)$, equality (3.6) leads to

$$n(\Gamma_n(2q) - \Delta_n(2q)) = \sum_{k=0}^{n-1} \pi_k^{2q} + \sum_{\ell=1}^{2q-1} C_q^\ell \sum_{k=0}^{n-1} \pi_k^{2q-\ell} \varepsilon_{k+1}^\ell.$$

We deduce from (3.7) together with the almost sure convergence of f_n to zero and Corollary 3.1 that

$$\sum_{k=0}^n \pi_k^{2q} = o(\log d_n) \quad \text{a.s.}$$

For all $\ell \in \{1, \dots, 2q-1\}$, let us write

$$\sum_{k=0}^{n-1} \pi_k^{2q-\ell} \varepsilon_{k+1}^\ell = P_n(\ell) + Q_n(\ell)$$

with

$$P_n(\ell) = \sum_{k=0}^{n-1} \pi_k^{2q-\ell} \sigma_k(\ell) \quad \text{and} \quad Q_n(\ell) = \sum_{k=0}^{n-1} \pi_k^{2q-\ell} e_{k+1}(\ell),$$

where $\sigma_n(\ell) = \mathbb{E}[\varepsilon_{n+1}^\ell | \mathcal{F}_n]$ and $e_{n+1}(\ell) = \varepsilon_{n+1}^\ell - \sigma_n(\ell)$. First, since the moments $\sigma_n(\ell)$ are almost surely bounded, it comes

$$|P_n(\ell)| = \mathcal{O}\left(\sum_{k=0}^{n-1} \pi_k^{2q-\ell}\right) = \mathcal{O}(\log d_n) \quad \text{a.s.}$$

Moreover, from the estimate

$$|Q_n(\ell)|^2 = \mathcal{O}(n \log d_n) \quad \text{a.s.}$$

we get

$$n^2 (\Gamma_n(2q) - \Delta_n(2q))^2 = \mathcal{O}(n \log d_n) \quad \text{a.s.},$$

which concludes the proof of Corollary 3.2. ■

It is now possible to deduce from Corollary 3.2 the asymptotic behavior of $(C_n(q))$. Under the assumptions of Corollary 3.2, the convergence (3.12) yields $C_n(q)/n$ converges a.s. to $\sigma(2q)$. Moreover since the conditional moment of order $a > 2p$ of (ε_n) is almost surely finite, Chow Lemma leads to

$$\left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k^{2q} - \sigma(2q) \right| = o(n^{c-1}) \quad \text{a.s.} \quad (3.13)$$

for all c such that $2pa^{-1} < c < 1$. Hence it follows from (3.12) and (3.13) that, if $\log d_n = o(n^c)$,

$$\left| \frac{1}{n} C_n(q) - \sigma(2q) \right|^2 = o(n^{c-1}) \quad \text{a.s.}$$

Before enouncing the result on the cumulative estimation error $(G_n(p))$, we need another corollary of theorem 2.1.

Corollary 3.3. *Under the assumptions of theorem 2.1, one has*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k \left((\hat{\theta}_k - \theta)^t S_k (\hat{\theta}_k - \theta) \right)^p = \ell(p) \quad a.s. \quad (3.14)$$

In addition, assume that it exists a positive definite symmetric matrix L such that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} S_n = L \quad a.s. \quad (3.15)$$

Then, we have

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} \left((\hat{\theta}_k - \theta)^t L (\hat{\theta}_k - \theta) \right)^p = \ell(p) \quad a.s. \quad (3.16)$$

Remark 3. Since L is positive definite, (3.16) immediately yields

$$G_n(p) = \mathcal{O}(\log n) \quad a.s.$$

Proof. From the definitions of S_n and of $\hat{\theta}_n$, it is easy to see that

$$(\hat{\theta}_n - \theta)^t S_n (\hat{\theta}_n - \theta) = V_n + g_n^2.$$

Hence, it follows from the convergence (2.2) and from the convergence of the explosion coefficient (f_n) to zero, together with Kronecker's Lemma that almost surely

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k \left((\hat{\theta}_k - \theta)^t S_k (\hat{\theta}_k - \theta) \right)^p = \lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k V_k^p = \ell(p).$$

Then, the convergence (3.14) is a straight forward consequence of theorem 2.1. Using the formel object \sqrt{L} , we get

$$\left((\hat{\theta}_n - \theta)^t L (\hat{\theta}_n - \theta) \right)^p \sim (n^{-2} m_n \beta_n)^p \sim (d^2 m_n \beta_n^{-1})^p \quad a.s.$$

Thus,

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} \left((\hat{\theta}_k - \theta)^t L (\hat{\theta}_k - \theta) \right)^p = \lim_{n \rightarrow \infty} \frac{d^{p+1}}{\log n} \sum_{k=1}^n \frac{m_k^p}{\beta_k} \quad a.s.$$

The classical Abel transform gives the decomposition

$$\sum_{k=1}^n \gamma_k m_k^p = \frac{m_n^p}{\beta_n} (\Sigma_n - d(n-1)) - \frac{m_1^p}{\beta_0} \Sigma_0 + r_n + d \sum_{k=1}^{n-1} \frac{m_k^p}{\beta_k}, \quad (3.17)$$

with

$$\Sigma_n = \sum_{k=1}^n \beta_k \gamma_k = \sum_{k=1}^n \Phi_k^t L^{-1} \Phi_k \sim \beta_n$$

and

$$r_n = \sum_{k=1}^{n-1} \left(\frac{m_k^p}{\beta_k} - \frac{m_{k+1}^p}{\beta_{k+1}} \right) (\Sigma_k - kd).$$

Moreover,

$$\frac{m_n^p}{\beta_n} (\Sigma_n - d(n-1)) - \frac{m_1^p}{\beta_0} \Sigma_0 = o(\log d_n) \quad \text{a.s.}$$

So, it only remains to prove that $r_n = o(\log n)$ a.s. Indeed, lemma 2.2 yields

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{m_k^p}{\beta_k} = \frac{1}{d} \lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \gamma_k m_k^p = \frac{\ell(p)}{d^{p+1}} \quad \text{a.s.}$$

Then, splitting r_n into two terms

$$r_n = \sum_{k=1}^{n-1} \frac{\Sigma_k - kd}{\beta_k} (m_k^p - m_{k+1}^p) + \sum_{k=1}^{n-1} \frac{\Sigma_k - kd}{\beta_k} \gamma_{k+1} m_{k+1}^p, \quad (3.18)$$

and using the proof of theorem 2.1 together with (2.12), we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^{n-1} \frac{\Sigma_k - kd}{\beta_k} \left(\beta_k^{-p} (\beta_k^p - \beta_{k-1}^p) m_k^p - w_{k+1} - b_{k+1} \right) = 0 \quad \text{a.s.}$$

The second term is almost surely $o(\log n)$: this is a mere consequence of Lemma 2.2 and of the a.s. convergence of $\Sigma_n - nd/\beta_n$ to zero. Finally, we have

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} \left((\hat{\theta}_k - \theta)^t L (\hat{\theta}_k - \theta) \right)^p = \ell(p) \quad \text{a.s.}$$

■

We shall apply now these asymptotic properties to autoregressive processes and to branching processes with immigration, which are both particular cases of the general stochastic regression process (3.1).

3.3 The linear autoregressive process

The linear autoregressive process is defined for all $n \geq 1$ by

$$X_{n+1} = \sum_{k=1}^d \theta_k X_{n-k+1} + \varepsilon_{n+1}. \quad (3.19)$$

Let C denote the companion matrix associated with (X_n)

$$C = \begin{pmatrix} \theta_1 & \theta_2 & \dots & \theta_{d-1} & \theta_d \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

We shall focus our attention on the stable case which means that we assume that $\rho(C) < 1$ where $\rho(C)$ is the spectral radius of the matrix C . In addition, we also assume that (ε_n) is a martingale difference sequence which satisfies (H_p) with $p \geq 1$. If $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s. and the matrix

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix},$$

then convergence (3.15) holds with the limiting matrix L given by

$$L = \sum_{k=0}^{\infty} C^k \Gamma (C^t)^k. \tag{3.20}$$

Moreover, one can immediately see that the matrix L is positive definite [7, 8]. Then we are in a position to state our following result.

Corollary 3.4. *Assume that (ε_n) satisfies (H_p) with $p \geq 1$. In addition, suppose that for some integer $1 \leq q \leq p$, $\mathbb{E}[\varepsilon_{n+1}^{2q} | \mathcal{F}_n] = \sigma(2q)$ a.s. Then, $\Gamma_n(2q)$ is a strongly consistent estimator of $\sigma(2q)$ with*

$$\left(\Gamma_n(2q) - \Delta_n(2q)\right)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \tag{3.21}$$

Moreover, we also have

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} \left((\hat{\theta}_k - \theta)^t L (\hat{\theta}_k - \theta) \right)^p = \ell(p) \quad a.s. \tag{3.22}$$

3.4 A branching process with immigration

3.4.1 Estimation of the mean

Branching process with immigration play an increasingly important role in statistical physics, computational biology and evolutionary theory. The concept of immigration is related to situation in which the population can be

enriched by exogenous contributions. The branching process with immigration (X_n) is given for all $n \geq 1$ by the recursive relation

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k} + I_{n+1}, \quad (3.23)$$

where $(Y_{n,k})$ and (I_n) are two independent sequences of i.i.d. nonnegative, integer-valued random variables. The initial ancestor is $X_0 = 1$. The distribution of $(Y_{n,k})$ is commonly called the offspring distribution, while that of (I_n) is known as the immigration distribution.

Define

$$\begin{aligned} \mathbb{E}[Y_{n,k}] &= m, & \mathbb{E}[I_n] &= \lambda, \\ \text{var}[Y_{n,k}] &= \sigma^2, & \text{var}[I_n] &= b^2. \end{aligned}$$

We are interested in the estimation of all the parameters $m, \lambda, \sigma^2, b^2$. Relation (3.23) may be rewritten as the autoregressive form

$$X_{n+1} = mX_n + \lambda + \varepsilon_{n+1} \quad (3.24)$$

where $\varepsilon_{n+1} = X_{n+1} - mX_n - \lambda$. Consequently, the branching process with immigration is a particular case of the stochastic regression process given by (3.1) with $\Phi_n^t = (X_n, 1)$ and $\theta^t = (m, \lambda)$. However, one can observe that the situation is a little bit more tricky as (ε_n) is a martingale difference sequence with unbounded conditional variance

$$\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2 X_n + b^2.$$

To circumvent this technical difficulty, we introduce the following regression process

$$Z_{n+1} = \theta^t \Psi_n + \xi_{n+1},$$

where the random variables Z_{n+1} , Ψ_n and ξ_n are given by

$$Z_{n+1} = c_n^{-1/2} X_{n+1}, \quad \Psi_n = c_n^{-1/2} \Phi_n, \quad \xi_{n+1} = c_n^{-1/2} \varepsilon_{n+1},$$

with

$$c_n = X_n + 1.$$

Hereafter, (ξ_n) is a martingale difference sequence with bounded conditional variance $\mathbb{E}[\xi_{n+1}^2 | \mathcal{F}_n] \leq \sigma^2 + b^2$ a.s. The mean vector $\theta^t = (m, \lambda)$ will be estimated by the conditional least-squares estimator

$$\hat{\theta}_n = S_n^{-1} \sum_{k=1}^n c_k^{-1} \Phi_k X_k,$$

where

$$S_n = I_2 + \sum_{k=0}^n c_k^{-1} \Phi_k \Phi_k^t.$$

In the subcritical case $m < 1$, Wei and Winnicki [18] have shown the almost sure convergence

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n = L \quad \text{a.s.}$$

where the limiting matrix L is given by

$$L = \begin{pmatrix} \mathbb{E} \left[\frac{X^2}{X+1} \right] & \mathbb{E} \left[\frac{X}{X+1} \right] \\ \mathbb{E} \left[\frac{X}{X+1} \right] & \mathbb{E} \left[\frac{1}{X+1} \right] \end{pmatrix}.$$

The notation X stands for a random variable sharing the same distribution as the stationary distribution of the Markov chain (X_n) . Consequently, the matrix L is positive definite and the following result holds.

Corollary 3.5. *Assume that (ε_n) satisfies (H_p) with $p \geq 1$. In addition, suppose that for some integer $1 \leq q \leq p$, $\mathbb{E}[\varepsilon_{n+1}^{2q} | \mathcal{F}_n] = \sigma(2q)$ a.s. Then, $\Gamma_n(2q)$ is a strongly consistent estimator of $\sigma(2q)$ with*

$$\left(\Gamma_n(2q) - \Delta_n(2q) \right)^2 = \mathcal{O} \left(\frac{\log n}{n} \right) \quad \text{a.s.} \quad (3.25)$$

Moreover, we also have

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} \left((\hat{\theta}_k - \theta)^t L (\hat{\theta}_k - \theta) \right)^p = \ell(p) \quad \text{a.s.} \quad (3.26)$$

3.4.2 Estimation of the variance

It follows from equation (3.24) that

$$\varepsilon_{n+1}^2 = \sigma^2 X_n + b^2 + V_{n+1},$$

where (V_n) is a martingale difference sequence satisfying

$$\mathbb{E}[V_{n+1}^2 | \mathcal{F}_n] = 2\sigma^4 X_n^2 + X_n(\tau^4 - 3\sigma^4 + 4b^2\sigma^2) + \nu^4 - b^4,$$

with

$$\tau^4 = \mathbb{E}[(Y_{n,k} - m)^4] \quad \text{and} \quad \nu^4 = \mathbb{E}[(I_n - \lambda)^4].$$

Consequently, we infer that

$$\mathbb{E}[c_n^{-2} V_{n+1}^2 | \mathcal{F}_n] \leq \tau^4 + 4b^2\sigma^2 + \nu^4.$$

As before, we estimate the vector of variances $\eta^t = (\sigma^2, b^2)$ by the following conditional least-squares estimator

$$\hat{\eta}_n = Q_n^{-1} \sum_{k=1}^n c_k^{-2} \Phi_k \hat{\varepsilon}_{k+1}^2 \quad \text{with} \quad \hat{\varepsilon}_{k+1} = X_{k+1} - \hat{\theta}_k \Phi_k,$$

where

$$Q_n = I_2 + \sum_{k=0}^n c_k^{-2} \Phi_k \Phi_k^t.$$

In the subcritical case $m < 1$, it was established by Winnicki [19] that

$$\lim_{n \rightarrow \infty} \frac{1}{n} Q_n = \Lambda$$

where Λ is the positive definite limiting matrix given by

$$\Lambda = \begin{pmatrix} \mathbb{E} \left[\frac{X^2}{(X+1)^2} \right] & \mathbb{E} \left[\frac{X}{(X+1)^2} \right] \\ \mathbb{E} \left[\frac{X}{(X+1)^2} \right] & \mathbb{E} \left[\frac{1}{(X+1)^2} \right] \end{pmatrix}.$$

Our last result is the following:

Corollary 3.6. *Assume that (ε_n) satisfies (H_p) for $p \geq 2$. Then*

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} ((\hat{\eta}_k - \eta)^t \Lambda (\hat{\eta}_k - \eta))^p = \ell(p) \quad a.s. \quad (3.27)$$

References

- [1] BERCU, B. On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. *Stochastic Processes and their applications* 111 (2004), 157–173.
- [2] BERCU, B., AND FORT, J. C. A moment approach for the almost sure central limit theorem for martingales. *Studia Scientiarum Mathematicarum Hungarica* 45, 1 (2008), 139–159.
- [3] BROSAMLER, G. A. An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.* 104, 3 (1988), 561–574.
- [4] CHAÂBANE, F. Invariance principles with logarithmic averaging for martingales. *Studia Sci. Math. Hungar.* 37, 1-2 (2001), 21–52.
- [5] CHAÂBANE, F., AND MAÂOUIA, F. Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab. Statist.* 4 (2000), 137–189.

- [6] CHAÂBANE, F., MAÂOUIA, F., AND TOUATI, A. Généralisation du théorème de la limite centrale presque-sûr pour les martingales vectorielles. *C. R. Acad. Sci. Paris Sér. I Math.* 326, 2 (1998), 229–232.
- [7] DUFLO, M. *Random Iterative Methods*. Springer-Verlag, 1997.
- [8] DUFLO, M., SENOUSI, R., AND TOUATI, A. Propriétés asymptotiques presque sûres de l'estimateur des moindres carrés d'un modèle autoregressif vectoriel. *Ann. Inst. Henri Poincaré* 27, 1 (1991), 1–25.
- [9] GOODWIN, G., AND SIN, K. *Adaptative Filtering Prediction and Control*. Prentice-Hall, Englewood Cliffs, N.J., 1984.
- [10] LACEY, M., AND PHILLIP, W. A note on the almost sure central limit theorem. *Statist. Probab. Letters* 9 (1990), 201–205.
- [11] LAI, T., AND WEI, C. Least-squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics* 10, 1 (1982), 154–166.
- [12] LAI, T., AND WEI, C. Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analysis* 13 (1983), 1–23.
- [13] LIFSHITS, M. A. Lecture notes on almost sure limit theorems. *Publications IRMA* 54 (2001), 1–25.
- [14] LIFSHITS, M. A. Almost sure limit theorem for martingales. In *Limit theorems in probability and statistics*. János Bolyai Math. Soc., Budapest, 2002, pp. 367–390.
- [15] SCHATTE, P. On strong versions of the central limit theorem. *Math. Nachr.* 137 (1988), 249–256.
- [16] WEI, C. Asymptotic properties of least-squares estimates in stochastic regression models. *The Annals of Statistics* 13, 4 (1985), 1498–1508.
- [17] WEI, C. Adaptative prediction by least squares predictors in stochastic regression models with applications to time series. *The Annals of Statistics* 15, 4 (1987), 1667–1682.
- [18] WEI, C. Z., AND WINNICKI, J. Estimation of the means in the branching process with immigration. *Ann. Statist.* 18, 4 (1990), 1757–1773.
- [19] WINNICKI, J. Estimation of the variances in the branching process with immigration. *Probab. Theory Rel. Fields* 88, 1 (1991), 77–106.



Centre de recherche INRIA Paris – Rocquencourt

Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex

Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier

Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq

Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique

615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex

Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex

Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur

INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-6399