A Note on Uniform Laws of Averages for Dependent Processes

Andrew Nobel [*]and Amir Dembo [†]

November, 1992

Abstract

If for a 'permissible' family of functions $\mathcal{F}$ and an i.i.d. process $\{X_i\}_{i=0}^{\infty}$

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) - Ef(X_0) \right| = 0$$

with probability one, then the same holds for every absolutely regular (weakly Bernoulli) process having the same marginal distribution. In particular, for any class of sets $\mathcal{C}$ having finite V-C dimension and any absolutely regular process $\{X_i\}_{i=0}^{\infty}$

$$\lim_{n\to\infty} \sup_{C\in\mathcal{C}} \left| \frac{1}{n} \sum_{i=0}^{n-1} I_C(X_i) - \mathbb{P}\{X_0 \in C\} \right| = 0$$

with probability one.

Appears in *Statistics and Probability Letters*, 17 169-172, 1993.

[*]Andrew Nobel is with the Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL. 61801 Email: nobel@assistant.beckman.uiuc.edu

[†]Amir Dembo is with the Statistics and Mathematics Departments, Stanford University, Stanford, CA 94305.

This work was partially supported by the National Science Foundation under Grant NCR-89-14538.

1 Introduction

Uniform laws of averages extend the classical strong law of large numbers (SLLN) from a single real-valued function to a collection of such functions. Let $X_0, X_1, \ldots$ be an i.i.d. sequence of random variables taking values in some measurable space $(\mathcal{X}, \mathcal{S})$, and let $\mathcal{F}$ be a class of real-valued measurable functions on $(\mathcal{X}, \mathcal{S})$. Although the SLLN applies individually to each function $f \in \mathcal{F}$, uniform laws of averages address the question of whether or not the sample averages of functions in $\mathcal{F}$ converge *uniformly* to their expected values; that is, whether or not

$$\lim_{n\to\infty} \sup_{\mathcal{F}} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) - Ef(X_0) \right| = 0 \quad \text{w.p.1.} \tag{1}$$

If the convergence in (1) obtains, $\mathcal{F}$ is said to be a *Glivenko-Cantelli* class. Alternatively, $\mathcal{F}$ is said to satisfy a uniform law of averages with respect to the sequence $X_0, X_1, \ldots$. Note that the term "uniform" here refers to the class $\mathcal{F}$ rather than the underlying sample space. Perhaps the best known result concerning uniform convergence is the classical Glivenko-Cantelli theorem which asserts that the class $\mathcal{F} = \{I_{(-\infty,t]} : t \in \mathbb{R}\}$ of indicator functions of half-infinite intervals is uniformly convergent with respect to any real-valued i.i.d. process $\{X_i\}$.

Vapnik and Cervonenkis (1981), found necessary and sufficient conditions under which a class of functions $\mathcal{F}$ is uniformly convergent with respect to an i.i.d. process $\{X_i\}$. The conditions, which involve the asymptotic behavior of covering numbers associated with the class $\mathcal{F}$ and sample sequences of the process $\{X_i\}$, will not be discussed here. For more details, we refer the reader to Vapnik and Cervonenkis (1981), or Pollard (1984).

In this paper we show that a class of functions satisfying a uniform law of averages with respect to an i.i.d. process will satisfy an analogous law with respect to any absolutely regular process having the same one-dimensional marginal distribution as the original process. The notion of absolute regularity (also known as weakly Bernoulli or β-mixing) is defined in the next section.

Yukich (1986) obtained uniform convergence with specified rates for ϕ-mixing random variables under bracketing entropy conditions on the class of functions $\mathcal{F}$. Philipp (1982) and Massart (1988) obtained invariance principles for α-mixing random variables when the α-coefficients tend to zero at an appropriate rate, and the class $\mathcal{F}$ satisfies bracketing entropy conditions.

Yu (1991) considered uniform convergence, rates of convergence, and functional central limit theorems for absolutely regular random variables. She obtained uniform strong laws for absolutely regular random variables under various conditions on the mixing coefficients $\beta(k)$ and the class $\mathcal{F}$. In Theorem 1 we obtain uniform convergence for absolutely regular random variables as a direct consequence of uniform convergence for independent random variables: no additional conditions are imposed on the class of functions, and no conditions are imposed on the mixing rate of the absolutely regular random variables.

Philipp (1986) gives an overview of invariance principles for dependent random variables. Functional central limit theorems for independent, but not necessarily identically distributed, random variables have been addressed by Dudley and Philipp (1983) and Alexander (1984).

2 Preliminaries

Throughout the remainder of this paper, we consider $\{X_i\}_{i=-\infty}^{\infty}$, a two-sided, strictly stationary sequence of random variables, each taking values in the measurable space $(\mathcal{X}, \mathcal{S})$. We assume the sequence $\{X_i\}$ is defined on the product space $(\mathcal{X}_{-\infty}^{\infty}, \mathcal{S}_{-\infty}^{\infty}, \mathbb{P})$ in the usual way.

Let $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ be any class of real-valued measurable functions on $(\mathcal{X}, \mathcal{S})$. A measurable function $F : \mathcal{X} \to \mathbb{R}$ is said to be an *envelope* for $\mathcal{F}$ if $|f(x)| \leq F(x)$ for every $f \in \mathcal{F}$ and every $x \in \mathcal{X}$. The class $\mathcal{F}$ is said to be *uniformly bounded* if for some $0 \leq K < \infty$, $F = K$ is an envelope for $\mathcal{F}$. If the class $\mathcal{F}$ is uncountable, it must satisfy regularity conditions in order to insure that no problems of measurability will arise when considering quantities such as the supremum in (1). Following Pollard we use the term *permissible* to indicate that a class $\mathcal{F}$ satisfies such conditions (for more details, see Dudley (1978) or Pollard (1984)).

Consider a process $\{X_i\}_{i=-\infty}^{\infty}$ having distribution $\mathbb{P}$ as above. Let $\mathbb{P}_{-\infty}^{0}$ and $\mathbb{P}_{1}^{\infty}$ be semi-infinite marginals of the distribution $\mathbb{P}$. Then $\mathbb{P}_{-\infty}^{0} \times \mathbb{P}_{1}^{\infty}$ represents a product distribution under which the two collections $\{\ldots, X_{-1}, X_0\}$ and $\{X_1, X_2, \ldots\}$ are independent, but are distributed individually as under $\mathbb{P}$. The dependence coefficients $\beta(k), k = 1, 2, \ldots,$ are defined as follows (cf. Volkonskii and Rozanov (1959, 1961)):

$$\beta(k) = \sup\{|\mathbb{P}(A) - \mathbb{P}_{-\infty}^{0} \times \mathbb{P}_{1}^{\infty}(A)| : A \in \sigma(X_{-\infty}^{0}, X_{k}^{\infty})\}.$$

As the coefficients $\beta(k)$ are bounded by 1, nonnegative and non-increasing, $\lim_{k\to\infty} \beta(k)$ always exists. If $\lim_{k\to\infty} \beta(k) = 0$ then the process $\{X_i\}_{i=-\infty}^{\infty}$ is said to be *absolutely regular*. In particular, if the process $\{X_i\}_{i=-\infty}^{\infty}$ is absolutely regular then it is strongly mixing in the ergodic theory sense (cf. Bradley (1986)). Additional properties of the β coefficients are discussed in Bradley (1983); alternative characterizations of the coefficients $\beta(k)$ may be found in Bradley (1986).

We require the following result, due to Steele (1978), which is a consequence of Kingman's subadditive ergodic theorem.

Lemma 1 *Let $\{X_i\}_{-\infty}^{\infty}$ be a stationary ergodic process and let $\mathcal{F}$ be a permissible class of functions with integrable envelope F. Then $\sup_{\mathcal{F}} |\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) - Ef(X_0)|$ converges almost surely to the constant*

$$\eta = \inf_n E\left[\sup_{\mathcal{F}}\left|\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) - Ef(X_0)\right|\right] \tag{2}$$

as n tends to infinity.

3 Uniform Laws of Averages for Absolutely Regular Processes

Consider a stationary process $\{X_i\}_{i=-\infty}^{\infty}$ having distribution $\mathbb{P}$ with one-dimensional marginal P. Let $\mathbb{P}_0 = \Pi_{-\infty}^{\infty} P$ denote the product distribution generated by P; $\mathbb{P}_0$ is the *independent approximation* of $\mathbb{P}$. We require the following elementary inequality.

Lemma 2 *If $A \in \sigma(X_0, X_k, \ldots, X_{(m-1)k})$ then $|\mathbb{P}(A) - \mathbb{P}_0(A)| \leq m\beta(k)$.*

Proof of Lemma 2: The event A is determined by a regularly spaced sequence of random variables. We may apply the definition of $\beta(k)$ repeatedly, first to A itself and then, by integrating, to sections of A determined by progressively longer initial segments of $(x_0, x_k, \ldots, x_{(m-1)k})$. At each stage we add an additional $\beta(k)$ to our estimate of the difference. □

We are now in a position to prove our result.

Theorem 1 *Let $\mathcal{F}$ be a permissible class of functions having envelope $F \in L_1(P)$. If $\mathcal{F}$ satisfies a uniform law of averages with respect to an i.i.d. process having distribution $\mathbb{P}_0 = \Pi_{-\infty}^{\infty} P$ then*

$$\sup_{\mathcal{F}}\left|\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) - Ef(X_0)\right| \to 0$$

with $\mathbb{P}$-probability one for every stationary, absolutely regular stochastic process $\{X_i\}_{i=-\infty}^{\infty}$ having distribution $\mathbb{P}$ with one-dimensional marginal P.

Thus, when investigating whether or not a class of functions is uniformly convergent with respect to some absolutely regular process, we may begin by evaluating the conditions for uniform convergence under the assumption that the process is i.i.d. . In this case the evaluation may be carried out without recourse to the precise nature of the dependence between the observations.

Proof: By arguments similar to those in Pollard (1984, pp.25-26) it is enough to consider a truncated version of the class $\mathcal{F}$ that is uniformly bounded. In addition, we may assume that $Ef = 0$ for every f in $\mathcal{F}$. Then we wish to show that $\sup_{\mathcal{F}} \left| \frac{1}{n} \sum f(X_i) \right| \to 0$ almost surely $\mathbb{P}$. First, note that if $n = km$ then

$$\sup_{\mathcal{F}} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(x_i) \right| \leq \frac{1}{k} \sum_{j=0}^{k-1} \sup_{\mathcal{F}} \left| \frac{1}{m} \sum_{l=0}^{m-1} f(x_{lk+j}) \right| \tag{3}$$

Let $\delta > 0$ and let the constant K be an envelope for $\mathcal{F}$. Taking expectations of both sides in (3) gives

$$\begin{aligned} E\left[\sup_{\mathcal{F}} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \right| \right] &\leq \frac{1}{k} \sum_{j=0}^{k-1} E\left[\sup_{\mathcal{F}} \left| \frac{1}{m} \sum_{l=0}^{m-1} f(X_{lk+j}) \right| \right] \\ &= E\left[\sup_{\mathcal{F}} \left| \frac{1}{m} \sum_{l=0}^{m-1} f(X_{lk}) \right| \right] \qquad (4) \\ &\leq \delta + K\mathbb{P}\left\{ \sup_{\mathcal{F}} \left| \frac{1}{m} \sum_{l=0}^{m-1} f(X_{lk}) \right| > \delta \right\} , \qquad (5) \end{aligned}$$

where the equality in (4) follows by stationarity. Lemma 2 allows us to express the second term of (5) in terms of the probability $\mathbb{P}_0$ plus an additional error term involving $\beta(k)$:

$$\begin{aligned} E\left[\sup_{\mathcal{F}} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \right| \right] &\leq \delta + Km\beta(k) + K\mathbb{P}_0 \left\{ \sup_{\mathcal{F}} \left| \frac{1}{m} \sum_{l=0}^{m-1} f(X_{lk}) \right| > \delta \right\} \\ &= \delta + Km\beta(k) + K\mathbb{P}_0 \left\{ \sup_{\mathcal{F}} \left| \frac{1}{m} \sum_{i=0}^{m-1} f(X_i) \right| > \delta \right\} \qquad (6) \end{aligned}$$

By assumption the last term above tends to zero as $m \to \infty$. Since $\beta(k)$ tends to zero as k tends to infinity, we can select increasing sequences of integers $\{m_i\}$ and $\{k_i\}$ such that $m_i \to \infty,\ k_i \to \infty$ and $m_i\beta(k_i) \to 0$. Let $n_i = m_i k_i$. It follows from (6) and the choice of $m_i,\ k_i$ that

$$\begin{aligned} \inf_n E\left[\sup_{\mathcal{F}} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \right| \right] &\leq \inf_i E\left[\sup_{\mathcal{F}} \left| \frac{1}{n_i} \sum_{j=0}^{n_i-1} f(X_j) \right| \right] \\ &\leq \delta . \qquad (7) \end{aligned}$$

As $\delta > 0$ is arbitrary, we have

$$\inf_n E\left[\sup_{\mathcal{F}}\left|\frac{1}{n}\sum_{i=0}^{n-1} f(X_i)\right|\right] = 0 ,$$

and the result follows from Lemma 1. □

Acknowledgements

The first author would like to thank Prof. Y. Katznelson, Prof. T. Cover and J. Gisladottir for their time and their advice.

References

[1] K. Alexander. Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*, 4:1041–1067, 1984.

[2] R.C. Bradley. Absolute regularity and functions of Markov chains. *Stochastic Process. Appl.*, 14:67–77, 1983.

[3] R.C. Bradley. Basic properties of strong mixing conditions. In E. Eberlein and M.S. Taqqu, editors, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 165–192, Boston, 1986. Birkhäuser.

[4] R.M. Dudley. Central limit theorems for empirical measures. *Ann. Probab.*, 6:899–929, 1978.

[5] R.M. Dudley and W. Philipp. Invariance principles for sums of Banach space valued random variables and empirical processes. *Probab. Theory and Related Fields*, 62:509–552, 1983.

[6] P. Massart. Invariance principles for empirical processes: the weakly dependent case. PhD thesis, University of Paris, 1988.

[7] W. Philipp. Invariance principles for sums of mixing random elements and the multivariate empirical process. In *Limit Theorems in Probability and Statistics*, volume 36, pages 843–873, Veszprém (Hungary), 1982. Colloquia Mathematica Societatis János Bolyai.

[8] W. Philipp. Invariance principles for independent and weakly dependent random variables. In E. Eberlein and M.S. Taqqu, editors, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 225–268. Birkhauser, 1986.

[9] D. Pollard. *Convergence of Stochastic Processes.* Springer, New York, 1984.

[10] J.M. Steele. Empirical discrepancies and subadditive processes. *Ann. Probab.*, 6:118–127, 1978.

[11] V.N. Vapnik and A.Ya. Chervonenkis. Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.*, 26:532–553, 1981.

[12] V.A. Volkonskii and Yu.A. Rozanov. Some limit theorems for random functions, i. *Theory Probab. Appl.*, 4:178–197, 1959.

[13] V.A. Volkonskii and Yu.A. Rozanov. Some limit theorems for random functions, ii. *Theory Probab. Appl.*, 6:186–198, 1961.

[14] B. Yu. Rates of convergence and central limit theorems for empirical processes of stationary mixing sequences. Technical Report 260, U.C. Berkeley, Department of Statistics, 1991.

[15] J.E. Yukich. Rates of convergence for classes of functions: the non-i.i.d. case. *J. Multivariate Anal.*, 20:175–189, 1986.