Stochastic Processes and their Applications 26 (1987) 73-85
North-Holland

ON THE STRONG LAW OF LARGE NUMBERS FOR MULTIVARIATE MARTINGALES

Heinz KAUFMANN
Lehrstuhl für Statistik, Universität Regensburg, Universitätsstr. 31, D-8400 Regensburg, FR Germany

Received 26 March 1987

The strong law of large numbers is considered for a multivariate martingale normed by a sequence of square matrices. In particular multivariate martingale extensions of the strong laws of Kolmogorov and Marcinkiewicz-Zygmund are presented. Convergence to zero in L_α is obtained under the same conditions. Norming by powers of the covariance matrix is considered in detail. Results are further used to derive conditions for strong consistency of the least squares estimator in linear regression with multivariate responses. These conditions do not necessarily assume square integrability of errors. They become particularly simple for polynomially bounded regressors. Two examples are treated, including polynomial regression.

AMS 1980 Subject Classifications: Primary 60G42; Secondary 62H12.

multivariate martingales * strong law of large numbers * multivariate linear regression * least squares * strong consistency

1. Introduction

In this paper, the strong law of large numbers is considered for a multivariate martingale $\{s_n, \mathcal{A}_n, n \geq 0\}$, i.e. a sequence of random vectors of dimension p, $1 \leq p < \infty$, satisfying the martingale properties. Additionally, we assume $s_0 = 0$. More specifically, we consider the normed martingale $\{B_n s_n, n \geq 1\}$, where B_n is an $\mathcal{A}_{n-1}$-measurable $p \times p$-matrix, and ask for conditions under which

$$B_n s_n \to 0 \quad \text{almost surely.} \tag{1.1}$$

Typically, (1.1) is needed in proving strong consistency of some estimator of a vector parameter. Indeed, most results on (1.1) have been given for the least squares estimator in the univariate linear regression model

$$y_n = z_n'\beta + \varepsilon_n, \quad n \geq 1,$$

where B_n resp. s_n are of the special form

$$B_n = \left(\sum_{i=1}^n z_i z_i'\right)^{-1} \qquad s_n = \sum_{i=1}^n z_i \varepsilon_i.$$

The weakest conditions in this direction have been presented by Lai, Robbins and Wei [11] and Lai and Wei [12]. For instance, if the regressors $\{z_n\}$ are nonrandom and the errors $\{\varepsilon_n\}$ form a martingale difference sequence with uniformly bounded second moments, then (1.1) is implied by the minimal assumption $B_n \to 0$, see [11, Corollary 5].

However, there are also interesting situations, for instance in the theory of maximum likelihood estimation, where the consistency question can be reduced to (1.1), without the special form of B_n resp. s_n treated by those authors. In the present paper, we take a totally different approach. Some comparison of those results to ours will be given in Section 5.

In the sequel, we will use the following notation. The martingale difference sequence corresponding to $\{s_n\}$ will be denoted by

$$x_n = s_n - s_{n-1}, \quad n \geq 1.$$

For any square matrix A, we write $|A|$ for the determinant of A and $\lambda_{\max} A$ resp. $\lambda_{\min} A$ for the largest resp. smallest eigenvalue of A, if A is additionally symmetric. The symbol $\|x\|$ resp. $\|A\|$ will be reserved for the Euclidean vector norm $(x'x)^{1/2}$ resp. the corresponding matrix norm $(\lambda_{\max} A'A)^{1/2}$. If A is positive semidefinite, then this norm reduces to $\|A\| = \lambda_{\max} A$.

In analogy to the scalar case (see the monographs of Hall and Heyde [9] and Stout [14]), (1.1) should be induced by some convergence condition relating $\{B_n\}$ and moments of $\{s_n\}$, together with some kind of monotone convergence of $\{B_n\}$ to zero. By componentwise application of the martingale convergence theorem it is not difficult to obtain conditions assuring the a.s. convergence of $\sum B_i x_i$. Some kind of multivariate Kronecker lemma would then give $B_n s_n \to 0$ a.s. However, the multivariate versions of the Kronecker lemma known so far lead to conditions which are very strong for $p > 1$, namely $\lambda_{\max} B_n = O(\lambda_{\min} B_n)$ (Anderson and Taylor [3], Anderson and Moore [1]) or B_n diagonal.

A more sensible recourse to the martingale convergence theorem can be based on the series

$$z_n(\alpha) = \sum_{i=1}^{n} (\|B_i s_i\|^\alpha - \|B_i s_{i-1}\|^\alpha), \quad 1 \leq \alpha < \infty,$$

respectively

$$w_n(\alpha) = \sum_{i=1}^{n} \|B_i x_i\|^\alpha, \quad 1 \leq \alpha \leq 2,$$

under the monotonicity condition

$$\|B_n s\| \geq \|B_{n+1} s\| \quad \text{for all } s \in R^p, \ n \geq 1. \tag{1.2}$$

This approach will be studied in Sections 2 and 3. Let us first discuss (1.2). It is equivalent to the requirement that the difference $B_n' B_n - B_{n+1}' B_{n+1}$ is positive semidefinite. If (1.2) does not hold for some sequence $\{B_n\}$ of interest, there may be related sequences for which it holds, e.g. $\{\|B_n\|^{1/2} B_n^{1/2}\}$ if $B_n - B_{n-1}$ is positive semidefinite. Such a replacement typically leads to conditions slightly stronger than $B_n \to 0$, as illustrated in Sections 4 and 5.

The series $z_n(\alpha)$, $n \geq 1$, is a multivariate analogue of the series appearing in the martingale generalization of the Hájek and Rényi [8] inequality given by Chow [4]. In fact, it is straightforward to extend this result to the present situation. This could be used to obtain a.s. results. In Section 2, we go another route and use the martingale convergence theorem to demonstrate that $\sup_n Ez_n(\alpha) < \infty$ for some $\alpha \geq 1$ implies that $\{\|B_n s_n\|\}$ converges a.s., with the further consequence that $B_n s_n \to 0$ a.s. and in probability are equivalent. If $B_n s_n \to 0$ in probability holds, this gives the desired result.

The series $Ew_n(\alpha)$, $n \geq 1$, is a multivariate analogue of series common in the theory of a.s. convergence. It leads to conditions of the Kolmogorov and Marcinkiewicz-Zygmund type. If $\alpha = 2$, then $Ez_n(\alpha) = Ew_n(\alpha)$ is valid, due to properties of square integrable martingales. In the remainder of Section 2, it is shown that $Ew_n(\alpha)$ dominates $Ez_n(\alpha)$ for $1 \leq \alpha \leq 2$. This is used to obtain a corollary with $\sup_n Ew_n(\alpha) < \infty$ assuring a.s. convergence of $\{\|B_n s_n\|\}$.

Sometimes $B_n s_n \to 0$ in probability is easily demonstrated, and results of Section 2 can then be used to conclude $B_n s_n \to 0$ a.e. Nevertheless, it is of interest to have general conditions assuring $B_n s_n \to 0$ a.s. For a nonrandom sequence $\{B_n\}$, such conditions are given in Section 3. Actually, multivariate martingale generalizations of the Kolmogorov [10] and Marcinkiewicz-Zygmund [13] strong laws are presented. Convergence in L_α is obtained as a by-product.

Sections 4 and 5 are independent of each other. Extending the scalar case, it seems natural to norm by powers of the covariance matrix of s_n. Section 4 deals with this norming. In Section 5 strong consistency of the least squares estimator is considered for a fairly general class of linear regression models for multivariate responses. This class includes the more familiar models described by Anderson [2]. Conditions are given not necessarily assuming square integrability of errors, and particular attention is paid to polynomially bounded regressors. Two examples are treated, including polynomial regression.

In Section 6, we conclude with some remarks of limitations and possible extensions of the approach presented in this paper.

2. Convergence of $\{\|B_n s_n\|\}$

In this section, we suppose without further mentioning that the sequence $\{B_n\}$ is decreasing in the sense (1.2). First we use this assumption to show the following martingale properties of the series $z_n(\alpha)$ essential in the proof of Theorem 1.

Lemma 1. *Let $1 \leq \alpha < \infty$. If $E\|B_n s_n\|^\alpha$ is finite for all $n \geq 1$, then $\{z_n(\alpha)\}$ is a nonnegative submartingale.*

Remark. Under (1.2), it can be shown that the assumption

$$E\|B_n s_n\|^\alpha < \infty, \quad n \geq 1,$$

is equivalent to

$$E\|B_nx_n\|^{\alpha}<\infty, \quad n\geq 1.$$

Proof. By partial summation, $z_n(\alpha)$ transforms into

$$z_n(\alpha)=\|B_ns_n\|^{\alpha}+\sum_{i=1}^{n-1}(\|B_is_i\|^{\alpha}-\|B_{i+1}s_i\|^{\alpha}). \tag{2.1}$$

Due to this equation and (1.2), $z_n(\alpha)$ is nonnegative. Further, (2.1) implies that $Ez_n(\alpha)<\infty$ for all $n\geq 1$ if and only if $E\|B_ns_n\|^{\alpha}<\infty$ for all $n\geq 1$. It remains to prove that

$$E(\|B_ns_n\|^{\alpha}\,|\,\mathscr{A}_{n-1})-\|B_ns_{n-1}\|^{\alpha}\geq 0, \quad n\geq 1.$$

This follows from the conditional Jensen inequality, since for $\alpha\geq 1$ the function $\|Bs\|^{\alpha}$ is a convex function of s for any matrix B. □

Theorem 1. *Let* $1\leq\alpha<\infty$ *and*

$$\sup_n Ez_n(\alpha)<\infty. \tag{2.2}$$

Then $\{\|B_ns_n\|\}$ *converges almost surely. Therefore* $B_ns_n\to 0$ *almost surely and in probability are equivalent.*

Proof. Under (2.2), the series $z_n(\alpha)$ converges a.s., by the martingale convergence theorem. Moreover, the remainder in (2.1), namely the series

$$\sum_{i=1}^{n-1}(\|B_is_i\|^{\alpha}-\|B_{i+1}s_i\|^{\alpha}),$$

is a.s. bounded above by $\sup_n z_n(\alpha)$. Since it is nondecreasing, it converges a.s. If both series converge, then also their difference and hence $\{\|B_ns_n\|\}$ converges a.s.

Since $B_ns_n\to 0$ is equivalent to $\|B_ns_n\|\to 0$ and the a.s. limit can be identified from the probability limit, $B_ns_n\to 0$ a.s. and in probability are equivalent. □

Now we turn to the question whether $Ew_n(\alpha)$ dominates $Ez_n(\alpha)$, $1\leq\alpha\leq 2$. The link between the two series is furnished by the following inequality.

Lemma 2. *Let* x *and* y *be p-vectors and* $1<\alpha\leq 2$. *Setting* $x'y\|y\|^{\alpha-2}=0$ *if* $y=0$, *the inequality*

$$\|x+y\|^{\alpha}-\|y\|^{\alpha}\leq 2^{2-\alpha}\|x\|^{\alpha}+\alpha x'y\|y\|^{\alpha-2} \tag{2.3}$$

holds, with equality if $\alpha=2$.

Proof. The cases $\alpha=2$ or $y=0$ can easily be checked directly. Let $1<\alpha<2$ and $y\neq 0$ in the sequel. Defining $\varphi=(x'x/y'y)^{1/2}$ and $\theta=x'y/y'y$, the inequality (2.3) can equivalently be transformed into

$$h(\varphi,\theta)=(\varphi^2+2\theta+1)^{\alpha/2}-1-2^{2-\alpha}\varphi^\alpha-\alpha\theta\leqslant 0, \tag{2.4}$$

where $\varphi\geqslant 0$ and $|\theta|\leqslant\varphi$, due to the Cauchy-Schwarz inequality.

For any fixed φ, set $g(\theta)=h(\varphi,\theta)$. This function is well defined and strictly concave on $\{\varphi^2+2\theta+1\geqslant 0\}$, which contains $\{|\theta|\leqslant\varphi\}$. On the larger set, $g(\theta)$ is maximized by $\theta_0=-\varphi^2/2$. If $\varphi\leqslant 2$, then θ_0 lies within $\{|\theta|\leqslant\varphi\}$, and it is easily checked that $g(\theta_0)\leqslant 0$. This implies $h(\varphi,\theta)\leqslant 0$ on $\{0\leqslant|\theta|\leqslant\varphi\leqslant 2\}$.

If $\varphi>2$, then θ_0 lies on the left of $\{|\theta|\leqslant\varphi\}$, and on this set, $g(\theta)$ is maximized by the left endpoint $\theta=-\varphi$, due to concavity. Inserting into (2.4), it remains to be shown that

$$h(\varphi,-\varphi)=(\varphi-1)^\alpha-1-2^{2-\alpha}\varphi^\alpha+\alpha\varphi\leqslant 0$$

on $\{\varphi>2\}$. By checking the derivatives of $h(\varphi,-\varphi)$, this function is found to be concave on $\{\varphi\geqslant 2\}$, with the maximum $2\alpha-4\leqslant 0$ obtained at $\varphi=2$. □

Corollary 1. *Let $1\leqslant\alpha\leqslant 2$. Then the inequality*

$$Ez_n(\alpha)\leqslant 2^{2-\alpha}Ew_n(\alpha),\quad n\geqslant 1, \tag{2.5}$$

holds. Consequently, if

$$\sup_n Ew_n(\alpha)<\infty,$$

then the conclusions of Theorem 1 hold.

Proof. If $\alpha=1$, (2.5) is a consequence of the triangle inequality. Let $1<\alpha\leqslant 2$ and assume without loss of generality that

$$E\|B_nx_n\|^\alpha<\infty,\quad n\geqslant 1.$$

If x and y of Lemma 2 are random vectors with $E\|x\|^\alpha<\infty$ and $E\|y\|^\alpha<\infty$, then the Hölder inequality implies that $Ex'y\|y\|^{\alpha-2}$ is finite. With $x=B_ix_i$, $y=B_is_{i-1}$, we can form expectations on both sides of (2.3) conditional on $\mathscr{A}_{i-1}$. Due to the martingale property, the mixed term on the right vanishes. By integration and summation (2.5) follows. □

3. Laws of large numbers

If the conditions of Theorem 1 hold for some $\alpha\geqslant 1$, one may suppose that $B_n\to 0$ a.s. is sufficient for $B_ns_n\to 0$ a.s. In general it is not clear whether this claim holds. However, it can be proved under the stronger conditions of Corollary 1, if $\{B_n\}$ is a nonrandom sequence. The following lemma on summability replaces use of the Kronecker lemma.

Lemma 3. *Let* $\{a_{ij}, 1 \leq i \leq j\}$ *be an infinite triangular array of real numbers such that*

$$\lim_{j\to\infty} a_{ij} = 0 \quad \textit{for any fixed } i \geq 1, \tag{3.1}$$

$$\sup_n \sum_{i=1}^{n} \sum_{j=i}^{n} |a_{ij} - a_{i,j+1}| < \infty. \tag{3.2}$$

Then

$$\lim_{j\to\infty} \sum_{i=1}^{j} a_{ij} = 0. \tag{3.3}$$

Proof. By partial summation,

$$\sum_{i=1}^{n} a_{in} = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n-1} \sum_{j=i}^{n-1} (a_{i,j+1} - a_{ij}). \tag{3.4}$$

Due to (3.2) and the theorem on rearrangement of double series, the rightmost sum has a limit which equals

$$\lim_{n\to\infty} \sum_{i=1}^{n} \sum_{j=i}^{\infty} (a_{i,j+1} - a_{ij}).$$

By (3.1),

$$\sum_{j=i}^{\infty} (a_{i,j+1} - a_{ij}) = -a_{ii}.$$

Hence $\lim_{n\to\infty} \sum_{i=1}^{n} a_{ii}$ exists. Inserting into (3.4), (3.3) follows. □

Now we can get the following multivariate martingale generalization of the strong laws of large numbers of Kolmogorov [10] and Marcinkiewicz and Zygmund [13]. Utilizing Lemmas 2 and 3, it is first shown that the conditions of Corollary 1 imply $B_n s_n \to 0$ in L_α. The a.s. result then follows from Corollary 1.

Theorem 2. *Let* $1 \leq \alpha \leq 2$ *and assume that*

$$\sup_n Ew_n(\alpha) < \infty. \tag{3.5}$$

If $\{B_n\}$ *is a nonrandom sequence with the monotonicity property* (1.2), *then* $B_n \to 0$ *implies*

$$B_n s_n \to 0 \quad \textit{almost surely and in } L_\alpha.$$

Proof. Apply Lemma 3 with

$$a_{ij} = E\|B_j x_i\|^\alpha, \quad 1 \leq i \leq j.$$

Due to (1.2), for fixed i the sequence $\{\|B_j x_i\|^\alpha, j \geq i\}$ is monotonically decreasing. Since $B_j \to 0$, we have $\|B_j x_i\|^\alpha \to 0$ a.s., for any fixed i. Due to (3.5), $E\|B_i x_i\|^\alpha < \infty$ holds and the monotone convergence theorem implies $E\|B_j x_i\|^\alpha \to 0$, i.e. (3.1).

Further, (1.2) implies that $a_{ij} - a_{i,j+1} \geq 0$, whence

$$\sup_n \sum_{i=1}^n \sum_{j=1}^n |a_{ij} - a_{i,j+1}| \leq \sup_n \sum_{i=1}^n a_{ii}.$$

Due to (3.5), the right hand expression is finite, and (3.2) is verified. From Lemma 3, we obtain

$$\sum_{i=1}^n E\|B_n x_i\|^\alpha \to 0. \tag{3.6}$$

Since B_n is nonrandom, we can apply Lemma 2. As in Corollary 1, it follows that

$$E\|B_n s_n\|^\alpha \leq 2^{2-\alpha} \sum_{i=1}^n E\|B_n x_i\|^\alpha, \quad n \geq 1. \tag{3.7}$$

Statements (3.6) and (3.7) together imply that $\|B_n s_n\| \to 0$ in L_α, and Corollary 1 gives the a.s. result. □

If second moments exist, analogy to scalar theorems becomes even more apparent in the following corollary.

Corollary 2. *Let the martingale* $s_n = \sum_{i=1}^n x_i$, $n \geq 1$, *be square integrable with* $\operatorname{cov} x_n = \Sigma_n$, *say. If* $\{B_n\}$ *is a nonrandom sequence satisfying* (1.2) *and* $B_n \to 0$, *then*

$$\sum_{n=1}^\infty \|B_n \Sigma_n B_n'\| < \infty \tag{3.8}$$

is sufficient for $B_n s_n \to 0$ *almost surely and in* L_2.

Remark. Since the matrices $B_n \Sigma_n B_n'$ are positive semidefinite, (3.8) is equivalent to $\sum_{n=1}^\infty \operatorname{trace} B_n \Sigma_n B_n' < \infty$, and this holds if and only if $\sum_{i=1}^n B_i \Sigma_i B_i'$ converges to a matrix with all elements finite.

Proof. First we consider the remark for a general sequence $\{A_n\}$ of positive semidefinite $p \times p$-matrices. Since $\|A_n\| = \lambda_{\max} A_n$ and all eigenvalues of A_n are nonnegative, $\|A_n\| \leq \operatorname{trace} A_n \leq p\|A_n\|$. Hence, if one of the series $\sum \|A_n\|$ or $\sum \operatorname{trace} A_n$ converges, then so does the other, by the monotonicity criterion. If $\sum_{i=1}^n A_i$ converges elementwise, clearly $\sum \operatorname{trace} A_n < \infty$. Conversely, if $\sum \operatorname{trace} A_n < \infty$, then the sequence $x' \sum_{i=1}^n A_i x$ converges for any fixed x, since it is nondecreasing and bounded above by $x'x \sum \operatorname{trace} A_n$. Choosing $x = (1, 0, \ldots, 0)', \ldots, (0, \ldots, 0, 1)'$ gives convergence of the diagonal elements. Convergence of the (i, j)-element is then obtained with the choice $x_i = x_j = 1$, and zero for the other components of x.

Applying the remark, it is easy to see that (3.5) and (3.8) are equivalent, and this finishes the proof. □

4. Norming by powers of the covariance matrix

If $\{s_n\}$ is a square integrable martingale with cov $s_n = F_n$, say, the a.s. behaviour of $F_n^{-\alpha} s_n$, $\alpha \geq \frac{1}{2}$, is of particular interest. Recall that for any positive definite matrix A with diagonalisation $A = P\Lambda P'$, $PP' = I$, Λ diagonal, powers are defined by $A^\alpha = P\Lambda^\alpha P'$, $\alpha \in \mathbb{R}$.

Results for $\alpha > \frac{1}{2}$ will be consequences of the result for $\alpha = \frac{1}{2}$. In estimating the magnitude of the standardized vector $F_n^{-1/2} s_n$, the following lemma is helpful. It is a generalization of Lemma 2(i) of Lai and Wei [12], who consider $r = 1$.

Lemma 4. *Let B and C be positive semidefinite matrices with rank $C = r$, say, and $A = B + C$ positive definite. Then*

$$1 - |B|/|A| \leq \operatorname{trace}(A^{-1/2} C A^{-1/2}) \leq r(1 - |B|/|A|). \tag{4.1}$$

Proof. Since $A^{-1/2} B A^{-1/2}$ is a symmetric matrix, there exists an orthogonal transformation P such that $B_* = TBT'$, with $T = PA^{-1/2}$, is a diagonal matrix $\operatorname{diag}(b_1, \ldots, b_p)$, say. Then

$$A_* = TAT' = I \quad \text{and} \quad C_* = TCT' = \operatorname{diag}(1 - b_1, \ldots, 1 - b_p).$$

Noting

$$|B|/|A| = |B_*| = \prod_1^p b_j$$

and

$$\operatorname{trace} A^{-1/2} C A^{-1/2} = \operatorname{trace} C_* = p - \sum_1^p b_j,$$

assertion (4.1) becomes

$$1 - \prod_1^p b_j \leq p - \sum_1^p b_j \leq r\left(1 - \prod_1^p b_j\right). \tag{4.2}$$

Due to the assumptions on B and C, $0 \leq b_j \leq 1$, $j = 1, \ldots, p$. By induction, it can be proved that the left hand inequality holds for any p such numbers. Due to rank $C = r$, exactly $p - r$ elements of $b_1, \ldots, b_p$ equal one, and r elements are less than one, say $b_1, \ldots, b_r$. Using this fact, the right hand inequality of (4.2) can be transformed into

$$\sum_1^r b_j \geq r \prod_1^r b_j, \tag{4.3}$$

where $0 \leq b_j < 1$, $j = 1, \ldots, r$. This inequality holds trivially if $r = 0$. Otherwise, inserting $z = \prod_1^r b_j$ into $z \leq z^{1/r}$ which holds if $z \leq 1$, $r \geq 1$, (4.3) follows from the well known inequality between arithmetic and geometric mean. □

If $\|F_n\| \to \infty$, the first part of Theorem 3 states that the magnitude of $F_n^{-1/2} s_n$ is $o((\log\|F_n\|)^\delta)$ a.s., for any $\delta > \frac{1}{2}$. This is well known in the scalar case. The second part gives a result on $F_n^{-\alpha} s_n$, $\alpha > \frac{1}{2}$. It follows easily from the first part, similarly to Theorem 1 of Lai and Wei [11] following from their Lemma 1.

Theorem 3. *Let the martingale $\{s_n\}$ be square integrable with $F_n = \operatorname{cov} s_n$ positive definite for $n \geq 1$.*

(i) *If $\|F_n\| \to \infty$, then, for any $\delta > \frac{1}{2}$,*

$$F_n^{-1/2} s_n = o((\log\|F_n\|)^\delta) \quad \text{a.s. } \textit{and in } L_2. \tag{4.4}$$

(ii) *If $\alpha > \frac{1}{2}$, then*

$$(\log\|F_n\|)^\gamma = o(\lambda_{\min} F_n) \quad \textit{for some } \gamma > (2\alpha - 1)^{-1} \tag{4.5}$$

implies $F_n^{-\alpha} s_n \to 0$ a.s. and in L_2.

Proof. (i) Set

$$B_n = (\log\|F_n\|)^{-\delta} F_n^{-1/2}, \quad n \geq 1.$$

The fact that $F_n - F_{n-1}$ is positive semidefinite implies that $\{B_n\}$ is decreasing in the sense (1.2). The assumption $\|F_n\| \to \infty$ implies $B_n \to 0$ and $|F_n| \to \infty$. Noting that $\log|F_n| \leq p \log\|F_n\|$ and using Lemma 4 with $A = F_n$, $B = F_{n-1}$, (3.8) follows. Hence Corollary 2 applies and yields the result.

(ii) If (4.5) holds, then $\lambda_{\min} F_n \to \infty$ and $\|F_n\| \to \infty$. Further, (4.5) and $1 - 2\alpha < 0$ imply that, for any $\varepsilon > 0$ and sufficiently large n,

$$\begin{aligned} \|F_n^{-\alpha} s_n\| &\leq (\lambda_{\min} F_n)^{(1-2\alpha)/2} \|F_n^{-1/2} s_n\| \\ &\leq \varepsilon (\log\|F_n\|)^{-\delta} \|F_n^{-1/2} s_n\|, \end{aligned}$$

where $\delta = (2\alpha - 1)\gamma/2 > \frac{1}{2}$. Hence (ii) follows from (i). □

Remark. In part (ii), it seems tempting to set $B_n = F_n^{-\alpha}$ and to replace (4.5) by the weaker assumption $\lambda_{\min} F_n \to \infty$. The convergence condition (3.8) can be demonstrated under this assumption in the important case $\alpha = 1$, for example. However, for $B_n = F_n^{-\alpha}$ with $\alpha > \frac{1}{2}$ the monotonicity condition (1.2) may fail.

5. Strong consistency in multivariate linear regression

In this section we consider the linear regression model

$$y_n = Z_n' \beta + \varepsilon_n, \quad n = 1, 2, \ldots, \tag{5.1}$$

where y_n is the observed q-dimensional response, $q \geq 1$, Z_n is a nonrandom $p \times q$ regressor matrix and β an unknown p-vector of parameters. The unobservable errors $\{\varepsilon_n\}$ are supposed to form a q-dimensional martingale difference sequence. Setting

$$Z'_n = \begin{bmatrix} z'_n & 0 & \cdots & 0 \\ 0 & z'_n & & 0 \\ 0 & \cdots & 0 & z'_n \end{bmatrix}, \qquad \beta' = (\beta'_1, \ldots, \beta'_q)$$

and assuming that the error vectors are i.i.d. and square integrable, it follows that the model of Anderson [2] is a particular case.

The least squares estimator from the first n observations, i.e. the estimator minimizing $\sum_{i=1}^n \|y_i - Z'_i\beta\|^2$, will be denoted by $\hat{\beta}_n$. With $s_n = \sum_{i=1}^n Z_i\varepsilon_i$, $V_n = \sum_{i=1}^n Z_iZ'_i$ nonsingular for $n \geq N$, say, it satisfies $\hat{\beta}_n - \beta = V_n^{-1}s_n$. Hence it is consistent if and only if $V_n^{-1}s_n \to 0$ a.s.

For nonrandom regressors, if the responses are univariate ($q = 1$) and $\sup_n E\varepsilon_n^2 < \infty$, then $\lambda_{\min} V_n \to \infty$ or equivalently $V_n^{-1} \to 0$ implies strong consistency, see Lai, Robbins and Wei [11]. Actually, those authors prove this statement for more general error sequences than martingale difference errors with uniformly bounded second moments. With the methods of the present paper, only a somewhat weaker assertion can be obtained, namely that $\sup_n E\varepsilon_n^2 < \infty$ and

$$(\log\|V_n\|)^\gamma = o(\lambda_{\min} V_n) \quad \text{for some } \gamma > 1 \tag{5.2}$$

is sufficient for strong consistency. Similar conditions are given in Lai und Wei [12] for $q = 1$ and random regressors Z_n supposed to be $\mathscr{A}_{n-1}$-measurable. However, we can get theorems where the assumptions on the error sequence are weakened in another direction, namely to martingale difference errors where some moment of order α, $1 \leq \alpha \leq 2$, is uniformly bounded, and results are easily extended to multivariate responses.

Theorem 4. *In the regression model* (5.1), *assume that* $\sup_n E\|\varepsilon_n\|^\alpha < \infty$ *holds for some* α, $1 \leq \alpha \leq 2$. *Then*

$$\sum_{n=N}^{\infty} (\|Z'_n V_n^{-1} Z_n\| / \lambda_{\min} V_n)^{\alpha/2} < \infty \tag{5.3}$$

implies that $\hat{\beta}_n \to \beta$ *a.s. and in* L_α.

Proof. The sequence

$$B_n = (\lambda_{\min} V_n)^{-1/2} V_n^{-1/2}, \quad n \geq N,$$

is decreasing in the sense (1.2), and

$$\|V_n^{-1}s_n\| \leq \|B_n s_n\|, \quad n \geq N.$$

With $s = \sup_n E\|\varepsilon_n\|^\alpha < \infty$, it is easy to see that

$$E\|B_n Z_n \varepsilon_n\|^\alpha \leq s(\|Z'_n V_n^{-1} Z_n\| / \lambda_{\min} V_n)^{\alpha/2}, \quad n \geq N.$$

Hence (5.3) implies (3.5), and Theorem 2 applies. □

In the first part of the following corollary we consider the practically important case of polynomially bounded regressor matrices. This restriction leads to a simple minimum growth rate for $\{\lambda_{\min} V_n\}$. The second part focuses on square integrable errors.

Corollary 3. *In the regression model* (5.1), *assume that* $\sup_n E\|\varepsilon_n\|^\alpha < \infty$ *holds for some* α, $1 \leqslant \alpha \leqslant 2$.

(i) *If* $\|Z_n\| = O(n^\delta)$ *for some* $\delta > 0$, *then*

$$n^{2/\alpha - 1}(\log n)^\gamma = O(\lambda_{\min} V_n) \quad \text{for some } \gamma > 2/\alpha \tag{5.4}$$

or, somewhat more strongly,

$$n^\gamma = O(\lambda_{\min} V_n) \quad \text{for some } \gamma > 2/\alpha - 1 \tag{5.5}$$

implies (5.3).

(ii) *If* $\alpha = 2$, *then* (5.2) *is sufficient for* (5.3). *If* $\alpha = 2$ *and* $\|Z_n\| = O(n^\delta)$ *for some real* δ, *then* (5.2) *is implied by*

$$(\log n)^\gamma = o(\lambda_{\min} V_n) \quad \text{for some } \gamma > 1.$$

Proof. The elementary inequality

$$x^\alpha y^{\alpha-1} \leqslant x + y^{-1}, \quad x \geqslant 0,\ y > 0,\ 0 < \alpha \leqslant 1,$$

implies

$$\|Z_n' V_n^{-1} Z_n\|^{\alpha/2} n^{\alpha/2-1} \leqslant \|Z_n' V_n^{-1} Z_n\| + n^{-1}, \quad n \geqslant N. \tag{5.6}$$

If $\|Z_n\| = O(n^\delta)$ for some $\delta > 0$, then $\|V_n\| = O(n^{\delta'})$ holds for some $\delta' > 0$, whence $\log\|V_n\| = O(\log n)$. Noting

$$\|Z_n' V_n^{-1} Z_n\| \leqslant \operatorname{trace} Z_n' V_n^{-1} Z_n = \operatorname{trace} V_n^{-1/2} Z_n Z_n' V_n^{-1/2},$$

Lemma 4 can be applied in a similar way as in the proof of Theorem 3. With $\log\|V_n\| = O(\log n)$, this yields

$$\sum_{n=N}^{\infty} \|Z_n' V_n^{-1} Z_n\| (\log n)^{-\gamma} < \infty \quad \text{for any } \gamma > 1.$$

Since

$$\sum_{n=N}^{\infty} n^{-1} (\log n)^{-\gamma} < \infty, \quad \gamma > 1,$$

inequality (5.6) leads to

$$\sum_{n=N}^{\infty} \|Z_n' V_n^{-1} Z_n\|^{\alpha/2} n^{\alpha/2-1} (\log n)^{-\gamma} < \infty \quad \text{for any } \gamma > 1.$$

Together with (5.4) or (5.5), this implies (5.3), and part (i) of Corollary 3 is proved. Part (ii) can be proved as Theorem 3. $\square$

The following examples illustrate the scope of Corollary 3 and demonstrate that sensible results can also be obtained if errors are not square integrable.

Examples. (i) If $q=1$ and the regressors are polynomials,

$$Z_n=(n^{c_1},\dots,n^{c_p})',\quad -\tfrac{1}{2}<c_1<c_2<\cdots<c_p,$$

the matrix V_n is nonsingular for $n\geq N=p$. Asymptotically, $n^{2c_1+1}=O(\lambda_{\min}V_n)$ holds (the converse $\lambda_{\min}V_n=O(n^{2c_1+1})$ is stated by Eicker [5, (3.5)]). Therefore (5.5) holds if $c_1>1/\alpha-1$, and $\hat{\beta}_n\to\beta$ a.s. follows if the error moments of some corresponding order α are uniformly bounded. Note that $\alpha>1$ allows for $c_i=0$ for some i, i.e. inclusion of a constant term.

(ii) If $q=1$, $p=2$ and $Z_n=(1,(\log n)^{\gamma})'$ with some $\gamma>0$, then $\lambda_{\min}V_n$ diverges at the same order as $n(\log n)^{-2}$, independently of $\gamma>0$. Hence (5.4) holds for $\alpha>1$, whereas it fails to hold if $\alpha=1$.

Proof. (i) Defining $D_n=\mathrm{diag}(n^{c_1+1/2},\dots,n^{c_p+1/2})$, we have that $D_n^{-1}V_nD_n^{-1}\to M$ positive definite, see Grenander and Rosenblatt [7], p. 246. The proof given there for $c_i=i-1$, $i=1,\dots,p$, carries over to the present situation. Positive definiteness of M and continuity of $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ imply that there exist constants c_1 and c_2 independently of x such that

$$0<c_1\leq x'V_nx/x'D_n^2x\leq c_2<\infty,\quad n\geq p,\ x\neq 0.$$

In particular, $\{\lambda_{\min}V_n\}$ diverges at the same order as $\{\lambda_{\min}D_n^2\}$, whence $n^{2c_1+1}=O(\lambda_{\min}V_n)$ and $\lambda_{\min}V_n=O(n^{2c_1+1})$.

(ii) If $p=2$, the inequality

$$|V_n|/\mathrm{trace}\ V_n\leq\lambda_{\min}V_n\leq 2|V_n|/\mathrm{trace}\ V_n,\quad n=1,2,\dots,$$

can be used to determine the order of divergence of $\{\lambda_{\min}V_n\}$. In particular, if $Z_n=(1,x_n)'$ with x_n scalar, then

$$|V_n|/\mathrm{trace}\ V_n=\sum(x_i-\bar{x}_n)^2/(1+n^{-1}\sum x_i^2).$$

Results of Fahrmeir and Kaufmann [6, p. 198] imply that $\sum(x_i-\bar{x}_n)^2$ diverges as $n(\log n)^{2\gamma-2}$, whereas $1+n^{-1}\sum x_i^2$ diverges as $(\log n)^{2\gamma}$, for any $\gamma>0$. Hence $\lambda_{\min}V_n$ diverges as $n(\log n)^{-2}$, for any $\gamma>0$. □

6. Some final remarks

The most drastic restrictions in this paper are consideration of finite dimensional martingales, the monotonicity condition (1.2) and nonrandom norming matrices from Section 3 on. However, not all of these seem limitations necessarily connected with the approach presented.

Since discussion is largely based on norms, at least some of the results should carry over to more general spaces without major difficulties. The monotonicity condition (1.2) plays an essential role in several places, and it does not seem easy to weaken it. Combining the methods given here with the stopping rule methods common for obtaining local theorems with random norming for scalar martingales, it should be possible to also obtain such theorems for multivariate martigales. This would lead to conditions for regression models with random regressors.

References

[1] B.D.O. Anderson and J.B. Moore, A matrix Kronecker Lemma, Linear Algebra Appl. 15 (1976) 227-234.

[2] T.W. Anderson, An Introduction to Multivariate Statistical Analysis (Wiley, New York, 2nd edition, 1984).

[3] T.W. Anderson and J.B. Taylor, Some theorems concerning strong consistency of least squares estimates in linear regression models, Working paper No. 47, Inst. Math. Studies Social Sci. Stanford Univ. (1974).

[4] Y.S. Chow, A martingale inequality and the law of large numbers, Proc. Amer. Math. Soc. 11 (1960) 107-111.

[5] F. Eicker, Asymptotic normality and consistency of the least squares estimators for families of linear regressions, Ann. Math. Statist. 34 (1963) 447-456.

[6] L. Fahrmeir and H. Kaufmann, Asymptotic inference in discrete response models, Statist. Papers 27 (1986) 179-205.

[7] U. Grenander and M. Rosenblatt, Statistical Analysis of Stationary Time Series (Almquist and Wiksell, Stockholm, 1956).

[8] J. Hájek and A. Rényi, Generalization of an inequality of Kolmogorov, Acta Math. Sci. Hungar. 6 (1955) 281-283.

[9] P. Hall and C.C. Heyde, Martingale Limit Theory and its Application (Academic Press, New York, 1980).

[10] A. Kolmogorov, Über die Summen durch den Zufall bestimmter unabhängiger Größen, Math. Ann. 99 (1928) 309-319.

[11] T.L. Lai, H. Robbins and C.Z. Wei, Strong consistency of least squares estimates in multiple regression II, J. Multivariate Anal. 9 (1979) 343-361.

[12] T.L. Lai and 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.

[13] J. Marcinkiewicz and A. Zygmund, Sur les fonctions indépendants, Fund. Math. 29 (1937) 60-90.

[14] W.F. Stout, Almost Sure Convergence (Academic Press, New York, 1974).