

## ON THE ALMOST SURE CENTRAL LIMIT THEOREM FOR VECTOR MARTINGALES: CONVERGENCE OF MOMENTS AND STATISTICAL APPLICATIONS

BERNARD BERCU,\* *Université Bordeaux I and INRIA Bordeaux Sud-Ouest*

PEGGY CÉNAC,\*\* *Université de Bourgogne*

GUY FAYOLLE,\*\*\* *INRIA Paris-Rocquencourt*

### 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.

*Keywords:* Almost-sure central limit theorem; vector martingale; moment, stochastic regression

2000 Mathematics Subject Classification: Primary 60B12

Secondary 46N30

### 1. Introduction

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

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

It follows from the ordinary central limit theorem that

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

which implies that, for any bounded continuous real function  $h$ ,

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

---

Received 15 July 2008; revision received 23 January 2009.

\* Postal address: Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, France.

Email address: bernard.bercu@math.u-bordeaux1.fr

\*\* Postal address: Université de Bourgogne, Institut de Mathématiques de Bourgogne, UMR 5584, 9 rue Alain Savary, BP 47870, 21078 Dijon Cedex, France. Email address: peggy.cenac@u-bourgogne.fr

\*\*\* Postal address: INRIA CR Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France. Email address: guy.fayolle@inria.fr

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

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

where  $\delta$  denotes the Dirac delta function, satisfies

$$G_n \Rightarrow G \quad \text{almost surely (a.s.).}$$

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 [16], and, in its present form, by Lacey and Phillip [11]. In contrast with the wide literature on the ASCLT for independent random variables, very few references are available on the ASCLT for martingales, apart from the recent work of Bercu and Fort [1], [2] and the important contributions of Chaâbane and Maâouia [4], Chaâbane [5], Chaâbane and Touati [6], and Lifshits [14], [15]. The ASCLT for scalar martingales is as follows. Let  $(\varepsilon_n)$  be a martingale difference sequence adapted to a filtration  $\mathbb{F} = (\mathcal{F}_n)$  with  $E[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$  a.s. Let  $(\Phi_n)$  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  $(\Phi_n)$ :

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

As soon as  $(f_n)$  goes to 0 a.s. and, under a reasonable assumption on the conditional moments of  $(\varepsilon_n)$ , the ASCLT for martingales asserts that the empirical measure

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

In other words, for any bounded continuous real function  $h$ ,

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

It is quite natural to overcome the case of unbounded functions  $h$ . To be more precise, we might wonder if convergence (1.2) remains true for unbounded functions  $h$ . It has recently been shown in [1] and [2] that, whenever  $(\varepsilon_n)$  has a finite conditional moment of order greater than  $2p$ , then 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  $(\varepsilon_n)$  is a martingale difference sequence such that  $E[\varepsilon_{n+1}^2 \mid \mathcal{F}_n] = \sigma^2$  a.s. and satisfying, for some integer  $p \geq 1$  and some real number  $a > 2p$ ,*

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

*If the explosion coefficient  $(f_n)$  tends to 0 a.s. then*

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

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  $\mathbb{F} = (\mathcal{F}_n)$ . Its increasing process is the sequence  $(\langle M \rangle_n)$  of symmetric, positive, semidefinite square matrices of order  $d$  given by

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

A first version of the ASCLT for discrete vector martingales was proposed in [5] and [7], 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 martingale 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  $(\Phi_n)$  denotes a sequence of random vectors of dimension  $d$ , adapted to  $\mathbb{F}$ . We also introduce

$$S_n = \sum_{k=0}^n \Phi_k \Phi_k^\top + S,$$

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

$$f_n = \Phi_n^\top S_n^{-1} \Phi_n = \frac{d_n - d_{n-1}}{d_n}.$$

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. In Section 3 we propose some statistical applications to estimation and prediction errors for linear autoregressive processes and for branching processes with immigration.

## 2. On the convergence of moments

As mentioned above, our main result is given in Theorem 2.1 and it clearly extends Theorem 1.1 to the vector case. Note that, 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  $(\varepsilon_n)$  be a martingale difference sequence satisfying the homogeneity condition  $E[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$  a.s. and such that, for some integer  $p \geq 1$  and some real number  $a > 2p$ ,*

$$\sup_{n \geq 0} E[|\varepsilon_{n+1}|^a | \mathcal{F}_n] < \infty \quad \text{a.s.} \tag{2.1}$$

*In addition, assume that the explosion coefficient  $(f_n)$  tends to 0 a.s., and that there exists a positive random sequence  $(\alpha_n)$  increasing a.s. to  $\infty$  and an invertible symmetric matrix  $L$  such that*

$$\lim_{n \rightarrow \infty} \frac{1}{\alpha_n} S_n = L \quad \text{a.s.} \tag{2.2}$$

*Then, the following limits hold a.s.:*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k (M_k^\top S_{k-1}^{-1} M_k)^p = \ell(p) = d\sigma^{2p} \prod_{j=1}^{p-1} (d + 2j), \tag{2.3}$$

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n ((M_k^\top S_{k-1}^{-1} M_k)^p - (M_k^\top S_k^{-1} M_k)^p) = \lambda(p) = \frac{p}{d} \ell(p). \tag{2.4}$$

**Remark 2.1.** The limit  $\ell(p)$  corresponds exactly to the moment of order  $2p$  of the norm of a Gaussian vector  $\mathcal{N}(0, \sigma^2 I_d)$ , where  $I_d$  is the identity matrix of order  $d$ . Consequently, Theorem 2.1 indeed shows the convergence of moments of order  $2p$  in the ASCLT for vector martingales. Here, the deterministic normalisation given in [5] has been replaced by the natural random normalisation associated with the increasing process.

**Remark 2.2.** The convergence hypothesis (2.2) clearly implies that  $f_n \rightarrow 0$  a.s. if and only if  $\alpha_n \sim \alpha_{n-1}$  a.s. As a matter of fact, we deduce from (2.2) that

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

### 2.1. Proof of the Theorem 2.1

For the sake of shortness, we shall define the following variables:

$$V_n = M_n^\top S_{n-1}^{-1} M_n, \quad \varphi_n = \alpha_n^{-1} \Phi_n^\top L^{-1} \Phi_n, \quad v_n = \alpha_{n-1}^{-1} M_n^\top L^{-1} M_n.$$

First of all, by using the symmetry of  $L$ , convergence (2.2) ensures that

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

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

Indeed, we have the decomposition

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

Moreover, the matrix  $R_n = \alpha_n L^{1/2} S_n^{-1} L^{1/2} - I_d$  is symmetric. Thus, if  $\rho_n$  stands for its spectral radius, we clearly have

$$|\alpha_n^{-1} \Phi_n^\top L^{-1/2} R_n L^{-1/2} \Phi_n| \leq \rho_n \varphi_n.$$

Since  $\rho_n$  converges to 0 a.s., the last inequality leads to (2.5). The proof of (2.6) goes along the same lines starting from the decomposition

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

Hence, by (2.6),  $V_n^p = v_n^p + o(v_n^p)$  a.s. In order to find the limit in (2.3), it is enough, by Toeplitz's lemma, to study the convergence

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

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

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

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

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

so that

$$m_{n+1} = (1 - \gamma_n) m_n + 2\delta_n \varepsilon_{n+1} + \gamma_n \varepsilon_{n+1}^2. \tag{2.7}$$

Theorem 2.1 relies essentially on the following lemma.

**Lemma 2.1.** *Under the assumptions of Theorem 2.1, we have*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n \gamma_k m_k^p = \frac{\ell(p)}{d^{p+1}} \quad \text{a.s.} \tag{2.8}$$

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

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (1 - f_k) g_k^2 m_k^{p-1} = \frac{\lambda(p)}{pd^{p-1}} \quad \text{a.s.} \tag{2.9}$$

*Proof.* Raising equality (2.7) to the power  $p$  implies that

$$m_{n+1}^p = \sum_{k=0}^p \sum_{\ell=0}^k 2^{k-\ell} C_p^k C_k^\ell \gamma_n^\ell \delta_n^{k-\ell} ((1 - \gamma_n) m_n)^{p-k} \varepsilon_{n+1}^{k+\ell}. \tag{2.10}$$

After some straightforward simplifications, we obtain the relation

$$m_{n+1}^p + \mathcal{A}_n(p) = m_1^p + \mathcal{B}_{n+1}(p) + \mathcal{W}_{n+1}(p), \tag{2.11}$$

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

*Case 1:  $p = 1$ .* First of all, we can observe that

$$\log \beta_n \sim \sum_{k=1}^n \gamma_k \quad \text{a.s.}$$

Hence, by Chow's lemma [8, Theorem 1.3.18, p. 22] we obtain

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

Now, the strong law of large numbers for martingales [8, Theorem 1.3.24, p. 26] together with the inequality  $\delta_n^2 \leq \gamma_n m_n$  implies that  $\mathcal{B}_{n+1}(1) = o(\mathcal{A}_n(1))$  a.s. In addition, from relation (2.30) of [18], it follows that  $m_{n+1} = o(\log \beta_n)$  a.s. Consequently, by (2.11),

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

But the basic convergence assumption, (2.2), immediately implies that

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

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 (2.8). As for the proof of (2.9), we can proceed in the same way, starting from the decomposition

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

Letting  $a_n(1) = M_n^\top S_{n-1}^{-1} M_n - M_n^\top S_n^{-1} M_n = V_n - h_n$ , we can write

$$V_{n+1} + A_n = V_1 + B_{n+1} + W_{n+1}, \tag{2.12}$$

where

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

By Riccati's formula,

$$g_n^2 = (1 - f_n) a_n(1). \tag{2.13}$$

Consequently, it follows from (2.13) together with the strong law of large numbers for martingales that

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

On the other hand,  $m_n = o(\log \beta_n)$  a.s. Hence, (2.6) gives the almost sure estimate  $V_n = o(\log d_n)$ . Furthermore, as  $(f_n)$  tends to 0 a.s., we also have

$$\log d_n \sim \sum_{k=1}^n f_k \quad \text{a.s.}$$

Therefore, Chow's lemma yields

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

We conclude the proof of Lemma 2.1 for  $p = 1$  by simply dividing both sides of (2.12) by  $\log d_n$  and letting  $n$  tend to  $\infty$ .

*Case 2:  $p \geq 2$ .* First of all, using Chow's lemma once again, we can write

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

Also, we shall show that

$$\mathcal{B}_{n+1}(p) = \frac{p}{d^{p+1}} \ell(p) \log d_n + o(\log d_n) + o(\mathcal{A}_n(p)) \quad \text{a.s.} \tag{2.14}$$

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

$$\varepsilon_{n+1}^\ell = e_{n+1}(\ell) + \mathbb{E}[\varepsilon_{n+1}^\ell \mid \mathcal{F}_n] = e_{n+1}(\ell) + \sigma_n(\ell),$$

we can split

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

with

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

First, for any  $\ell$  such that  $1 \leq \ell \leq p - 1$ , using the strong law of large numbers for martingales once again and Equation (2.30) of [18], we have

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

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

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

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

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

Now suppose that  $p \leq \ell \leq 2p - 1$ . 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 find that, 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  $\mathcal{C}_{n+1}(\ell) = o(v_n(\ell))$  a.s., where

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

For  $\ell > p$ , we apply Hölder's inequality with exponents  $\ell/p$  and  $\ell/(\ell - p)$ . Then,  $v_n(\ell) = o(\log d_n)$  a.s. In the particular case  $p = \ell$ , we find by the strong law of large numbers for martingales that  $|\mathcal{C}_{n+1}(\ell)|^2 = \mathcal{O}(\tau_n(p) \log \tau_n(p))$  a.s., after having set

$$\tau_n(p) = \sum_{k=1}^n b_k(p)^2 = \mathcal{O}\left(\sum_{k=1}^n \gamma_k^p m_k^p\right).$$

Then we can deduce that  $\mathcal{C}_{n+1}(\ell) = o(\log d_n)$  a.s., since by Equation (2.30) of [18],  $m_n^p = o((\log d_n)^\delta)$  a.s. for some  $0 < \delta < 1$ . As for the last term  $\mathcal{D}_n(2)$ , we need 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 show that

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

Then the induction assumption and Toeplitz's lemma imply that

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

Consequently, we find that

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

We can observe that this convergence is also valid for  $p = 2$ , which leads to (2.14). On the other hand, still applying Equation (2.30) of [18], 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

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

we finally obtain

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

which completes the proof of convergence (2.8). As for the second part of Lemma 2.1, 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  $g_n = \Phi_n^\top S_{n-1}^{-1} M_n$  and  $h_n = M_n^\top S_n^{-1} M_n$ .

The proof of Theorem 2.1 is completed as relations (2.3) and (2.4) are clearly direct consequences of (2.8) and (2.9), respectively. Indeed, since  $\beta_n \sim d\alpha_n$  a.s., we have, a.s.,

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

Therefore, convergence (2.8) immediately yields (2.3). Moreover, in the particular case  $p = 1$ , the second convergence (2.9) is exactly (2.4). Now, for  $p \geq 2$ , the elementary expansion

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

leads to

$$\begin{aligned} a_n(p) &= (M_n^\top S_{n-1}^{-1} M_n)^p - (M_n^\top S_n^{-1} M_n)^p \\ &= a_n(1) V_n^{p-1} \sum_{q=0}^{p-1} \left( \frac{V_n - a_n(1)}{V_n} \right)^{p-1-q}. \end{aligned} \quad (2.15)$$

Riccati's formula implies that

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

As  $(f_n)$  tends to 0 a.s., we clearly have  $a_n(1) = o(V_n)$  a.s. and, by (2.15),

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

Finally, we obtain

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

In most statistical applications encountered so far, the convergence assumption (2.2) is satisfied. However, this technical hypothesis somehow circumvents the vector problem, which in its full generality is not yet solved. Indeed, (2.2) entails that all eigenvalues of the matrix  $S_n$  grow to  $\infty$  at the same speed  $\alpha_n$ . Thus, our method of proof has 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 (2.2).

**Conjecture 2.1.** *Let  $(\varepsilon_n)$  be a martingale difference sequence satisfying the condition that  $E[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$  a.s. and assumption (2.1) introduced in Theorem 2.1 for some integer  $p \geq 1$ . Then*

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

### 3. Applications

#### 3.1. Linear regression models

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

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

where  $\theta \in \mathbb{R}^d$  is the unknown parameter. The random variables  $X_n$ ,  $\Phi_n$ , and  $\varepsilon_n$  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  $\theta$ , we shall investigate the asymptotic performance of  $\hat{\theta}_n^\top \Phi_n$  as a predictor of  $X_{n+1}$ . More precisely, we shall focus on the prediction error,  $X_{n+1} - \hat{\theta}_n^\top \Phi_n$ , and on the estimation error,  $\hat{\theta}_n - \theta$ . In fact, it is more relevant and efficient [10] to consider the cumulative prediction and estimation errors defined, for  $p \geq 1$ , as

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

and

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

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 estimator

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^n \Phi_{k-1} X_k. \tag{3.2}$$

It turns out that Theorem 2.1 allows us to improve the results of [1] and [2]. To the best of the authors' knowledge, only partial results in the particular case  $p = 1$  have been obtained, namely in [8] and [18], where the authors derived the asymptotics of  $(C_n(p))$  and  $(G_n(p))$ . For the proofs of the strong consistency of the least-squares estimator for general linear autoregressive processes, we refer the reader to [9], [12], and [17]. Also, various results on the asymptotic behaviour of the empirical estimator of the covariance associated with process (3.1) can be found in [9], [12], [13], and [17].

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 (3.2), that

$$\hat{\theta}_n - \theta = S_{n-1}^{-1} M_n, \tag{3.3}$$

where

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

with  $M_0 = -S\theta$ . If

$$\pi_n = (\theta - \hat{\theta}_n)^\top \Phi_n = X_{n+1} - \hat{\theta}_n^\top \Phi_n - \varepsilon_{n+1}, \tag{3.4}$$

relations (3.3) and (3.4) yield

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

It follows from Riccati's formula given in [8, pp. 96 and 99] that

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

Hence, we can write

$$a_n(1) = M_n^\top S_{n-1}^{-1} M_n - M_n^\top S_n^{-1} M_n = (1 - f_n) \pi_n^2. \tag{3.5}$$

It is often difficult to obtain asymptotic information on the explosion coefficient  $f_n$ . Nevertheless, in our framework, it is possible to show that  $(f_n)$  converges a.s. to 0. So, the asymptotic behaviours of  $(G_n(p))$  and  $(C_n(p))$  can be deduced from the properties of  $(a_n(1))^p$ , under some suitable moment conditions on the driven noise  $(\varepsilon_n)$ . By the same token, the moments of order  $2p$  can also be estimated.

**Corollary 3.1.** *Under the assumptions of Theorem 2.1, we have, a.s.,*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n (a_k(1))^p = \begin{cases} 0 & \text{if } p > 1, \\ \sigma^2 & \text{if } p = 1. \end{cases} \tag{3.6}$$

*Proof.* In the particular case  $p = 1$ , convergence (3.6) is exactly given by (2.4). Now suppose that  $p > 1$ . Since

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

and  $f_n \rightarrow 0$  a.s., we obtain, once again by Lemma 2.1 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 that  $(\varepsilon_n)$  is a martingale difference sequence such that  $E[\varepsilon_{n+1}^2 \mid \mathcal{F}_n] = \sigma^2$  a.s. for all  $n \geq 1$ , and let

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

If  $(\varepsilon_n)$  has a conditional moment of order  $a > 2$  then the strong law of large numbers for martingales implies the almost sure convergence of  $\Delta_n$  to  $\sigma^2$ . Under the assumptions of Theorem 2.1 with  $p = 1$ , convergence (3.6) leads to the strong consistency of the estimator of  $\sigma^2$ ,

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

since

$$\lim_{n \rightarrow \infty} \frac{n}{\log d_n} (\Gamma_n - \Delta_n) = \sigma^2 \quad \text{a.s.} \tag{3.7}$$

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

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

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

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

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

**Corollary 3.2.** *Assume that  $(\varepsilon_n)$  satisfies (2.1) with  $p \geq 1$ . In addition, suppose that, for some integer  $1 \leq q \leq p$ ,  $E[\varepsilon_{n+1}^{2q} | \mathcal{F}_n] = \sigma(2q)$  a.s. Then,  $\Gamma_n(2q)$  is a strongly consistent estimator of  $\sigma(2q)$  with*

$$(\Gamma_n(2q) - \Delta_n(2q))^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \quad (3.8)$$

*Proof.* We have already shown via (3.7) that Corollary 3.2 holds for  $q = 1$ . Now assume that  $q \geq 2$ . By expanding the expression of  $\Gamma_n(2q)$ , equality (3.4) leads to

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

Using the almost sure convergence of  $f_n$  to 0 and Corollary 3.1, we deduce from (3.5) 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  $\sigma_n(\ell) = E[\varepsilon_{n+1}^\ell | \mathcal{F}_n]$  and  $e_{n+1}(\ell) = \varepsilon_{n+1}^\ell - \sigma_n(\ell)$ . First, since the conditional moments  $\sigma_n(\ell)$  are a.s. bounded, we have

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

Moreover, from the estimate

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

we obtain

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

which concludes the proof of Corollary 3.2.

**Remark 3.1.** It is now possible to deduce from Corollary 3.2 the asymptotic behaviour of  $(C_n(q))$ . Under the assumptions of Corollary 3.2, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} C_n(q) = \sigma(2q) \quad \text{a.s.}$$

Moreover, since the conditional moment of order  $a > 2p$  of  $(\varepsilon_n)$  is a.s. finite, Chow's lemma leads to

$$\left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k^{2q} - \sigma(2q) \right| = o(n^{c-1}) \quad \text{a.s.} \quad (3.9)$$

for all  $c$  such that  $2pa^{-1} < c < 1$ . Hence, it follows from (3.8) and (3.9) 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 stating 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, we have*

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \sum_{k=1}^n f_k((\hat{\theta}_k - \theta)^\top S_k(\hat{\theta}_k - \theta))^p = \ell(p) \quad \text{a.s.} \quad (3.10)$$

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

$$\lim_{n \rightarrow +\infty} \frac{1}{n} S_n = L \quad \text{a.s.} \quad (3.11)$$

Then, we have

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n k^{p-1} ((\hat{\theta}_k - \theta)^\top L(\hat{\theta}_k - \theta))^p = \ell(p) \quad \text{a.s.} \quad (3.12)$$

**Remark 3.2.** Since  $L$  is positive definite, (3.12) immediately yields

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

*Proof of Corollary 3.3.* From the definitions of  $S_n$  and  $\hat{\theta}_n$ , it is easy to see that

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

Hence, it follows from convergence (2.3) and from the convergence of the explosion coefficient  $(f_n)$  to 0, together with Kronecker's lemma, that, a.s.,

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

Then, convergence (3.10) is a straightforward consequence of Theorem 2.1. In addition, it is not difficult to see that

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

The classical Abel transform gives the decomposition

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

where

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

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 2.1 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, i.e.

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

and using the proof of Theorem 2.1 together with (2.10), we obtain

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

The second term is a.s.  $o(\log n)$ . This is a mere consequence of Lemma 2.1 and of the almost sure convergence of  $(\Sigma_n - nd)/\beta_n$  to 0. Finally, we have

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

We shall now apply 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

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

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  $(\varepsilon_n)$  is a

martingale difference sequence which satisfies (2.1) with  $p \geq 1$ . If  $E[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$  a.s. and

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

then convergence (3.11) holds with the limiting matrix  $L$  given by

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

Moreover, we can immediately see that the matrix  $L$  is positive definite [8], [9]. Then we are in position to state our following result.

**Corollary 3.4.** *Assume that  $(\varepsilon_n)$  satisfies (2.1) with  $p \geq 1$ . In addition, suppose that, for some integer  $1 \leq q \leq p$ ,  $E[\varepsilon_{n+1}^{2q} | \mathcal{F}_n] = \sigma(2q)$  a.s. Then,  $\Gamma_n(2q)$  is a strongly consistent estimator of  $\sigma(2q)$  with*

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

Moreover, we also have

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

### 3.4. A branching process with immigration

3.4.1. *Estimation of means.* Branching processes with immigration play an increasingly important role in statistical physics, computational biology, and evolutionary theory. The concept of immigration is related to the 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

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k} + I_{n+1}, \tag{3.13}$$

where  $(Y_{n,k})$  and  $(I_n)$  are two independent sequences of independent, identically distributed, 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

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

We are interested in the estimation of all the parameters  $m$ ,  $\lambda$ ,  $\sigma^2$ , and  $b^2$ . Relation (3.13) may be rewritten in the autoregressive form

$$X_{n+1} = mX_n + \lambda + \varepsilon_{n+1}, \tag{3.14}$$

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  $\Phi_n^\top = (X_n, 1)$  and



$\theta^\top = (m, \lambda)$ . However, we can observe that the situation is a little more tricky as  $(\varepsilon_n)$  is a martingale difference sequence with unbounded conditional variance

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

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

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

where the random variables  $Z_{n+1}$ ,  $\Psi_n$ , and  $\xi_{n+1}$  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,  $(\xi_n)$  is a martingale difference sequence with bounded conditional variance  $E[\xi_{n+1}^2 \mid \mathcal{F}_n] \leq \sigma^2 + b^2$  a.s. We estimate the vector of means  $\theta^\top = (m, \lambda)$  by the conditional least-squares estimator

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

where

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

In the subcritical case,  $m < 1$ , Wei and Winnicki [19] showed 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

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

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  $(\varepsilon_n)$  satisfies (2.1) with  $p \geq 1$ . In addition, suppose that, for some integer  $1 \leq q \leq p$ ,  $E[\varepsilon_{n+1}^{2q} \mid \mathcal{F}_n] = \sigma(2q)$  a.s. Then,  $\Gamma_n(2q)$  is a strongly consistent estimator of  $\sigma(2q)$  with*

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

Moreover, we also have

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

3.4.2. *Estimation of variances.* It follows from (3.14) that

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

where  $(V_n)$  is a martingale difference sequence satisfying

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

where

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

Consequently, we infer that

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

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

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

where

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

In the subcritical case,  $m < 1$ , it was established in [20] that

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

where  $\Lambda$  is the positive definite limiting matrix given by

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

Our last result is as follows.

**Corollary 3.6.** *Assume that  $(\varepsilon_n)$  satisfies (2.1) for  $p \geq 2$ . Then*

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

### References

- [1] BERCU, B. (2004). On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. *Stoch. Process. Appl.* **111**, 157–173.
- [2] BERCU, B. AND FORT, J.-C. (2008). A moment approach for the almost sure central limit theorem for martingales. *Studia Sci. Math. Hung.* **45**, 139–159.
- [3] BROSAMLER, G. A. (1988). An almost everywhere central limit theorem. *Math. Proc. Camb. Phil. Soc.* **104**, 561–574.
- [4] CHAËBANE, F. (2001). Invariance principles with logarithmic averaging for martingales. *Studia Sci. Math. Hung.* **37**, 21–52.

- [5] CHAÂBANE, F. AND MAÂOUIA, F. (2000). Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Prob. Statist.* **4**, 137–189.
- [6] CHAÂBANE, F. AND TOUATI, A. (2001). On averaging methods for identification of linear regression models. *C. R. Acad. Sci. Paris* **333**, 133–138.
- [7] CHAÂBANE, F., MAÂOUIA, F. AND TOUATI, A. (1998). Généralisation du théorème de la limite centrale presque-sûr pour les martingales vectorielles. *C. R. Acad. Sci. Paris* **326**, 229–232.
- [8] DUFLO, M. (1997). *Random Iterative Methods* (Appl. Math. **34**). Springer, Berlin.
- [9] DUFLO, M., SENOSSI, R. AND TOUATI, A. (1991). Propriétés asymptotiques presque sûres de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel. *Ann. Inst. H. Poincaré Prob. Statist.* **27**, 1–25.
- [10] GOODWIN, G. AND SIN, K. (1984). *Adaptive Filtering Prediction and Control*. Prentice-Hall, Englewood Cliffs, NJ.
- [11] LACEY, M. T. AND PHILLIP, W. (1990). A note on the almost sure central limit theorem. *Statist. Prob. Lett.* **9**, 201–205.
- [12] LAI, T. L. AND WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10**, 154–166.
- [13] LAI, T. L. AND WEI, C. Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.* **13**, 1–23.
- [14] LIFSHITS, M. A. (2001). Lecture notes on almost sure limit theorems. *Publ. IRMA* **54**, 1–25.
- [15] LIFSHITS, M. A. (2002). Almost sure limit theorem for martingales. In *Limit Theorems in Probability and Statistics*, Vol. II, János Bolyai Mathematical Society, Budapest, pp. 367–390.
- [16] SCHATTE, P. (1988). On strong versions of the central limit theorem. *Math. Nachr.* **137**, 249–256.
- [17] WEI, C. Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *Ann. Statist.* **13**, 1498–1508.
- [18] WEI, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* **15**, 1667–1682.
- [19] WEI, C. Z. AND WINNICKI, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.* **18**, 1757–1773.
- [20] WINNICKI, J. (1991). Estimation of the variances in the branching process with immigration. *Prob. Theory Relat. Fields* **88**, 77–106.