1

A note on convergence rates in the strong law of large numbers for associated sequences

Cecília Azevedo
Centro de Matemática da Universidade do Minho
Department of Mathematics
Campus de Gualtar
4710-057 Braga Portugal
cecilia@math.uminho.pt

***Abstract*—Let X_n, $n \in \mathbb{N}$, be a strictly stationary sequence of centered and associated real random variables. Sufficient conditions for the strong law of large numbers to hold are known, but no rates of convergence where given. We derive an upper bound for this convergence rate. This rate is made explicit for geometrically decreasing covariances.**

***Index Terms*—association; kernel estimation; convergence rates**

I. Introduction

LET $X_n, n \in \mathbb{N}$, be a strictly stationary sequence of centered random variables and define, for each $n \in \mathbb{N}$, $S_n = X_1 + \cdots + X_n$.
If the variables are independent it is well known that

$$\frac{1}{n} S_n \to 0 \ \ a.s.$$

and the optimal rate of this convergence is characterized by the law of iterated logarithm.
Extensions of these characterizations to dependent sequences have been considered in the literature.
In this note we will be interested in the case where the sequence X_n, $n \in \mathbb{N}$, is associated, that is,

Definition 1.1: The sequence X_n, $n \in \mathbb{N}$ is such that, given any real-valued cordinatewise increasing functions g and h defined on $\mathbb{R}^I$, where I is any finite subset of $\mathbb{N}$,

$$\text{Cov}\ (g(X_i, i \in I), h(X_j, j \in I)) \geq 0,$$

whenever this covariance exists.

A strong law of large numbers for strictly stationary associated random variables has been proved by
Newman [9] under the assumption

$$\frac{1}{n} \sum_{j=1}^{n} \text{Cov}\,(X_1,\, X_j) \longrightarrow 0.$$

An extension of this result to nonstationary sequences has been proved by Birkel [5]. The results referred give no indication about convergence rates. The usual approach to the treatment of such characterizations requires the use of exponential inequalities, in order to optimize the rates. Such type of inequalities have became available for associated sequences only quite recently, proved by Ioannides, Roussas [8], thus providing the tools to take some steps in identifying rates of convergence.

The research was partially supported by CMAT - Centro de Matemática and FCT - Fundação para a Ciência e Tecnologia under the program POCI 2010.

The literature on association is extensive.
Positive association has found applications in reliability, statistical mechanics, probability, stochastic processes, and statistics. So has negative association but to a lesser degree. Many more applications are to be anticipated in a host of areas, and, in particular, those areas where spatial statistics play an important role. Such areas are, for example, analysis of agricultural field experiments, geostatistical analysis, image analysis, oceanographic applications, and signal processing.

Positive association, called just association, was introduced by Esary et al.[7], who also derived an abundance of properties, discussed equivalent characterizations, and presented several applications in probability and statistics. Their motivation stemmed from the usefulness of Positive association in the context of systems reliability.

The main result in this note proves an upper bound for the convergence rate for the strong law of large numbers under positive association. The identification of the exact convergence rate seems out of reach of the method of proof used.
In fact, the proof presented in next section uses the first Borel-Cantelli Lemma to prove the almost sure convergence of the various sequences treated. For the identification of a lower bound for the convergence rate it would be necessary to use some sort of version of the second Borel-Cantelli Lemma which, as far as the author knows, is not available for associated sequences.

II. Results

In order to present our results we need to introduce some notation required by the use of the exponential inequality by Ioannides, Roussas [8], upon which our approach is based.

For each $n \in \mathbb{N}$ let $p_n \leq n$ be an integer, such that $p_n \to +\infty$,

and let r_n be the largest integer less or equal to $\frac{n}{2p_n}$.
It is also required that $r_n \to +\infty$.

Theorem 2.1: Let $X_n,\, n \in \mathbb{N}$ be strictly stationary centered and associated random variables with finite variances. Suppose that

- there exists a constant $M > 0$ such that
$$|X_n| \leq M,\, n \in \mathbb{N}$$
- the covariances $\mathrm{Cov}\,(X_1,\, X_n)$ are decreasing
- $\mathbb{E}(S_n^2) = O(n)$
- the variables satisfy, for some constant $c > 0$,
$$\mathrm{Cov}\,(X_1,\, X_{p_n+1}) \leq \exp\left(-c r_n \varepsilon_n\right),$$
with
$$\varepsilon_n^2 = O\left(\frac{\log\log n}{r_n}\right). \tag{1}$$

Then
$$\mathbb{P}\left(\limsup \frac{S_n}{n\varepsilon_n} > 1\right) = 0.$$

The assumptions on the variables and on the covariances used in this theorem are inherited from the conditions for the use of the exponential inequality by Ioannides, Roussas [8].
As a consequence of this theorem, we may identify an upper bound for the convergence rate of the strong law of large numbers when the covariances are geometrically decreasing.

Corollary 2.2: Let $X_n,\, n \in \mathbb{N}$, be strictly stationary centered and associated random variables with finite variances. Suppose there exists a constant $M > 0$ such that $|X_n| \leq M,\, n \in \mathbb{N}$, and there exists some $\rho_0 > 0$ and $\rho \in (0,1)$ such that
$\mathrm{Cov}\,(X_1,\, X_n) = \rho_0 \rho^n$. Then,
$$\mathbb{P}\left(\limsup \frac{n^{1/3}}{(\log\log n)^{2/3}}\,\frac{1}{n} S_n > 1\right) = 0.$$

III. Auxiliary results

In this section we quote the inequalities that are used in course of proof of our results.

First the exponential inequality.

Theorem 3.1: (Ioannides, Roussas [8])
Let $X_n,\, n \in \mathbb{N}$, be a strictly stationary sequence of centered and associated variables for which there exists a constant $M > 0$ such that
$$|X_n| \leq M.$$

Suppose further that the sequence of covariances $\mathrm{Cov}\,(X_1,\, X_n)$ is decreasing and satisfy, for some $\alpha > 1$ and some sequence $\varepsilon_n \to 0$,
$$\mathrm{Cov}\,(X_1,\, X_{p_n+1}) \leq \exp\left(-\frac{4(M+1)}{3M}\,\frac{\alpha^{1/2}}{2}\, r_n \varepsilon_n\right).$$

Then, for n sufficiently large, there exists a constant $c_0 > 0$ such that
$$\mathbb{P}\left(\frac{1}{n}\,|S_n| \geq \varepsilon_n\right) \leq c_0 \exp\left(-\frac{2M^2}{9}\, r_n \varepsilon_n^2\right). \tag{2}$$

In order to deduce the almost sure convergence we must choose the sequence ε_n such that
$$r_n \varepsilon_n^2 = \frac{9}{2M^2}\,\alpha \log n,$$
with $\alpha > 1$ to derive that
$$\sum_n \mathbb{P}\left(\frac{1}{n}\,|S_n| > \varepsilon_n\right) < \infty$$
and use the Borel-Cantelli Lemma to conclude the convergence.

For the optimization of the rate of convergence we shall use the following Kolmogorov type inequality.

Theorem 3.2: (Newman, Wright [10])
Let $X_n,\, n \in \mathbb{N}$, be centered random variables with finite variances. Define
$$S_m^* = \max\left(0, S_1, \ldots, S_m\right)$$
and
$$s_m^2 = \mathbb{E}\big(S_m^2\big).$$
Then,
$$\mathbb{E}\big((S_m^*)^2\big) \leq \mathrm{Var}\,(S_m)$$
and
$$\begin{aligned} P\Big(\max\left(|S_1|, \ldots, |S_m|\right) \geq \lambda s_m\Big) &\leq \\ \leq 2P\Big(|S_m| \geq (\lambda - \sqrt{2}) s_m\Big)&. \end{aligned} \tag{3}$$

IV. Proofs

Proof of Theorem 2.1: Take $t_n = \sqrt{n}\,\epsilon_n$.
According to Theorem 3.1,
$$P(S_n \geq n\varepsilon_n) \leq C_0 \exp(-c r_n \varepsilon_n^2),$$
where $c = \dfrac{2\,M^2}{9}$.
Taking account of (1), there exist some constant $c' > 0$, such that,
$$P(S_n \geq n\varepsilon_n) = P(S_n \geq \sqrt{n}\, t_n) \leq C_0 (\log n)^{-c'}.$$

To prove this theorem it is enough that to verify that, for every $\delta > 0$, we have
$$P\left(\limsup\{S_n \geq (1+\delta)\sqrt{n}\, t_n\}\right) = 0. \tag{4}$$

From the assumptions of the theorem it follows that there exist two positive constants c_1, c_2 such that
$$c_1 n \leq \mathbb{E}(S_n^2) \leq c_2 n.$$

Next we are more precise about the choice of the sequence ε_n.

Choose this sequence as

$$\varepsilon_n^2 = \left(\frac{c_2}{c_1}\right)^2 \frac{1}{c} \frac{\log\log n}{r_n}.$$

Now, using (2)

$$\begin{aligned} P(S_n \geq (1+\delta)\sqrt{n}\, t_n) \leq \\ \leq C_0 \exp\left(-\left(\tfrac{c_2}{c_1}\right)^2 (1+\delta)^2 \log\log n\right) \sim \\ \sim C_0 (\log n)^{-\left(\frac{c_2}{c_1}\right)^2 (1+\delta)^2}. \end{aligned} \tag{5}$$

If we choose $n_k = [\theta^k]$, for some $\theta > 1$, where $[x]$ represents the largest integer that is less ou equal than x, we find

$$(\log n_k)^{-\left(\frac{c_2}{c_1}\right)^2 (1+\delta)^2} \sim k^{-\left(\frac{c_2}{c_1}\right)^2 (1+\delta)^2},$$

so that the left hand side of (5), along the subsequence indicated, defines a convergent series and it follows that, again along the same subsequence, (4) is verified.

Next we need to control the remaining terms in the sequence.

For each $k \in \mathbb{N}$, define $S^*_{n_k} = \max_{n \leq n_k} S_n$, and suppose that $n_{k-1} < n \leq n_k$.
Then, we have that

$$P(S_n > (1+\delta)\sqrt{n}\, t_n) \leq P(S^*_{n_k} > (1+\delta)\sqrt{n_{k-1}}\, t_{n_{k-1}}).$$

But given the choice of the subsequence of indexes and of the sequence ε_n,

$$\begin{aligned} \sqrt{n_{k-1}}\, t_{n_{k-1}} = n_{k-1}\epsilon_{n_{k-1}} \sim \\ \sim \tfrac{n_k}{\theta}\, \epsilon_{n_k} \sqrt{\tfrac{\log\log \theta^{k-1}}{r_{\theta^{k-1}}} \tfrac{r_{\theta^k}}{\log\log \theta^k}} \end{aligned}$$

Suppose now that $\frac{r_{\theta^k}}{r_{\theta^{k-1}}} \sim \tilde{c}$.
It follows then that

$$\sqrt{n_{k-1}} \quad t_{n_{k-1}} \sim n_k \epsilon_{n_k} \frac{\tilde{c}}{\theta}.$$

Now, it is always possible choose θ in such way that

$$1 < \frac{\tilde{c}}{\theta} < 1 + \delta.$$

Then, $\dfrac{1+\delta}{\theta/\tilde{c}} = 1 + \delta'$, for some $\delta' < \delta$.

Finally, for all sufficiently large k,

$$\begin{aligned} P(S_n > (1+\delta)\sqrt{n}\, t_n) \leq \\ \leq P(S^*_{n_k} > (1+\delta')\sqrt{n_k}\, t_{n_k}). \end{aligned} \tag{6}$$

Using now (3)

$$\begin{aligned} P(S^*_{n_k} > (1+\delta')\sqrt{n_k}\, t_{n_k}) \leq \\ \leq 2P\left(S_{n_k} > (\tfrac{1+\delta}{c_2} - \tfrac{\sqrt{2}}{t_{n_k}}) c_1 \sqrt{n_k}\, t_{n_k}\right). \end{aligned}$$

If we use now (2) on this last expression, we find

$$\begin{aligned} P(S_n > (1+\delta)\sqrt{n}\, t_n) \leq \\ \leq 2C_0 \exp\left(-\left(\tfrac{1+\delta'}{c_2} - \tfrac{\sqrt{2}}{t_{n_k}}\right)^2 c_1^2 \log\log n_k\right) \leq \\ \leq 2C_0 (\log n_k)^{-\left(\frac{1+\delta'}{c^2} - \frac{\sqrt{2}}{t_{n_k}}\right)^2 c_1^2} \sim \\ \sim k^{-\left(\frac{1+\delta'}{c_2} - \frac{\sqrt{2}}{t_{n_k}}\right)^2 c_1^2}. \end{aligned}$$

If $t_{n_k} \to +\infty$, that is, $\sqrt{n}\,\varepsilon_n \to +\infty$, the right side of the above inequality defines a convergent series. ■

The condition on the sequence ε_n translates into a condition on the sequence r_n introduced earlier:

$$\frac{n \log\log n}{r_n} \longrightarrow +\infty.$$

As we have used the exponential inequality stated in Theorem 3.1, the sequence p_n must be such that

$$\begin{aligned} \mathrm{Cov}\,(X_1, X_{p_n+1}) \leq \\ \leq \exp\left(-\tfrac{4(M+1)}{3M}\, \tfrac{\alpha^{1/2}}{2}\, r_n \varepsilon_n\right) = \\ = \exp\Big(-M_2' (r_n \log\log n)^{1/2}\Big), \end{aligned}$$

or, written in a more convenient way,

$$r_n \leq M_3 \frac{\log^2 \mathrm{Cov}\,(X_1, X_{p_n+1})}{\log\log n}. \tag{7}$$

Proof of Corollary 2.2:
Suppose that $\mathrm{Cov}\,(X_1, X_{k+1}) = \rho_0 \rho^k$, for some $\rho \in (0,1)$. In this case (7) may be written as

$$r_n \leq M_3 \frac{p_n^2 \log^2 \rho}{\log\log n} + \frac{\log^2 \rho_0}{\log\log n}. \tag{8}$$

Of course, the second term of the sum is irrelevant for the behavior of r_n, as it converges to zero.

We are interested to choose the sequence r_n as large as possible, so the case of interest is when equality holds in (8). Now, the way the sequences p_n and r_n were defined implies that

$$p_n = \frac{1}{2x_n} \frac{n}{r_n},$$

for some $x_n \to 1$.

Inserting this into (8) it follows that

$$r_n = O\Big(\frac{n^{2/3}}{(\log\log n)^{1/3}}\Big)$$

and

$$p_n = O(n \log\log n)^{1/3}.$$

In these conditions, the rate of convergence to zero of

$\frac{1}{n}S_n$ is given by

$$\frac{1}{\varepsilon_n},$$

that is

$$\frac{n^{1/3}}{(\log\log n)^{2/3}},$$

as announced.

It remains to check the assumption in Theorem 2.1 on $\mathbb{E}(S_n^2)$. But, taking account of the nonnegativity of the covariances it follows easily that

$$\rho_0 n \leq \mathbb{E}(S_n^2) \leq \rho_0 \frac{\rho}{1-\rho}\, n. \qquad \blacksquare$$

V. Conclusions

By taking $\{X_n\}_{n\in\mathbb{N}}$, a strictly stationary sequence of centered and associated real random variables, we know sufficient conditions for the strong law of large numbers to hold but no rates of convergence where given.

In this note we derived an upper bound for this convergence rate. This rate was made explicit for geometrically decreasing covariances.

In fact we proved, under certain conditions on the covariances of the random variables and by using the exponential inequality by Ioannides, Roussas [8], that

$$\mathbb{P}\left(\limsup \frac{S_n}{n\varepsilon_n} > 1\right) = 0.$$

Also we proved that if there exists a constant $M > 0$ such that $|X_n| \leq M$, $n \in \mathbb{N}$, and there exists some $\rho_0 > 0$ and $\rho \in (0,1)$ such that

$$\mathrm{Cov}\,(X_1,\, X_n) = \rho_0 \rho^n$$

then,

$$\mathbb{P}\left(\limsup \frac{n^{1/3}}{(\log\log n)^{2/3}}\,\frac{1}{n}S_n > 1\right) = 0.$$

Acknowledgments

The research was partially supported by CMAT and FCT under the program POCI 2010.
Author would like to thank the referees for a careful reading of this paper.

References

[1] Amini, M.; Bozorgnia, A.; The Strong Law of Large Numbers and Asymptotic Behavior of Quantiles for Negatively Dependent Random Variables, *2000 WSES International Conference on Applied and Theoretical Mathematics*, Edited by N.E.Mastorakis, ISBN: 960-8052-20-3, pp. 1771-1774, 2000

[2] Azevedo, C.; Covariance Estimation of the limit process of the empirical process under association , *2000 WSES International Conference on Applied and Theoretical Mathematics*, Edited by N.E.Mastorakis, ISBN: 960-8052-20-3, pp. 2051-2056, 2000

[3] Billingsley, P.; Probability and Measure, *Wiley, New York*.

[4] Birkel, T.; Moment bounds for associated sequences, *Ann. Probab.*, 16:1184-1193, 1988.

[5] Birkel, T.; A note on the strong law of large numbers for positively dependent random variables, *Statist. Probab. Letters*, 14:355-362, 1992.

[6] Devroye, L.; Exponential inequalities in nonparametric estimation, in G.G. Roussas (ed.),*Nonparametric functional estimation and related topics*, Kluver Academic Publishers, Dordrecht, 31-44, 1991.

[7] Esary,J.; Proschan, F.; Walkup, D.; Association of random variables with applications, *Ann. Math. Statist.* 38, pp. 1466-1474, 1967.

[8] Ioannides, D.A. and Roussas,G.; Exponential inequality for associated random variables, *Statist. Probab. Lett.*, 42:423-431, 1999.

[9] Newman, C.M.; Asymptotic independence and limit theorems for positively and negatively dependent random variables, *Inequalities in Statistics and Probability*, IMS Lecture Notes - Monograph Series 5:127-140, 1984.

[10] Newman, C.M. and Wright, A.L.; An invariance principle for certain dependent sequences, *Ann. of Probab.*, 9:671-675, 1981.

[11] Roussas, G.G.; Kernel estimates under association: strong uniform consistency, *Statist. Probab. Lett.*, 12:393-403, 1991.

[12] Zarei, H.; Complete convergence for negatively dependent random variables, *Proceedings of the 9th WSEAS International Conference on Applied Mathematics, Istanbul, Turkey*, pp. 338-341, 2006.