ON THE GLIVENKO-CANTELLI THEOREM FOR WEIGHTED EMPIRICALS BASED ON INDEPENDENT RANDOM VARIABLES

BY RADHEY S. SINGH

Michigan State University

For $X_1, \cdots, X_n$ independent real valued random variables and for $\alpha \in [0, 1]$, let $F_j(x) = \alpha P[X_j < x] + (1-\alpha)P[X_j \leqq x]$ and $Y_j(x) = \alpha I_{[X_j < x]} + (1-\alpha)I_{[X_j \leqq x]}$, where I_A is the indicator function of the set A. For numbers $w_1, w_2, \cdots, w_n$, let $D_n = \sup_{x,\alpha} \max_{N \leqq n} |\sum_1^N w_j(Y_j(x) - F_j(x))|$. We will obtain an exponential bound for $P[D_n \geqq a]$ and a rate for *almost sure* convergence of D_n. When $w_j \equiv 1$ the bound and the rate become, respectively, $4a \exp\{-2((a^2/n) - 1)\}$ and $O((n \log n)^{\frac{1}{2}})$.

1. Introduction. Let $X_1, \cdots, X_n$ be independent real valued random variables. If $X_1, \cdots, X_n$ are *identically* distributed, Theorem 2 of Kiefer (1961) leads to

$$(1)\qquad \sup_x |\textstyle\sum_1^n (I_{[X_j < x]} - P[X_1 < x])| = O((n \log\log n)^{\frac{1}{2}}) \quad \text{w.p. } 1\,,$$

where (and hereinafter) indicator of a set A is denoted by I_A and convergence is wrt $n \to \infty$. The analogue of (1) in the non-identically distributed case would be

$$(2)\qquad \sup_x |\textstyle\sum_1^n (I_{[X_j < x]} - P[X_j < x])| = O((n \log\log n)^{\frac{1}{2}}) \quad \text{w.p. } 1\,,$$

but this is unproved. Neither of the two proofs given in Kiefer (1961) for (1) works for (2). Nor are we able to supply a proof here. However, using a simple proof, we derive a result whose specialization shows that, if for $\alpha \in [0, 1]$

$$F_j(x) = \alpha P[X_j < x] + (1-\alpha)P[X_j \leqq x]$$

and

$$Y_j(x) = \alpha I_{[X_j < x]} + (1-\alpha)I_{[X_j \leqq x]}\,,$$

then

$$(3)\qquad \sup_{x,\alpha} \max_{N \leqq n} |\textstyle\sum_1^N (Y_j(x) - F_j(x))| = O((n \log n)^{\frac{1}{2}}) \quad \text{w.p. } 1\,.$$

THEOREM. *Let $w_1, \cdots, w_n$ be any numbers. Set $||\mathbf{w}_n|| = \sum_1^n |w_j|$ and $||\mathbf{w}_n||_2^2 = \sum_1^n w_j^2$. For any sequence $\{a_n, n \geqq 1\}$ for which $a_n \geqq ||\mathbf{w}_n||_2$ and*

$$(4)\qquad \textstyle\sum_1^\infty \left\{a_n \frac{||\mathbf{w}_n||}{||\mathbf{w}_n||_2^2} \exp\left(-2\left(\frac{a_n}{||\mathbf{w}_n||_2}\right)^2\right)\right\} < \infty\,,$$

we have

$$(5)\qquad D_n = \sup_{x,\alpha} \max_{N \leqq n} |\textstyle\sum_1^N w_j(Y_j(x) - F_j(x))| = O(a_n) \quad \text{w.p. } 1\,.$$

2. Proof of Theorem. We assume, without loss of generality, that $w_j \geqq 0$

Received April 16, 1974; revised July 22, 1974.

AMS 1970 *subject classifications.* Primary 60F10; Secondary 60F15.

Key words and phrases. Glivenko-Cantelli theorem, weighted empiricals, independent non-identically distributed, Borel-Cantelli lemma.

(since, otherwise, we can always work with w_j^+ and w_j^- separately). In view of the Borel-Cantelli lemma and our assumption (4) we complete the proof of the theorem by proving the following lemma.

LEMMA. *For each $n \geqq 1$ and for any $a \geqq ||\mathbf{w}_n||_2$,*

$$P[D_n \geqq a] < \frac{4a||\mathbf{w}_n||}{||\mathbf{w}_n||_2^2} \exp\left\{-2\left(\left(\frac{a}{||\mathbf{w}_n||_2}\right)^2 - 1\right)\right\} . \tag{6}$$

PROOF OF THE LEMMA. For each $j = 1, \cdots, n$, let $w_j' = w_j/||\mathbf{w}_n||_2$ and $W = \sum_1^n w_j'$. Set $H_n = \sum_1^n w_j' F_j$, $H_n^* = \sum_1^n w_j' Y_j$ and $S = \sup_{x,\alpha} \max_{N \leq n} |H_N^*(x) - H_N(x)|$. Thus, to complete the proof of the lemma, it suffices to show that, with $M = a/||\mathbf{w}_n||_2$,

$$P[S \geqq M] < 4WM \exp(-2(M^2 - 1)) . \tag{7}$$

Let $\Delta = \max_{N \leq n} (H_N^* - H_N)$ and $S^+ = \sup_{x,\alpha} \Delta(x)$. The remark following (2.17) of Hoeffding (1963) page 17, and Theorem 2 therein, applied to random variables $w_j' Y_j$ with $\alpha = 1$ give

$$P[\Delta(x-) \geqq \eta] \leqq \exp(-2\eta^2) \qquad \forall x \in R \text{ and } \forall \eta > 0 . \tag{8}$$

Fix (temporarily) $0 < \gamma \leqq M$ and partition R into k intervals with endpoints $-\infty = x_0 < x_1 < \cdots < x_k = \infty$ such that $H_n(x_{j-1}, x_j) \leqq \gamma$ for $j = 1, \cdots, k$. Since $0 \leqq H_n(\cdot) \leqq W$, we can (and do) take $k < W\gamma^{-1} + 1$. Since $H_N(x_{j-1}, x_j) \leqq H_n(x_{j-1}, x_j) \leqq \gamma$ for $N \leqq n$, using the monotonicity of H_N and H_N^*, we get

$$\begin{aligned} \sup_{x_{j-1} < x < x_j} \Delta(x) &\leqq \max_{N \leq n}(H_N^*(x_j-) - H_N(x_{j-1}+)) \\ &\leqq \Delta(x_j-) + \gamma . \end{aligned} \tag{9}$$

Note that the rhs of (9) is independent of α.

Now observe that $\Delta(x) \leqq \Delta(x+) \vee \Delta(x-) \leqq \sup_{x \in A} \sup_\alpha \Delta(x)$, where A is any dense subset of R. Therefore, $S^+ = \sup_{x,\alpha} \Delta(x) \leqq \sup_\alpha \max_{1 \leq j \leq k} \sup_{x_{j-1} < x < x_j} \Delta(x)$, and from (9), (8) and $\Delta(x_k-) = 0$, we have

$$\begin{aligned} P[S^+ \geqq M] &\leqq P(\textstyle\bigcup_1^{k-1} [\Delta(x_j-) \geqq M - \gamma]) \\ &< W\gamma^{-1} \exp(-2(M - \gamma)^2) . \end{aligned} \tag{10}$$

Since the lhs of (10) is independent of γ, substituting γ on the rhs of (10) by $\gamma_0 = M(1 - (1 - M^{-2})^{\frac{1}{2}})$ and noting that $\gamma_0^{-1} \leqq 2M$, we get from (10)

$$P[S^+ \geqq M] < 2WM \exp(-2(M^2 - 1)) . \tag{11}$$

Let S^- be defined by interchanging H_N^* and H_N in S^+. Then, since $S^-(\mathbf{X}_n) = S^+(-\mathbf{X}_n)$ where $\mathbf{X}_n = (X_1, \cdots, X_n)$, the arguments used for (11) lead to

$$P[S^- \geqq M] < \text{rhs of (11)} . \tag{12}$$

Since $S = S^+ \vee S^-$, the proof of (7) (and hence of the lemma) is complete by (11) and (12).

3. Remarks. Consideration of certain nonparametric test-statistics is the motivation of the present consideration of weighted empiricals, (e.g., the weights $w_1, \cdots, w_n$ could be regression constants of certain nonparametric test statistics based on $X_1, \cdots, X_n$). The result of the theorem with $\alpha = 0$ and $w_j \equiv 1$ is needed in another paper (Singh (1974)) on nonparametric estimation of derivatives of average of densities.

The choice, γ_0, of γ in the proof of the lemma is made so that $\gamma_0^{-1}\exp(-2M - \gamma_0)^2)$ is quite close to the $\inf_{0<\gamma\leq M} \gamma^{-1}\exp(-2(M-\gamma)^2)$, and the resulting bound is not a complicated one. This choice is suggested by Professor James F. Hannan.

When in the theorem $a_n = ||\mathbf{w}_n||_2\{1 + \log(n|\mathbf{w}_n|^{\frac{1}{2}})\}^{\frac{1}{2}}$ where $|\mathbf{w}_n| = ||\mathbf{w}_n||/||\mathbf{w}_n||_2$ and $n|\mathbf{w}_n|^{\frac{1}{2}} \geqq 1$, then, since by the c_r inequality (Loève (1963) page 155) $\{1 + \log(n|\mathbf{w}_n|^{\frac{1}{2}})\}^{\frac{1}{2}} \leqq 1 + \{\log(n|\mathbf{w}_n|^{\frac{1}{2}})\}^{\frac{1}{2}}$, (4) reduces to a simple condition

$$\sum_1^\infty n^{-2}\{\log(n|\mathbf{w}_n|^{\frac{1}{2}})\}^{\frac{1}{2}} < \infty . \tag{4'}$$

Thus, as a special case of the theorem we have: If $n|\mathbf{w}_n|^{\frac{1}{2}} \geqq 1$ for all sufficiently large n, and if (4′) holds, then

$$D_n = O(||\mathbf{w}_n||_2\{1 + \log(n|\mathbf{w}_n|^{\frac{1}{2}})\}^{\frac{1}{2}}) \quad \text{w.p. } 1 . \tag{5'}$$

In particular, with $w_j \equiv 1$, (5′) gives (3).

It is proved by Dvoretzky, Kiefer and Wolfowitz (1956) (and later generalized to the multivariate case by Kiefer and Wolfowitz (1958)) that there is a universal constant c such that, for all $r \geqq 0$, $P[\text{lhs } (1) \geqq r] \leqq c\exp(-2r^2/n)$. This bound is stronger than the one obtained for the *larger probability* in (6) (with $w_j \equiv 1$), (omission of the condition that $a \geqq ||\mathbf{w}_n||_2$ in the lemma here results in a slight change in the bound). The question of whether an inequality of the type $P[\text{lhs } (2) \geqq r] \leqq c_1\exp(-c_2 r^2/n)$, where c_1, c_2 are universal constants, holds and that whether lhs of (2) is $O((n \log\log n)^{\frac{1}{2}})$ w.p. 1 are still open. The affirmative answers of these questions, however, may not lead to similar results concerning the lhs of (3) (and hence of (5)), because (3) is a special case of (5) and the lhs of (3) could be much larger than that of (2).

Acknowledgment. The author wishes to express his sincere appreciation to Professor James F. Hannan for his comments and help which led to a refinement to an earlier version of the paper.

REFERENCES

[1] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[2] Hoeffding, Wassily (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[3] Kiefer, J. (1961). On large deviations of the empiric df of vector chance variables and a law of iterated logarithm. *Pacific J. Math.* **11** 649–660.

[4] Kiefer, J. and Wolfowitz, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Soc.* **87** 173–186.

[5] Loève, Michel (1963). *Probability Theory* 3rd ed. Van Nostrand, Princeton.

[6] SINGH, RADHEY S. (1974). Nonparametric estimation of derivatives of the average of n μ-densities, and convergence rates in n. Submitted to *Ann. Statist.*

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
WELLS HALL
EAST LANSING, MICHIGAN 48824