A Weak Law of Large Numbers Under Weak Mixing

Bruce E. Hansen*
University of Wisconsin†

January 2019

Abstract

This paper presents a new weak law of large numbers (WLLN) for heterogenous dependent processes and arrays. The dependence requirements are notably weaker than the best available current results (due to Andrews (1988)). Specifically, we show that the WLLN holds when the process is weak mixing, only requiring that the mixing coefficients Cesàro sum to zero. This is weaker than the conventional assumption of strong mixing.

*Research supported by the National Science Foundation and the Phipps Chair.
†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.

1 Introduction

One of the foundations for asymptotic inference is the weak law of large numbers (WLLN). For dependent and strictly stationary time series, the most flexible and powerful result is the Ergodic Theorem, which states that sample means converge almost surely to the population mean under the minimal condition that the process is ergodic. The latter only requires that separated events are on average asymptotically independent. For many applications, however, the assumption of strict stationarity is too restrictive. It does not allow for heterogeneous time series, nor allow for random array structures. In such settings the best available dependence conditions for the WLLN are due to Andrews (1988), who showed that uniform integrability plus strong mixing are sufficient for the WLLN. His result is particularly powerful as it does not require any rate of convergence for the mixing coefficients.

A limitation with Andrews' WLLN is that strong mixing may be unnecessarily restrictive. In classical ergodic theory, distinctions are made between ergodic processes, weak mixing processes, and strong mixing processes. Ergodicity is the weakest requirement (and broadest class), strong mixing the strongest requirement (and most narrow class). The distinction between weak mixing and strong mixing is that the former requires that the Cesàro sum of mixing coefficients is zero, while strong mixing requires the coefficients to limit to zero. The Cesàro limit requirement is weaker, as it can hold even when the mixing coefficients do not converge to zero.

This may seem like a minor difference in practice, but classical ergodic theory has argued that it is a major difference. Specifically, the view is that weak mixing processes are relatively generic, while strong mixing processes are relatively special. For example, the textbook of Petersen (1983, p. 71-72) states that "there is a sense in which *almost every measure preserving transformation is weakly mixing but not strong mixing*.... Halmos (1944) proved that *with respect to the weak topology, the set of weakly mixing measure preserving transformation systems is residual* (i.e. the complement of a first category set); while Rokhlin (1948) showed that *with respect to the weak topology the set of all strongly mixing transformations is of the first category*. Thus, in this particular sense, *the 'generic' measure preserving transformation is weakly mixing but not strongly mixing*." (emphasis as quoted).

For examples of processes which are weak mixing but not strong mixing, see Section 4.5 of Petersen (1983), which describes the examples of Kakutani (1973) and Chacon (1969). Other examples are provided by Maruyama (1949) and Katok and Stepin (1967).

By showing that the WLLN holds for heterogenous weak mixing processes and arrays, our result brings the theory for heterogeneous dependent processes closer to the classical Ergodic Theorem.

One of the interesting features of our result is that the proof is elementary. It uses the standard representation of the variance of the trimmed mean as the weighted Cesàro sum of the covariances, and bounds the latter using the mixing inequality for bounded random variables. The deviation of the mean from the trimmed mean is bounded conventionally. This proof method is notably different from that of Andrews (1988) who approximated the sample mean by the sum of M martingale difference means, bounded the latter using moment bounds for martingale difference sequences, and bounded the deviation by a mixingale inequality.

In the econometrics literature, Andrews' WLLN has been generalized to allow for trending moments by Davidson (1993) and De Jong (1995, 1998). These papers do not weaken, however, the dependence conditions. For strong laws of large numbers for dependent heterogenous processes, see Hansen (1991, 1992), De Jong (1995, 1996), and Davidson and De Jong (1997). For a textbook treatment see Davidson (1994).

In Section 2 we review the concepts of ergodicity and mixing for stationary processes. In Section 3 we discuss mixing for heterogeneous arrays. Section 4 presents our WLLN. Section 5 presents the proof of the result.

2 Ergodicity and Mixing

Let $(\Omega, \boldsymbol{F}, P)$ denote a probability space. A stochastic process $\{X_t \in \mathbb{R} : -\infty < t < \infty\}$ is a measurable mapping from Ω to $\mathbb{R}^\infty$. Letting $\omega \in \Omega$, we can write $X_t(\omega)$ to indicate that the process depends on the element ω. We can then define the shift transformation T by $X_{t+1}(\omega) = X_t(T\omega)$.

An event $E \in \boldsymbol{F}$ is **invariant** if $E = T^{-1}E$. The process X_t is called **ergodic** if all invariant events have probability either 0 or 1. Intuitively, an ergodic process visits all parts of the probability space, and never gets stuck in a subspace.

There are several equivalent ways to characterize an ergodic process. One is that X_t is ergodic if and only if for all $A, B \in \boldsymbol{F}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \big(P(T^{-m}A \cap B) - P(A)P(B)\big) = 0.$$

See, for example, Theorem 1.4 of Billingsley (1965) or Corollary 1.14.2 of Walters (1982). This means, intuitively, that the translation $T^{-m}A$ becomes independent of B on average.

For some purposes it is desirable to work with somewhat stronger characterizations of asymptotic independence. The process X_t is called **weakly mixing** if for all $A, B \in \boldsymbol{F}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \big|P(T^{-m}A \cap B) - P(A)P(B)\big| = 0$$

and is called **strongly mixing** if

$$\lim_{m\to\infty} \big|P(T^{-m}A \cap B) - P(A)P(B)\big| = 0.$$

Weak mixing also can be interpreted as stating that $T^{-m}A$ becomes independent of B provided we neglect a few instances. Strong mixing can be interpreted as stating that $T^{-m}A$ is asymptotically independent of B.

From these expressions it is evident that ergodicity implies weak mixing, and weak mixing implies strong mixing. It is known that this nesting is strict, as there are examples of ergodic transformations which are not weak mixing, and weak mixing transformations which are not strong mixing. Strong mixing requires that the probabilities $P(T^{-m}A \cap B) - P(A)P(B)$ limit to zero, but this is not required by the Cesàro summability of weak mixing. For concrete examples of transformations which are weak but not strong mixing see the references in the introduction.

3 Mixing for Heterogenous Arrays

In econometrics we are frequently interested in heterogenous stochastic processes and arrays (indexed by sample size n). For this purpose the most commonly used dependence tool are mixing coefficients. For a random array $\{X_{nt} : t = 1, ..., n\}$ the mixing coefficents are defined as

$$\alpha_n(m) = \sup_{-\infty<t<\infty} \sup_{A\in\boldsymbol{F}^n_{-\infty,t}, B\in\boldsymbol{F}^n_{t+m,\infty}} |P(A \cap B) - P(A)P(B)|$$

where $\boldsymbol{F}^n_{-\infty,t} = \sigma\left(..., X_{n,t-1}, X_{n,t}\right)$ and $\boldsymbol{F}^n_{t+m,\infty} = \sigma\left(X_{n,t+m}, X_{n,t+m+1}, ...\right)$. The latter are σ-fields generated by the past and future values of the stochastic process, respectively, separated by m time periods. The mixing coefficients $\alpha_n(m)$ measure the serial dependence as the degree of separation is increased. For stationary stochastic processes the coefficients do not depend on n.

It is standard to say that the stochastic process X_{nt} is **strong mixing** if

$$\sup_{n\geq1} \alpha_n(m) \to 0$$

as $m \to \infty$. This is an analog of the ergodic theory concept of strong mixing.

We now introduce an analog of the ergodic theory concept of weak mixing.

Definition 1. *X_{nt} is **weak mixing** if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \alpha_n(m) = 0. \tag{1}$$

The expression (1) states that the Cesàro sum of the mixing coefficients is zero.

Since convergence implies Cesàro convergence, weak mixing implies strong mixing. Thus weak mixing is a strictly broader class of stochastic processes than strong mixing. For example, consider the mixing coefficient sequence $\alpha_n(m) = \mathbf{1}\left(\sqrt{m} = [\sqrt{m}]\right) = \{1, 0, 0, 1, 0, 0, 0, 0, 1, ...\}$. This does not have a limit, but its Cesàro sum limits to zero. Hence it is weak mixing but not strong mixing.

For another example using arrays, take the process $X_{nt} = e_t + e_{t-q(n)}$ with e_t i.i.d. If $q(n) \to \infty$ yet $q(n)/n \to 0$ as $n \to \infty$ then this process is weak mixing but not strong mixing.

4 Weak Law of Large Numbers

Define the sample mean

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_{nt}.$$

Theorem 1. *If X_{nt} is weak mixing and*

$$\lim_{B\to\infty}\sup_n \frac{1}{n}\sum_{t=1}^{n} E\left|X_{nt}\mathbf{1}\left(|X_{nt}|>B\right)\right| = 0 \tag{2}$$

then

$$E\left|\overline{X}_n - E\left(\overline{X}_n\right)\right| \to 0 \tag{3}$$

and

$$\overline{X}_n - E\left(\overline{X}_n\right) \to_p 0 \tag{4}$$

as $n\to\infty$.

Theorem 1 shows that the sample mean converges in L_1 and converges in probability. The condition (2) is an average uniform integrability condition. It is implied if X_{nt} is uniformly integrable:

$$\lim_{B\to\infty}\sup_n \sup_{1\le t\le n} E\left|X_{nt}\mathbf{1}\left(|X_{nt}|>B\right)\right| = 0$$

or if X_{nt} has a uniformly bounded moment:

$$\lim_{B\to\infty}\sup_n \sup_{1\le t\le n} E\left|X_{nt}\right|^r < \infty$$

for some $r>1$.

Theorem 1 generalizes the WLLN for strong mixing processes of Andrews (1988) (his Theorem 2, example 4). Primarily, Theorem 1 relaxes the assumption of strong mixing to that of weak mixing. Theorem 1 shows that weak mixing is sufficient for consistent estimation.

5 Proof

We show (3). (4) follows by Markov's inequality.

Without loss of generality assume $E(X_{nt})=0$. Fix $\varepsilon>0$. Pick B large enough such that

$$\sup_n \frac{1}{n}\sum_{t=1}^{n} E\left|X_{nt}\mathbf{1}\left(|X_{nt}|>B\right)\right| \le \frac{\varepsilon}{4} \tag{5}$$

which is feasible under (2). Define

$$\begin{aligned} W_{nt} &= X_{nt}\mathbf{1}\left(|X_{nt}| \le B\right) - E\left(X_{nt}\mathbf{1}\left(|X_{nt}| \le B\right)\right) \\ Z_{nt} &= X_{nt}\mathbf{1}\left(|X_{nt}| > B\right) - E\left(X_{nt}\mathbf{1}\left(|X_{nt}| > B\right)\right) \end{aligned}$$

so that

$$E\left|\overline{X}_n\right| = E\left|\overline{W}_n + \overline{Z}_n\right| \le E\left|\overline{W}_n\right| + E\left|\overline{Z}_n\right|. \tag{6}$$

By the triangle inequality and (5)

$$\begin{aligned} E\left|\overline{Z}_n\right| &\le \frac{1}{n}\sum_{t=1}^{n} E\left|Z_{nt}\right| \\ &\le \frac{2}{n}\sum_{t=1}^{n} E\left|X_{ni}\mathbf{1}\left(|X_{nt}| > B\right)\right| \\ &\le \frac{\varepsilon}{2}. \end{aligned} \tag{7}$$

It is useful to observe that W_{nt} satisfies the bound $|W_{nt}| \le 2B$ and has the same mixing coefficients as X_{nt}. By the mixing inequality for bounded random variables (e.g. Theorem A.5 of Hall and Heyde (1980)), and the fact that W_{nt} are mean zero,

$$\left|E\left(W_{nt}W_{nj}\right)\right| \le 16B^2\alpha_n\left(|t-j|\right). \tag{8}$$

By Jensen's inequality, (8) and $\alpha_n(0) \le 1/4$,

$$\begin{aligned} \left(E\left|\overline{W}_n\right|\right)^2 &\le E\left|\overline{W}_n\right|^2 \\ &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{n} E\left(W_{nt}W_{nj}\right) \\ &\le \frac{1}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{n} \left|E\left(W_{nt}W_{nj}\right)\right| \\ &\le \frac{16B^2}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{n} \alpha_n\left(|t-j|\right) \\ &= 16B^2\left(\frac{\alpha_n(0)}{n} + \frac{2}{n}\sum_{m=1}^{n-1}\left(1-\frac{m}{n}\right)\alpha_n(m)\right) \\ &\le 16B^2\left(\frac{1}{4n} + \frac{2}{n}\sum_{m=1}^{n-1}\alpha_n(m)\right) \\ &\le \frac{\varepsilon^2}{4}. \end{aligned}$$

The final inequality holds for n large enough since X_{nt} is weak mixing. Thus

$$E\left|\overline{W}_n\right| \leq \frac{\varepsilon}{2}. \tag{9}$$

Together, (6), (7) and (9) show that $E\left|\overline{X}_n\right| \leq \varepsilon$ which establishes (3) as claimed.

References

[1] Andrews, Donald W. K. (1988): "Laws of large numbers for dependent non-identically distributed random variables," *Econometric Theory*, 4, 458-467.

[2] Billingsley, Patrick (1965): *Ergodic Theory and Information*, Wiley.

[3] Chacon, R. V. (1969): "Weakly mixing transformations which are not strongly mixing," *Proceedings of the American Mathematical Society*, 22, 559-562.

[4] Davidson, James (1993) "An L1-convergence theorem for heterogeneous mixingale arrays with trending moments", *Statistics and Probability Letters*, 16, 301-306.

[5] Davidson, James (1994): *Stochastic Limit Theory*, Oxford University Press.

[6] Davidson, James and Robert M. De Jong (1997): "Strong laws for near epoch dependent functions of mixing processes: a synthesis of new results," *Econometric Reviews*, 16, 251-280.

[7] De Jong, Robert M. (1995): "Laws of large numbers for dependent heterogeneous processes," *Econometric Theory*, 11, 347-358.

[8] De Jong, Robert M. (1996): "A strong law of large numbers for triangular mixingale arrays," *Statistics and Probability Letters*, 27, 1-9.

[9] De Jong, Robert M. (1998): "Weak laws of large numbers for mixingales," *Annales d'Economie et de Statistiques*, 51, 209-225.

[10] Hall, Peter and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*, Academic Press.

[11] Hansen, Bruce E. (1991): "Strong laws for dependent heterogeneous processes," *Econometric Theory*, 7, 213-221.

[12] Hansen, Bruce E. (1992): "Erratum," *Econometric Theory*, 8, 421-422.

[13] Halmos, Paul R. (1944): "In general a measure preserving transformation is mixing," *Annals of Mathematics*, 45, 786-792.

[14] Kakutani, Shizuo (1973): "Examples of ergodic measure preserving transformations which are weakly mixing but not strongly mixing," *Recent Advances in Topological Dynamics*, Lecture Notes in Mathematics 318, Springer-Verlag, 143-149.

[15] Katok, A. B. and Stepin, Anatolii M. (1967): "Approximations in ergodic theory," *Russian Mathematical Surveys*, 22, 77-102.

[16] Maruyama, Gisiro (1949): "The harmonic analysis of stationary stochastic processes," *Memoirs of the Faculty of Science Kyushu University Series A*, 4, 45-106.

[17] Petersen, Karl (1983): *Ergodic Theory*, Cambridge University Press.

[18] Rokhlin, Vladimir Abramovich (1948): "A 'general' measure-preserving transformation is not mixing," *Doklady Akademii Nauk SSSR*, 60, 349-351.

[19] Walters, Peter (1982): *An Introduction to Ergodic Theory*, Springer-Verlag.