JOURNAL OF MULTIVARIATE ANALYSIS 7, 594–597 (1977)

A Uniform Bound for the Deviation of Empirical Distribution Functions

LUC P. DEVROYE*

Department of Electrical Engineering, University of Texas, Austin, Texas 78712 and School of Computer Science, McGill University, Montreal, Canada H3C 3G1

Communicated by M. Rosenblatt

If $X_1, \ldots, X_n$ are independent R^d-valued random vectors with common distribution function F, and if F_n is the empirical distribution function for $X_1, \ldots, X_n$, then, among other things, it is shown that

$$P\{\sup_x |F_n(x) - F(x)| \geqslant \epsilon\} \leqslant 2e^2(2n)^d e^{-2n\epsilon^2}$$

for all $n\epsilon^2 \geqslant d^2$. The inequality remains valid if the X_i are not identically distributed and $F(x)$ is replaced by $\sum_i P\{X_i \leqslant x\}/n$.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent identically distributed random vectors in R^d with distribution function F. If $F(x) = P\{X_1 \leqslant x\}$, F_n is the empirical distribution function for $X_1, \ldots, X_n$ (i.e., $F_n(x) = \sum_{i=1}^n I_{\{X_i \leqslant x\}}/n$ where I is the indicator function) and if

$$D_n = \sup_x |F_n(x) - F(x)| \tag{1.1}$$

then Dvoretzky *et al.* [1] for $d = 1$ show that there exists a universal constant C such that

$$P\{D_n \geqslant \epsilon\} \leqslant Ce^{-2n\epsilon^2}. \tag{1.2}$$

For $d \geqslant 1$, Kiefer and Wolfowitz [2] show that

$$P\{D_n \geqslant \epsilon\} \leqslant C_1(d)\, e^{-C_2(d)n\epsilon^2}, \tag{1.3}$$

Received April 3, 1977; revised May 12, 1977.

AMS 1970 subject classification: 62H99.

Key words and phrases: Random vectors, empirical distribution function, uniform consistency, probability inequality.

* This research was supported in part by the Air Force under Grant AFOSR-72-2371.

ISSN 0047-259X

where $C_1(d)$, $C_2(d)$ are positive constants only depending upon d. In [3] this result is improved; in particular, it is shown that for all $b \in (0, 2)$

$$P\{D_n \geqslant \epsilon\} \leqslant C_3(b, d)\, e^{-(2-b)n\epsilon^2}, \tag{1.4}$$

where $C_3(b, d)$ is a constant only depending upon b and d. In none of these papers are expressions derived for C, C_1, or C_3. We show that

$$P\{D_n \geqslant \epsilon\} \leqslant 2(2n)^d\, e^{2\alpha\epsilon^2} e^{-2n\epsilon^2} \tag{1.5}$$

for all n with $n\epsilon^2 \geqslant d^2/\alpha$ and $n \geqslant \alpha$. Using the fact that $D_n \leqslant 1$ and letting $\alpha = 1$ yields

$$P\{D_n \geqslant \epsilon\} \leqslant 2e^2(2n)^d\, e^{-2n\epsilon^2}, \tag{1.6}$$

which is valid for all $n\epsilon^2 \geqslant d^2$.

Singh has proved an inequality of the same type in [4, Theorem 2.1, inequality (2.9)]. In the next section we will point out an error in Singh's argument which makes his proof invalid for $d > 1$. The inequality of Singh is not disproved. However, an obvious correction of his proof, also found in the next section, leads to inequality (1.7) which is much weaker than Singh's original inequality:

$$P\{D_n \geqslant \epsilon\} \leqslant 2e^2((1 + 2nd\epsilon)^d - 1)\, e^{-2n\epsilon^2}, \qquad \text{all } n \geqslant 1/\epsilon^2. \tag{1.7}$$

Inequality (1.6) is tighter than (1.7) for $d\epsilon \geqslant 1 - 1/2n$, that is, for large dimensions d. The detailed proof of (1.6) is given in the next section. We note here that both (1.6) and (1.7) can be used to derive expressions for $C_3(b, d)$ in (1.4). Also, both inequalities are strong enough to imply that $D_n = O(\log n/n)^{1/2}$ with probability one. To see this, use

$$\sum_{n=1}^{\infty} P\{D_n \geqslant [(1 + 2d \log n)/n]^{1/2}\} < \infty$$

and the Borel–Cantelli lemma.

Finally we notice that (1.5)–(1.7) remain valid if $X_1, \ldots, X_n$ are independent but not identically distributed and if $F(x)$ is replaced by

$$n^{-1} \sum_{i=1}^{n} P\{X_i \leqslant x\}.$$

2. Proofs

Proof of (1.5). Let $X_i = (X_{i1}, \ldots, X_{id})$, $1 \leqslant i \leqslant n$, and let $Y_1, \ldots, Y_{n^d}$ be the random vectors obtained by considering all $(X_{i_1 1}, \ldots, X_{i_d d})$ where $(i_1, \ldots, i_d) \in \{1, \ldots, n\}^d$. For each $x = (x_1, \ldots, x_d) \in R^d$ and for each $a = (a_1, \ldots, a_d) \in \{0, 1\}^d$, let $A(x, a)$ be the set of all $y = (y_1, ..., y_d) \in R^d$ such that $y_j \leqslant x_j$ for all j for which $a_j = 1$, and $y_j < x_j$ for all j for which $a_j = 0$. Define then

$$F^a(x) = P\{X_1 \in A(x, a)\}$$

and

$$F_n^a(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \in A(x,a)\}}.$$

Clearly,

$$\sup_x |F_n(x) - F(x)| = \sup_a \sup_i |F_n^a(Y_i) - F^a(Y_i)|$$

by the monotonicity of F_n and F and the fact that F_n is a staircase function with flat levels on rectangles that have some of the Y_i as vertices. Thus,

$$P\{D_n \geqslant \epsilon\} \leqslant 2^d n^d \sup_a \sup_i P\{|F_n^a(Y_i) - F^a(Y_i)| \geqslant \epsilon\}.$$

To every $i \leqslant n^d$ corresponds an $(i_1, ..., i_d) \in \{1, ..., n\}^d$ with $d' \leqslant d$ different components i_j. Thus, if $X_1, ..., X_n, X_{n+1}, ..., X_{n+d'}$ are independent identically distributed random vectors, then we have for every fixed a,

$$\begin{aligned} &P\{|F_n^a(Y_i) - F^a(Y_i)| \geqslant \epsilon\} \\ &\leqslant P\Bigg\{\Bigg| n^{-1} \sum_{\substack{k=1 \\ k \notin \{i_1, ..., i_d\}}}^{n+d'} (I_{\{X_k \in A(Y_i, a)\}} - P\{X_k \in A(Y_i, a)\}) \Bigg| \\ &\quad + \Bigg| n^{-1} \sum_{k \in \{i_1, ..., i_d\}} I_{\{X_k \in A(Y_i, a)\}} - n^{-1} \sum_{k=n+1}^{n+d'} I_{\{X_k \in A(Y_i, a)\}} \Bigg| \geqslant \epsilon \Bigg\}. \end{aligned}$$

The second term in the latter sum is never larger than $d'/n \leqslant d/n$. If n is so large that $d/n \leqslant \gamma\epsilon$ (where $\gamma \leqslant 1$), then by the independence of Y_i and $X_1, ..., X_{n+d'}$ with $X_{i_1}, ..., X_{i_d}$ omitted,

$$\begin{aligned} &P\{|F_n^a(Y_i) - F^a(Y_i)| \geqslant \epsilon\} \\ &\leqslant \sup_x P\Bigg\{\Bigg| n^{-1} \sum_{k=1}^{n} (I_{\{X_k \in A(x,a)\}} - P\{X_k \in A(x, a)\}) \Bigg| \geqslant (1 - \gamma)\epsilon\Bigg\} \end{aligned}$$

which by Hoeffding's inequality [5] can be upper bounded by

$$2e^{-2n(1-\gamma)^2\epsilon^2}.$$

Since this bound is uniform over all a and i, we obtain

$$P\{D_n \geqslant \epsilon\} \leqslant 2(2n)^d e^{-2n(1-\gamma)^2\epsilon^2}.$$

Let $n \geqslant \alpha$ and $(1 - \gamma)^2 = 1 - \alpha/n$ (i.e., $\gamma = 1 - (1 - \alpha/n)^{1/2}$). Resubstitution gives the desired inequality (1.5), which is valid for $d \leqslant n\gamma\epsilon \leqslant n(1-(1-(\alpha/n)^{1/2})\epsilon = \epsilon(\alpha n)^{1/2}$.

If the X_i are independent but have different distribution functions, replace $F(x)$ and $F^\alpha(x)$ by

$$n^{-1} \sum_{i=1}^{n} P\{X_i \leqslant x\}$$

and

$$n^{-1} \sum_{i=1}^{n} P\{X \in A(x, a)\}. \qquad \text{Q.E.D.}$$

Proof of (1.7). The proof of Lemma 2.1 of Singh [4] is picked up at Eq. (2.6). For the sake of brevity, we will use Singh's notation (with p replaced by d) without redefining all the symbols. In (2.6), Singh upper bounds $\sup_{x,\alpha} \Delta(x)$ by

$$\sup_{\alpha} \max_{1 \leqslant j \leqslant r} \sup \left\{ \Delta(x) \,\middle|\, x \in \bigtimes_{i=1}^{d} (x_{i\,j-1}, x_{i\,j}) \right\},$$

which is incorrect. However, $\sup_{x,\alpha} \Delta(x)$ can be upper bounded by

$$\begin{aligned} &\sup_{\alpha} \max_{(i_1,\dots,i_d)\in\{1,\dots,r\}^d} \sup\{\Delta(x) \mid x \in (x_{1\,i_1-1}, x_{1\,i_1}) \times \cdots \times (x_{d\,i_d-1}, x_{d\,i_d})\} \\ &\quad \leqslant \max_{(i_1,\dots,i_d)} \Delta((x_{1\,i_1}, \dots, x_{d\,i_d})-) + \gamma, \\ &\quad \text{not all components} = r \end{aligned}$$

if an argument is used as in [4, (2.5)]. Following (2.7), we obtain

$$P\{W+ \geqslant K\} \leqslant ((1 + dc/\gamma)^d - 1)\, e^{-2(K-\gamma)^2}.$$

Taking γ as suggested by Singh gives

$$P\{W \geqslant K\} \leqslant 2((1 + 2dcK)^d - 1)\, e^2 e^{-2K^2}$$

which, upon letting $K = \epsilon^{1/2} n$ and $c = n^{1/2}$ proves (1.7). Q.E.D.

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

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

[4] Singh, R. S. (1976). On the rate for uniform strong consistency of empirical distributions of independent nonidentically distributed multivariate random variables. *J. Multivariate Analysis* **6** 338–342.

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