Elect. Comm. in Probab. **3** (1998) 13–19

ELECTRONIC COMMUNICATIONS in PROBABILITY

THE LAW OF LARGE NUMBERS FOR U–STATISTICS UNDER ABSOLUTE REGULARITY

MIGUEL A. ARCONES
Department of Mathematics
University of Texas
Austin, TX 78712–1082.
email: `arcones@math.utexas.edu`
web: `http://www.ma.utexas.edu/users/arcones/`

submitted September 15, 1997; *revised* March 4, 1998.

AMS 1991 Subject classification: 60F15.
Keywords and phrases: Law of large numbers, U–statistics, absolute regularity.

Abstract
We prove the law of large numbers for U–statistics whose underlying sequence of random variables satisfies an absolute regularity condition (β–mixing condition) under suboptimal conditions.

1 Introduction.

We consider the law of large numbers for U–statistics whose underlying sequence of random variables satisfies a β–mixing condition. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables with values in a measurable space $(S, \mathcal{S})$. Given a kernel h, i.e. given a function h from S^m into $\mathbb{R}$, symmetric in its arguments, the U–statistic with kernel h is defined by

$$U_n(h) := \frac{(n-m)!}{n!} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}). \tag{1.1}$$

We refer to Serfling (1980), Lee (1990), and Koroljuk and Borovskich (1994) for more in U–statistics. For i.i.d.r.v.'s, assuming that $E[|h(X_1, \ldots, X_m)|] < \infty$, Hoeffding (1961; see also Berk, 1966) proved the law of large numbers for U–statistics:

$$\frac{(n-m)!}{n!} \sum_{1 \le i_1 < \cdots < i_m \le n} (h(X_{i_1}, \ldots, X_{i_m}) - E[h(X_{i_1}, \ldots, X_{i_m})]) \to 0 \text{ a.s.} \tag{1.2}$$

Several authors have studied limit theorems for U–statistics under different dependence conditions. Sen (1972), Yoshihara (1976) and Denker and Keller (1983) proved a central limit theorem and a law of the iterated logarithm for U–statistics under different types of dependence conditions. Qiying (1995) and Aaronson, Burton, Dehling, Gilat, Hill, and Weiss (1996) studied the law of large numbers for U–statistics for stationary sequences of dependent r.v.'s.

Aaronson, Burton, Dehling, Gilat, Hill, and Weiss (1996) gave several sufficient conditions for the law of large numbers over a ergodic stationary sequence of r.v.'s. It is shown in this paper (Example 4.1) that even the weak law of large numbers for U–statistics is not true just assuming finite first moment and ergodicity, that is the ergodic theorem is not true for U–statistics. Thus further conditions must be imposed.
Qiying (1995) considered the law of large numbers under ϕ^*–mixing. But, there is a gap in his proofs. In Equation (11), he claims that

$$\sum_{k=1}^{\infty} 2^{-2k} \sup_{m\geq 2} E|h(X_1,X_m)|^2 I_{(|h(X_1,X_m)|\leq 2^{2k})} \leq A \sup_{m\geq 2} E|h(X_1,X_m)|,$$

where A is an arbitrary constant. Qiying is using that there exist a universal constant A such that for any sequence of r.v.'s $\{\xi_m\}$,

$$\sum_{k=1}^{\infty} 2^{-2k} \sup_{m\geq 2} E\xi_m^2 I_{(|\xi_m|\leq 2^{2k})} \leq A \sup_{m\geq 2} E|\xi_m|.$$

This claim is not true. Let us take ξ_m such that $\Pr(\xi_m = 2^{2m}) = 2^{-2m}$ and $\Pr(\xi_m = 0) = 1 - 2^{-2m}$. Then,

$$\sup_{m\geq 2} E|\xi_m| = 1$$

and

$$\sum_{k=1}^{\infty} 2^{-2k} \sup_{m\geq 2} E\xi_m^2 I_{(\xi_m\leq 2^{2k})} \geq \sum_{k=1}^{\infty} 2^{-2k} E\xi_k^2 I_{(\xi_k\leq 2^{2k})} = \infty.$$

A similar comment applies to Equation (11) in Qiying (1995).

Instead of using ϕ^*–mixing, we use β–mixing. ϕ^*–mixing is one of the stronger mixing conditions. The ϕ^*–mixing coefficient is bigger than the β–mixing. The dependence condition we will consider is known as absolute regularity. Given a strictly stationary sequence $\{X_i\}_{i=1}^{\infty}$ with values in a measurable space $(S,\mathcal{S})$, let $\sigma_1^l = \sigma(X_1,\ldots,X_l)$ and let $\sigma_l^{\infty} = \sigma(X_l, X_{l+1},\ldots)$, the β–mixing sequence is defined by

$$\beta_k := 2^{-1} \sup\{\sum_{i=1}^{I}\sum_{j=1}^{J} |\Pr(A_i\cap B_j) - \Pr(A_i)\Pr(B_j)| : \{A_i\}_{i=1}^{I} \text{ is a partition in } \sigma_1^l \tag{1.3}$$

$$\text{and } \{B_j\}_{j=1}^{J} \text{ is a partition in } \sigma_{k+l}^{\infty}, l\geq 1\}.$$

We refer to Ibragimov and Linnik (1971) and Doukhan (1994) for more information in this type of dependence condition.
We present the following theorem:

Theorem 1. *Let $\{X_i\}_{i=1}^{\infty}$ be a strictly stationary sequence of random variables with values in a measurable space $(S,\mathcal{S})$. Let $h : S^m \to I\!R$ be a symmetric function. Suppose that at least one of the following conditions is satisfied:*
(i) For some $\delta > 2$, $\sup_{1\leq i_1<\cdots<i_m<\infty} E[|h(X_{i_1},\ldots,X_{i_m})|^{\delta}] < \infty$ and $\beta_n \to 0$.
(ii) For some $0 < \delta \leq 1$ and some $r > 2\delta^{-1}$, $\sup_{1\leq i_1<\cdots<i_m<\infty} E[|h(X_{i_1},\ldots,X_{i_m})|^{1+\delta}] < \infty$ and $\beta_n = O((\log n)^{-r})$

(iii) For some $0 < \delta \le 1$ *and some* $r > 0$,
$\sup_{1 \le i_1 < \cdots < i_m < \infty} E[|h(X_{i_1}, \dots, X_{i_m})|(\log^+ |h(X_{i_1}, \dots, X_{i_m})|)^{1+\delta}] < \infty$ *and* $\beta_n = O(n^{-r})$.
Then,

$$n^{-m} \sum_{1 \le i_1 < \cdots < i_m \le n} (h(X_{i_1}, \dots, X_{i_m}) - E[h(X_{i_1}, \dots, X_{i_m})]) \to 0 \text{ a.s.}$$

Observe that the conditions in the previous theorem are very close to being optimal.

2 Proofs.

c will denote an arbitrary constant that may change from line to line. Given a r.v. Y, we define $\|Y\|_p = (E[|Y|])^{1/p}$, for and $1 \le p < \infty$; and we define $\|Y\|_\infty = \inf\{t > 0 : |Y| \le t \text{ a.s.}\}$.
We need to recall some notation on U–statistics. We define

$$\pi_{k,m} h(x_1, \dots, x_k) = (\delta_{x_1} - P) \cdots (\delta_{x_k} - P) P^{m-k} h, \tag{2.1}$$

where $Q_1 \cdots Q_m h = \int \cdots \int h(x_1, \dots, x_m)\, dQ_1(x_1) \cdots dQ_m(x_m)$. We say that a kernel h is P–canonical if it is symmetric and

$$E[h(x_1, \dots, x_{m-1}, X_m)] = 0 \quad \text{a.s.} \tag{2.2}$$

It is known that

$$U_n(h) = \sum_{k=0}^{m} \binom{m}{k} U_n(\pi_{k,m} h). \tag{2.3}$$

Previous inequality is known as the Hoeffding decomposition (Hoeffding, 1948, Section 5). Observe that the Hoeffding decomposition is a decomposition in U–statistics of canonical kernels ($\pi_{k,m} h$ is a canonical kernel).
The β–mixing condition allows to compare probabilities of the initial sequence with respect to a sequence of r.v.'s with independent blocks. Explicitly, we have the following lemma:

Lemma 2. *Let* $\{X_j\}_{j=1}^\infty$ *be a stationary sequence of r.v.'s with values in a measurable space* $(S, \mathcal{S})$. *Let* f *be a measurable function on* S^m. *Let* $(m(i,j))_{\substack{1 \le i \le k \\ 1 \le j \le r_i}}$ *be integers such that*

$$m(1,1) < \cdots < m(1,r_1) < m(2,1) < \cdots < m(2,r_2) < \cdots < m(k,1) < \cdots < m(k,r_k).$$

Let $r = \sum_{i=1}^k r_i$. *Let* $\{\xi_j\}_{j=1}^r$ *be a sequence of identically distributed r.v.'s with the distribution of* X_1 *such that*

$$\begin{aligned} &\mathcal{L}(\xi_{m(1,1)}, \dots, \xi_{m(1,r_1)}, \xi_{m(2,1)}, \dots, \xi_{m(2,r_2)}, \cdots, \xi_{m(k,1)}, \dots, \xi_{m(k,r_k)}) \\ &\quad = \mathcal{L}(X_{m(1,1)}, \dots, X_{m(1,r_1)}) \otimes \cdots \otimes \mathcal{L}(X_{m(k,1)}, \dots, X_{m(k,r_k)}). \end{aligned}$$

Then,
(i)

$$|E[f(X_{m(1,1)}, \dots, X_{m(k,r_k)})] - E[f(\xi_{m(1,1)}, \dots, \xi_{m(k,r_k)})]| \le 2 \sum_{i=1}^{k-1} \beta(m(i+1,1) - m(i,r_i)) \|f\|_\infty.$$

(ii) If $1 < p < \infty$,

$$|E[f(X_{m(1,1)}, \dots, X_{m(k,r_k)})] - E[f(\xi_{m(1,1)}, \dots, \xi_{m(k,r_k)})]|$$
$$\leq 4(\sum_{i=1}^{k-1} \beta(m(i+1,1) - m(i,r_i)))^{(p-1)/p}$$
$$\times \max(\|f(X_{m(1,1)}, \dots, X_{m(k,r_k)})\|_p, \|f(\xi_{m(1,1)}, \dots, \xi_{m(k,r_k)})\|_p).$$

Part (i) in previous lemma follows directly from the definition of β mixing (see the characterization of β–mixing on page 193 in Volkonskii and Rozanov, 1961) and induction (see Lemma 2 in Eberlein, 1984). Part (ii) follows directly from part (i) (see for example Lemma 2 in Arcones, 1995).

The following lemma gives a bound on the second moment of a U–statistic over a degenerated kernel.

Lemma 3. *There is a universal constant c, depending only on m, such that for each canonical kernel h and each $p > 2$,*

$$E\left[\left(\sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m})\right)^2\right] \leq cn^m M^2 (1 + \sum_{j=1}^{n-1} j^{m-1} \beta_j^{(p-2)/p})$$

where

$$M := \sup_{1 \leq i_1 < \cdots < i_m < \infty} (E[|h(X_{i_1}, \dots, X_{i_m})|^p]^{1/p}.$$

Proof. We have that

$$E\left[\left(\sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m})\right)^2\right]$$
$$\leq \sum_{\sigma \in \Gamma(2m)} \sum_{1 \leq i_1 \leq \cdots \leq i_{2m} \leq n} |E[h(X_{i_{\sigma(1)}}, \dots, X_{i_{\sigma(m)}}) h(X_{i_{\sigma(m+1)}}, \dots, X_{i_{\sigma(2m)}})]|$$

where $\Gamma(2m)$ is the collection of all permutations of $2m$ elements. Let $j_1 = i_2 - i_1$, let $j_l = \min(i_{2l-1} - i_{2l-2}, i_{2l} - i_{2l-1})$ for $2 \leq l \leq m-1$, and let $j_m = i_{2m} - i_{2m-1}$. If $j_1 = \max(j_1, \dots, j_m)$, we compare the initial sequence $\{X_1, \dots, X_n\}$ with the one having the independent blocks $\{i_1\}$, $\{i_2, \dots, i_{2m}\}$ and the same block distribution. We claim that by Lemma 2, we get that

$$\sum_{\substack{1 \leq i_1 \leq \cdots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} |E[h(X_{i_{\sigma(1)}}, \dots, X_{i_{\sigma(m)}}) h(X_{i_{\sigma(m+1)}}, \dots, X_{i_{\sigma(2m)}})]|$$
$$\leq cn^m M^2 (1 + \sum_{k=1}^{n-1} k^{m-1} \beta_k^{(p-2)/p}).$$

Observe that if $i_2 = i_1 + k$, i_1 can take at most n different values. Assume that $i_3 - i_2 \le i_4 - i_3$, then $i_3 - i_2 \le k$, so i_3 can take at most k values and i_4 can take at most n values. If $i_4 - i_3 \le i_3 - i_2$, then i_3 can take at most n values and i_4 can take at most k values. Proceeding in this way we obtain that the possible values for the variables $i_1 \le \cdots \le i_{2m}$ (under the assumptions $1 \le i_1 \le \cdots \le i_{2m} \le n$ and $k = j_1 \ge j_2, \dots, j_m$) is bounded by $n^m k^{m-1}$.
If $j_l = \max(j_1, \dots, j_m)$, for some $2 \le l \le m-1$, we compare the initial sequence with the one with the independent blocks $\{i_1, \dots, i_{2l-2}\}$, $\{i_{2l-1}\}$ and $\{i_{2l}, \dots, i_{2m}\}$. A similar argument applies to this case.
If $j_m = \max(j_1, \dots, j_m)$, we compare the initial sequence with the one with the independent blocks $\{i_1, \dots, i_{2m-1}\}$ and $\{i_{2m}\}$. □

Now, we are ready to prove Theorem 1.

Proof of Theorem 1. First, we consider the case (iii). We may assume that $0 < r < m$. A standard argument gives that it suffices to show that for each $\alpha > 1$,

$$n_k^{-m} \sum_{1 \le i_1 < \cdots < i_m \le n_k} h(X_{i_1}, \dots, X_{i_m}) \to E[h(X_{i_1}, \dots, X_{i_m})] \text{ a.s.,} \tag{2.4}$$

where $n_k = [\alpha^k]$. Now, by the Hoeffding decomposition, it suffices to prove (2.4) for canonical kernels. We are going to prove (2.4) by induction on m. The case $m = 1$ is the ergodic theorem (see for example Theorem 6.21 in Breiman, 1992).
It is easy to see that it suffices to show that

$$n_k^{-m} \sum_{i_m = n_{k-1}+1}^{n_k} \sum_{1 \le i_1 < \cdots < i_{m-1}}^{i_m - 1} h(X_{i_1}, \dots, X_{i_m}) \to 0 \text{ a.s.}$$

Take $p > 2$ and $\tau > 0$ such that

$$2\tau(p-1) < r(p-2). \tag{2.5}$$

Next we prove that

$$n_k^{-m} \sum_{i_m = n_{k-1}+1}^{n_k} \sum_{1 \le i_1 < \cdots < i_{m-1}}^{i_m - 1} h(X_{i_1}, \dots, X_{i_m}) I_{|h(X_{i_1}, \dots, X_{i_m})| \ge n_k^\tau} \to 0 \text{ a.s.} \tag{2.6}$$

We have that

$$E[\sum_{k=1}^{\infty} n_k^{-m} \sum_{i_m = n_{k-1}+1}^{n_k} \sum_{1 \le i_1 < \cdots < i_{m-1}}^{i_m - 1} |h(X_{i_1}, \dots, X_{i_m})| I_{|h(X_{i_1}, \dots, X_{i_m})| \ge n_k^\tau}] \tag{2.7}$$
$$\le c \sum_{k=1}^{\infty} (\log n_k^\tau)^{-\delta - 1} < \infty.$$

Therefore, (2.6) follows.
Thus, we must prove that

$$n_k^{-m} \sum_{i_m = n_{k-1}+1}^{n_k} \sum_{1 \le i_1 < \cdots < i_{m-1}}^{i_m - 1} (h(X_{i_1}, \dots, X_{i_m}) I_{|h(X_{i_1}, \dots, X_{i_m})| < n_k^\tau} \tag{2.8}$$

$$-E[h(X_{i_1},\dots,X_{i_m})I_{|h(X_{i_1},\dots,X_{i_m})|<n_k^\tau}] \to 0 \text{ a.s.}$$

Using that

$$\begin{aligned} &\delta_{x_1}\cdots\delta_{x_m} - P^m \\ &= (\delta_{x_1}-P)P^{m-1} + P(\delta_{x_2}-P)P^{m-2} + \cdots + P^{m-1}(\delta_{x_m}-P) \\ &+(\delta_{x_1}-P)(\delta_{x_2}-P)P^{m-2} + \cdots + (\delta_{x_1}-P)\cdots(\delta_{x_m}-P), \end{aligned}$$

we get that (2.8) decomposes in sums of terms of the form

$$n_k^{-m}\sum_{i_m=n_{k-1}+1}^{n_k}\sum_{1\le i_1<\cdots<i_{m-1}}^{i_m-1} P^{j_0}(\delta_{x_{i_{\alpha_1}}}-P)P^{j_1}\cdots(\delta_{x_{i_{\alpha_l}}}-P)P^{j_l}hI(|h|<n_k^\tau), \tag{2.9}$$

where $1\le\alpha_1<\cdots<\alpha_l\le m$, $1\le l\le m$, $0\le j_0,\dots,j_l$ and $l+j_0+\cdots+j_l=m$.
For $1\le l\le m-1$, using that h is canonical,

$$\begin{aligned} &P^{j_0}(\delta_{x_{i_{\alpha_1}}}-P)P^{j_1}\cdots(\delta_{x_{i_{\alpha_l}}}-P)P^{j_1}hI(|h|<n_k^\tau) \\ &= P^{j_0}(\delta_{x_{i_{\alpha_1}}}-P)P^{j_1}\cdots(\delta_{x_{i_{\alpha_l}}}-P)P^{j_l}hI(|h|\ge n_k^\tau). \end{aligned}$$

Thus, (2.9) is bounded in absolute value by

$$n_k^{-m}\sum_{1\le i_1<\cdots<i_m\le n_k} P^{j_0}(\delta_{x_{i_{\alpha_1}}}+P)P^{j_1}\cdots(\delta_{x_{i_{\alpha_l}}}+P)P^{j_l}|h|I(|h|\ge n_k^\tau).$$

Again, decomposing terms, we get that we have to deal with

$$\begin{aligned} &n_k^{-m}\sum_{1\le i_1<\cdots<i_m\le n_k} P^{j_0}\delta_{x_{i_{\alpha_1}}}P^{j_1}\cdots\delta_{x_{i_{\alpha_l}}}P^{j_l}|h|I(|h|\ge n_k^\tau) \\ &\le cn_k^{-l}\sum_{1\le i_1<\cdots<i_l\le n_k} P^{j_0}\delta_{x_{i_1}}P^{j_1}\cdots\delta_{x_{i_l}}P^{j_l}|h|I(|h|\ge n_k^\tau), \end{aligned}$$

which goes to zero a.s. by the induction hypothesis.
To get the case $l=m$,

$$n_k^{-m}\sum_{i_m=n_{k-1}+1}^{n_k}\sum_{1\le i_1<\cdots<i_{m-1}}^{i_m-1}\pi_{m,m}(hI(|h|<n_k^\tau)(X_{i_1},\dots,X_{i_m})\to 0 \quad\text{a.s.} \tag{2.10}$$

By Lemma 3,

$$\begin{aligned} &E[(n_k^{-m}\sum_{i_m=n_{k-1}+1}^{n_k}\sum_{1\le i_1<\cdots<i_{m-1}}^{i_m-1}\pi_{m,m}(hI(|h|<n_k^\tau)(X_{i_1},\dots,X_{i_m}))^2] \qquad (2.11) \\ &\le cn_k^{-m}(1+\sum_{j=1}^{n_k}j^{m-1}\beta_j^{(p-2)/p})(\sup_{i_1<\cdots<i_m}E[|h(X_{i_1},\dots,X_{i_m})|^pI(|h|<n_k^\tau)])^{2/p} \\ &\le cn_k^{-r(p-2)p^{-1}+\tau(p-1)2p^{-1}}, \end{aligned}$$

which by (2.5) implies (2.10).

The proof in the case (ii) follows similarly, instead of truncating at n_k^τ we truncate at $k^{(1+\epsilon)/\delta}$, where $2^{-1}\delta r - 1 > \epsilon > 0$. We take $p > 2$ such that $r > 2(p-1-\delta)(1+\epsilon)\delta^{-1}(p-2)^{-1}$. It is easy to see that (2.7) and (2.11) hold.
In the case (iii), we truncate at n_k and we take $p = \delta$. It is easy to see that (2.11) is bounded by

$$cn_k^{-m}(1 + \sum_{j=1}^{n_k} j^{m-1}\beta_j^{(p-2)/p}),$$

which goes to zero. □

References

[1] Aaronson, J.; Burton, R.; Dehling, H.; Gilat, D.; Hill, T. and Weiss, B. (1996). Strong laws for L– and U–statistics. *Trans. Amer. Math. Soc.* **348** 2845–2866.

[2] Arcones, M. A. (1995). On the central limit theorem for U–statistics under absolute regularity. *Statist. Probab. Lett.* **24** 245–249.

[3] Berk, R.H. (1966). Limiting behavior of posterior distributions where the model is incorrect. *Ann. Math. Statis.* **37** 51–58.

[4] Breiman, L. (1992). *Probability.* SIAM, Philadelphia.

[5] Denker, M. and Keller, G. (1983). On U–statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahsrsch. verw. Geb.* **64** 505–522.

[6] Doukhan, P. (1994). *Mixing: Properties and Examples.* Lectures Notes in Statistics, 85. Springer–Verlag, New York.

[7] Eberlein, E. (1984). Weak rates of convergence of partial sums of absolute regular sequences. *Statist. Probab. Lett.* **2** 291–293.

[8] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

[9] Hoeffding, W. (1961). The strong law of large numbers for U–statistics. Inst. Statist. Univ. of North Carolina, Mimeo Report, No. 302.

[10] Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff Publishing, Groningen, The Netherlands.

[11] Koroljuk, V. S. and Borovskich, Yu. V. (1994). *Theory of U–statistics.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

[12] Lee, A. J. (1990). *U–statistics, Theory and Practice.* Marcel Dekker, Inc., New York.

[13] Qiying, W. (1995). The strong law of U–statistics with ϕ^*–mixing samples. *Stat. Probab. Lett.* **23** 151–155.

[14] Sen, P. K. (1974) Limiting behavior of regular functionals of empirical distributions for stationary $*$–mixing processes. *Z. Wahrsch. verw. Gebiete* **25** 71–82.

[15] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[16] Volkonskii, V. A. and Rozanov, Y. A. (1961). Some limit theorems for random functions II. *Theor. Prob. Appl.* **6** 186–198.

[17] Yoshihara, K. (1976). Limiting behavior of U–statistics for stationary, absolute regular processes. *Z. Wahrsch. verw. Geb.* **35** 237–252.