Prob. Th. Rel. Fields 81, 235–238 (1989)

Probability Theory and Related Fields



# Exponential Convergence Under Mixing*

Roberto H. Schonmann**

Instituto de Matemática e Estatística, USP Caixa Postal 20570 São Paulo, SP, Brasil

**Summary.** We show that for a $\phi$-mixing stationary sequence of bounded random variables, the average of the first $n$ variables converges exponentially fast with $n$ to the mean value of these random variables.

A recurring pattern in the development of probability theory is the establishment of results first for sequences of independent random variables and later the generalization of such results for "weakly dependent" sequences. Usually the notions of weak dependence are expressed by various types of mixing conditions (see for instance [3, 5] or [7]).

Surprisingly enough there seems to be a lack in the literature of generalizations of Cramer-Chernoff [1, 2] large deviation theorem to the case of weakly dependent sequences of random variables.

In this note we provide a result of this type for $\phi$-mixing (also called uniformly mixing in [5], p. 308) stationary sequences of bounded random variables. Let $\ldots, \xi_{-2}, \xi_{-1}, \xi_0, \xi_1, \xi_2, \ldots$ be a stationary sequence of random variables, bounded above by $A<\infty$, i.e., $P(\xi_0 \geqq A)=0$, and with finite first moment, i.e., $E(\xi_0):=\rho \neq \mp\infty$, where $E(\cdot)$ denotes expectation. Define $X_n=n^{-1}(\xi_1+\ldots+\xi_n)$. For $m \leqq n$ let $\mathscr{F}_m^n$ be the $\sigma$-field generated by $\xi_m, \xi_{m+1}, \ldots, \xi_n$. Set also $\mathscr{F}_{-\infty}^n = \bigcup_{m \leqq n} \mathscr{F}_m^n$, $\mathscr{F}_m^\infty = \bigcup_{n \geqq m} \mathscr{F}_m^n$, and define

$$\phi(l) := \sup_{\substack{E \in \mathscr{F}_{-\infty}^0,\, F \in \mathscr{F}_l^\infty \\ P(E)>0}} |P(F \mid E) - P(F)|.$$

**Theorem.** *Under the conditions above, if $\phi(l) \to 0$ as $l \to \infty$, then for any $\varepsilon>0$ there exist $\gamma>0$ and $C<\infty$ such that*

$$P(X_n \geqq \rho+\varepsilon) \leqq C e^{-\gamma n}.$$

* Work supported by the U.S. Army Research Office through the Mathematical Sciences Institute at Cornell and a NSF Grant to Cornell University

** On leave from São Paulo University, Brazil

Obviously if $\xi_0$ is bounded below ($P(\xi_0 \leqq B)=0$ for some $B>-\infty$) then for any $\varepsilon>0$, $P(X_n \leqq \rho-\varepsilon)$ decays also exponentially with $n$.

The Theorem above will be proven by a block (renormalization) argument. By taking large enough blocks of indices we can use the ergodicity of the sequence to assure that the average inside each block is probably close to $\rho$. Taking these blocks far from each other, the mixing condition can be used to show that even disasters in many blocks do not increase very much the probability of a disaster in the next block and we can then consider the blocks as so nearly independent that very rough exponential estimates suffice to finish the proof.

*Proof of the Theorem.* Given two positive integers $L$ and $N$ we define the blocks

$$R_1 := \{1, \ldots, L\},$$
$$R_j := R_1 + NL(j-1) = \{1+NL(j-1), \ldots, L+NL(j-1)\}, \quad j=2,3,\ldots.$$

And set

$$S_k := \bigcup_{i=1}^{k} R_i,$$
$$Y_j := L^{-1} \sum_{i \in R_j} \xi_i,$$
$$Z_k := (kL)^{-1} \sum_{i \in S_k} \xi_i = k^{-1} \sum_{j=1}^{k} y_j.$$

Let $\delta := 3^{-2(A-\rho)/\varepsilon}$. By the ergodicity of the sequence $(\xi_i)$ which is implied by its mixing property [7], there exists $L$ such that

$$P(Y_1 \geqq \rho + \varepsilon/2) \leqq \delta/2. \tag{1}$$

By the mixing property, $\lim_{l \to \infty} \phi(l)=0$, there is $N$ such that $\phi(NL) \leqq \delta/2$ and then, for any choice of $j_1 < j_2 < \ldots < j_s < j$

$$P(Y_j \geqq \rho + \varepsilon/2 \mid Y_{j_r} \geqq \rho + \varepsilon/2, r=1, \ldots, s) \leqq \delta. \tag{2}$$

Now, by induction, it follows from (1) and (2) that

$$P(Y_{j_r} \geqq \rho + \varepsilon/2, r=1, \ldots, s) \leqq \delta^s. \tag{3}$$

If the events $\{Y_j \geqq \rho + \varepsilon/2\}$ occur for less than $[k\varepsilon/(2(A-\rho))]$ indices $j \in \{1, \ldots, k\}$, then

$$\begin{aligned} kZ_k &\leqq A \cdot (k\varepsilon/(2(A-\rho))) + (k - k\varepsilon/(2(A-\rho)))(\rho+\varepsilon/2) \\ &= k(\rho+\varepsilon) - k\varepsilon^2/(4(A-\rho)) \\ &\leqq k(\rho+\varepsilon). \end{aligned}$$

Hence

$P(Z_k \geqq \rho+\varepsilon) \leqq P(\{Y_j \geqq \rho+\varepsilon/2\}$ for at least $[k\varepsilon/(2(A-\rho))]$ indices $j \in \{1, \ldots, k\})$.

But there are no more than $2^k$ ways of choosing $[k\varepsilon/(2(A-\rho)]$ out of $k$. Therefore, using (3)

$$P(Z_k \geqq \rho+\varepsilon) \leqq 2^k \delta^{[k\varepsilon/(2(A-\rho))]} \leqq \delta^{-1} \cdot (2/3)^k. \tag{4}$$

Consider now the translations of $S_k$:

$$S_{k,q} := S_k + qL, \qquad q = 0, 1, 2, \ldots, N-1.$$

And let $Z_{k,q}$ be the average of the $\xi_i$ in $S_{k,q}$:

$$Z_{k,q} := (kL)^{-1} \sum_{i \in S_{k,q}} \xi_i.$$

Then, by the stationarity of $(\xi_i)$, it follows from (4) that

$$P(Z_{k,q} \geqq \rho+\varepsilon) \leqq \delta^{-1} \cdot (2/3)^k, \qquad q = 0, 1, 2, \ldots, N-1.$$

Now, if $n = kNL$, then

$$P(X_n \geqq \rho+\varepsilon) \leqq P\left(\bigcup_{q=0}^{N-1} \{Z_{k,q} \geqq \rho+\varepsilon\}\right) \leqq N\delta^{-1}(2/3)^k = Ce^{-\gamma n},$$

where $\gamma > 0$ and $C < \infty$ depend on $N$ and $L$ (and hence on $\varepsilon$) but not on $k$. This finishes the proof in case $n$ is a multiple of $NL$. The extension to general $n$ is routine. □

One can observe that the only uniformity in the mixing condition, that we needed for the proof was that on $E$. We chose to state the theorem in the form above simply to match the usual definition of $\phi$-mixing. Besides relaxing the uniformity in $F$ we can weaken further the mixing condition by restricting $E$ to the type of events (depending on blocks) that we really need in the proof; a mixing condition of this type was indeed needed for an application to a Gibbs measure in [8].

Generalizations to higher dimensional index sets, i.e., to random fields $\{\xi_i)_{i \in Z^d}$ are straightforward, provided that one has good enough mixing. (See the remarks after the proof of Lemma 2 in [8], for more details).

The result in the Theorem above is weaker than in Cramer-Chernoff Theorem for the i.i.d. case since we are not able to identify the rate of exponential decay in general. Nevertheless, under extra assumptions we can provide a complete generalization of Cramer-Chernoff Theorem. For this purpose we say that the stationary sequence $(\xi_i)$ is associated (or is FKG) if for any natural $n$ and any pair of functions $f, g: \mathbb{R}^n \to \mathbb{R}$ which are coordinatewise non-decreasing,

$$E(f(\xi_{i_1}, \ldots, \xi_{i_n})) \cdot E(g(\xi_{i_1}, \ldots, \xi_{i_n})) \leqq E(f(\xi_{i_1}, \ldots, \xi_{i_n})\, g(\xi_{i_1}, \ldots, \xi_{i_n})),$$

for any $i_1 < \ldots < i_n$. Combining the results and methods in [4] and [6] with the Theorem in this paper, it follows that if the random variables $\xi_i$ are bounded on both sides by some $A < \infty$, i.e., $P(|\xi_0| \geqq A) = 0$, and $(\xi_i)$ is associated, then there exists a convex function $\varphi$: $R \to [0, \infty]$ such that $\varphi(x) = 0$ if and only if $x = \rho$, $\varphi(x) = \infty$ if $x \notin [-A, A]$, and for any pair $a < b$,

$$\lim_{n \to \infty} n^{-1} \log P(X_n \in (a, b)) = - \inf_{a < x < b} \varphi(x).$$

Moreover also the following limit exists for any $h \in \mathbb{R}$

$$\lim_{n \to \infty} n^{-1} \log E(\exp(hnX_n)) := \pi(h),$$

and $\varphi$ and $\pi$ are dual convex functions:

$$\varphi(x) = \sup_h \{xh - \pi(h)\},$$

$$\pi(h) = \sup_x \{xh - \varphi(x)\}.$$

(In [6] the $\xi_i$ were assumed to take values on a finite space, but this was so only because of our motivation there and the same proofs apply to the case of bounded continuous variables). The reader is referred to [4] and references there for another sufficient condition for these results to hold (superconvolutiveness).

*Acknowledgement.* It is a pleasure to thank F. Avram, Y. Derriennic, R. Durrett, G. Grimmett and H. Kesten for discussions.

## References

1. Chernoff, H.: A measure of asymptotic efficiency for tests of hypothesis based on sums of observations. Ann. Math. Stat. **23**, 493–507 (1952)
2. Cramer, H.: Sur un nouveau théorèm limite de la théorie des probabilités. Colloqium on theory of probability. Paris: Hermann (1937)
3. Dürr, D., Goldstein, S.: Remarks on the central limit theorem for weakly dependent random variables. In: Dolb, A., Eckmann, B. (eds.) Stochastic processes, mathematics and physics. Lect. Notes Math. Vol. 1158, pp. 104–118. Berlin Heidelberg: Springer 1958
4. Grimmett, G.: Large deviations in subadditive processes and first passage percolation. In: Durrett, R. (ed.) Particle systems, random media and large deviations. Contemp. Math., Vol. 41. pp. 175–194. Providence: AMS 1985
5. Ibragimov, I.A., Linnik, Yu.V.: Independent and stationary sequences of random variables. Groningen: Noordhoff 1969
6. Lebowitz, J.L., Schonmann, R.H.: Pseudo free energies and large deviations for non-Gibbsian FKG measures. Probab. Th. Rel. Fields **77**, 49–64 (1988)
7. Rosenblatt, M.: Dependence and asymptotic independence for random processes. In: Rosenblatt, M. (ed.) Studies in Probability Theory, MAA Studies in Mathematics, vol. 18, pp. 24–45 1978
8. Schonmann, R.H.: Second order large deviation estimates for ferromagnetic systems in the phase coexistence region. Commun. Math. Phys. **112**, 409–422 (1987)

Received December 16, 1987