may be desirable to consider the distribution of K for $k \geqq n/3$, where the results are likely to be less simple and neat.

REFERENCE

[1] WILLIAM FELLER, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1950.

ON THE CONVERGENCE OF EMPIRIC DISTRIBUTION FUNCTIONS[1]

BY J. R. BLUM

Indiana University

1. Summary. Let μ be a probability measure on the Borel sets of k-dimensional Euclidean space E_k. Let $\{X_n\}$, $n = 1, 2, \cdots$, be a sequence of k-dimensional independent random vectors, distributed according to μ. For each $n = 1, 2, \cdots$ let μ_n be the empiric distribution function corresponding to $X_1, \cdots, X_n$, i.e., for every Borel set $A \varepsilon E_k$, we define $\mu_n(A)$ to be the proportion of observations among $X_1, \cdots, X_n$ which fall in A.

Let $\mathfrak{A}$ be the class of Borel sets in E_k defined below. The object of this paper is to prove that $P\{\lim_{n\to\infty} \sup_{A \epsilon \mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1$.

2. Introduction. Let $F(x)$ be a distribution function on the real line and let $\{X_n\}$, $n = 1, 2, \cdots$, be a sequence of independent random variables distributed according to F. For each $n = 1, 2, \cdots$ let $F_n(x)$ be the empiric distribution function corresponding to $X_1, \cdots, X_n$. The well-known theorem of Glivenko-Cantelli (see, e.g., Fréchet [1]) states that

$$P\{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\} = 1.$$

Fortet and Mourier [2] have proved several theorems on the convergence of empiric distribution functions in a separable metric space E. In particular, they show that if E is a Euclidean space and μ is a probability measure on E which is absolutely continuous with respect to Lebesgue measure, then

$$P\{\lim_{n\to\infty} \sup_{A \epsilon \mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1, \tag{2.1}$$

where $\mathfrak{A}$ is the collection of open half-spaces in E. Wolfowitz [3] proved that (2.1) holds without any assumptions on μ. In this note we prove that if μ is absolutely continuous with respect to Lebesgue measure, then (2.1) holds for a considerably more general class of sets.

To avoid repetition we shall assume from now on that every set considered is

Received October 29, 1954.

[1] Work done with the support of the Office of Ordnance Research, U.S. Army.

a Borel subset of E_k. Let $\mathfrak{A}_1$ be the class of sets A each of which possesses the following property. If $x = (x_1, \cdots, x_k) \varepsilon A$ and $y = (y_1, \cdots, y_k)$ is such that $y_i < x_i$ for $i = 1, \cdots, k$, then $y \varepsilon A$. Let $\mathfrak{A}_j$, $j = 2, \cdots, 2^k$, be the $2^k - 1$ classes of sets which can be obtained by reversing, one at a time, the k inequalities occurring in the definition of $\mathfrak{A}_1$. Let $\mathfrak{A} = \bigcup_{j=1}^{2^k} \mathfrak{A}_j$. In this note we shall prove the following theorem.

THEOREM. *If μ is absolutely continuous with respect to Lebesgue measure, then*

$$P\{\lim_{n\to\infty} \sup_{A\varepsilon\mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1.$$

3. Proof of the theorem. In proving the theorem we shall restrict ourselves to the class $\mathfrak{A}_1$. The method of proof also applies to each of the classes $\mathfrak{A}_2, \cdots, \mathfrak{A}_{2^k}$, and consequently, from elementary considerations, the theorem holds for $\mathfrak{A}$.

The method of proof depends on the following lemma.

LEMMA 1. *Let $\mathfrak{B}$ be a class of sets and suppose for each $\rho > 0$ there exists a finite class of sets $\mathfrak{B}(\rho)$ such that for each $B \varepsilon \mathfrak{B}$ there exist sets B_1 and B_2 in $\mathfrak{B}(\rho)$ satisfying*

i) $$B_1 \subset B \subset B_2,$$

ii) $$\mu(B_2) - \mu(B_1) \leqq \rho.$$

Then $P\{\lim_{n\to\infty} \sup_{B\varepsilon\mathfrak{B}} |\mu_n(B) - \mu(B)| = 0\} = 1.$

The proof of the lemma is a direct consequence of the strong law of large numbers and is omitted here.

In proving the theorem we shall assume that $k = 2$. It will be clear from the sequel that the method of proof applies to arbitrary k, although the details become vastly more complicated.

Let R be a closed square in the plane which is subdivided into m^2 subsquares of equal area by dividing each side into m equal length intervals. Let $\mathfrak{A}_1(R)$ be the class of sets of the form $A \cap R$, with $A \varepsilon \mathfrak{A}_1$. For each set $T \varepsilon \mathfrak{A}_1(R)$ let $B(T)$ be the set of boundary points of T with the exception of those lying on the south and west boundaries of R. If $x = (x_1, x_2) \varepsilon R$, we shall say that x lies in a subsquare if it lies in the interior or on the north or east boundary of the subsquare. Let $N(T)$ be the number of distinct subsquares in which the points of $B(T)$ lie. Then we have the following lemma.

LEMMA 2. *For every $T \varepsilon \mathfrak{A}_1(R)$, $N(T) \leqq 2m - 1$.*

PROOF. We may assume that the coordinates of the corners of the subsquares are of the form (i, j) with $i = 0, \cdots, m$; $j = 0, \cdots, m$. Now consider the $2m - 1$ lines of the form $f(x) = x + k$, with $k = -m + 1, -m + 2, \cdots, m - 1$. By identifying each subsquare with the coordinates of its northeast corner it is easily seen that through each subsquare passes one and only one of these lines. Let $T \varepsilon \mathfrak{A}_1(R)$. We shall show that on every line of the form $f(x) = x + k$ there lies at most one point of $B(T)$. For suppose $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two distinct points of $B(T)$, and both lying on a line $f(x) = x + k$. Assume that $x_i < y_i$, $i = 1, 2$. Then we can find a point $z = (z_1, z_2) \varepsilon T$, with $x_i < z_i$, $i = 1, 2$. But this contradicts the fact that $x \varepsilon B(T)$. From this it follows that each line $f(x) = x + k$ passes through at most one subsquare containing points of $B(T)$. Since there are $2m - 1$ such lines the lemma follows.

Let (i_1, j_1) be the coordinates of a corner of a subsquare with either $i_1 = 0$ or $j_1 = m$, and let (i_2, j_2) be the coordinates of another corner of a subsquare with either $i_2 = m$ or $j_2 = 0$. By a path P in R we shall mean a linear continuum of points connecting (i_1, j_1) with (i_2, j_2) and satisfying in addition:

i) Every point $x = (x_1, x_2) \varepsilon P$ lies on the boundary of a subsquare of R.

ii) If $x = (x_1, x_2) \varepsilon P$ and $y = (y_1, y_2) \varepsilon P$, and if $x_1 \leqq y_1$, then $x_2 \geqq y_2$.

By induction on m it is easily verified that there are at most finitely many paths P in R. To each path P we associate two sets $T_1(P)$ and $T_2(P)$ in $\mathfrak{A}_1(R)$ with $B(T_1) = B(T_2) = P$ and such that $T_1(P)$ contains all points of P and $T_2(P)$ contains no points of P. Let $\mathfrak{A}_{1,m}(R)$ be the class of all sets obtained in this manner for all possible paths P. Then $\mathfrak{A}_{1,m}(R)$ is clearly a finite class of sets for each integer m.

Let $T \varepsilon \mathfrak{A}_1(R)$, and let ρ be a positive number. For any positive integer m we may then choose two sets T_1 and T_2 in $\mathfrak{A}_{1,m}(R)$ such that $T_1 \subset T \subset T_2$, and such that if T' and T'' are in $\mathfrak{A}_{1,m}(R)$ and if $T_1 \subset T' \subset T \subset T'' \subset T_2$, then $T_1 = T'$ and $T_2 = T''$. From the choice of T_1 and T_2 it is clear that $T_2 - T_1$ is contained in the set of subsquares which contain $B(T)$. Let $L(U)$ be the Lebesque measure of a set U. Then from Lemma 2 it follows that $L(T_2 - T_1) \leqq L(R)\, N(T) / m^2 \leqq L(R)\,(2m - 1) / m^2$. Since μ is absolutely continuous with respect to L, we may choose an integer m such that $\mu(T_2 - T_1) < \rho$. Applying Lemma 1 we obtain the following lemma.

LEMMA 3. $P\{\lim_{n\to\infty} \sup_{T\varepsilon\mathfrak{A}_1(R)} |\mu_n(T) - \mu(T)| = 0\} = 1$.

Let ρ be a positive number. Let $A \varepsilon \mathfrak{A}_1$, and let R be a square with $\mu(R) > 1 - \rho/4$. Write $A = A_1 \cup A_2$, where $A_1 = A \cap R$, $A_2 = A \cap \bar{R}$, and where $\bar{R}$ is the complement of R. By virtue of Lemma 3 it suffices to show that $\lim_{n\to\infty} \sup_{A_2} |\mu_n(A_2) - \mu(A_2)| = 0$ on a set of sample sequences of probability one. Now consider any sample sequence in the set of probability one for which $\lim_{n\to\infty} \mu_n(\bar{R}) = \mu(\bar{R})$. Choose n so large that $\mu_n(\bar{R}) < \rho/2$. Since $0 \leqq \mu_n(A_2) \leqq \mu_n(\bar{R})$ and $0 \leqq \mu(A_2) \leqq \mu(\bar{R})$, we obtain $|\mu_n(A_2) - \mu(A_2)| < \rho/2$, uniformly in $A_2 \subset \bar{R}$, and the proof of the theorem is complete.

It appears to be a reasonable conjecture that the theorem is true without the condition of absolute continuity. One can easily construct examples which show that the method used in this note will no longer apply in the general situation. It would be of some interest to extend the result to the general case.

REFERENCES

[1] MAURICE FRÉCHET, "Généralités sur les Probabilités. Élements aléatoire", Gauthier-Villars, Paris, 1950.

[2] R. FORTET AND E. MOURIER, "Convergence de la répartition empirique vers la répartition théorique," Ann. Sci. Ecole Norm. Sup., (3) Vol. 70 (1953).

[3] J. WOLFOWITZ, "Generalization of the theorem of Glivenko-Cantelli," Ann. Math. Stat., Vol. 25, (1954), pp. 131–138.