The Annals of Probability
1974, Vol. 2, No. 1, 163-166

NEAREST RANDOM VARIABLES WITH GIVEN DISTRIBUTIONS

BY GEZA SCHAY

University of Massachusetts at Boston

A new proof, based on the duality theorem of linear programming, is given of a theorem of V. Strassen, which states essentially that the minimum distance between random variables with given distributions equals the Prokhorov distance of their distributions.

In 1965 V. Strassen [5] proved that for any two Borel probability measures μ and ν on a complete separable metric space S with metric d, there is a Borel probability measure λ on $S \times S$ with μ and ν as marginals, such that $\min\{\varepsilon : \lambda\{d(x, y) > \varepsilon\} \leqq \varepsilon\}$ equals the Lévy–Prokhorov distance of μ and ν. If we introduce a probability space and S-valued random variables, then this can be restated as saying that there exist nearest random variables distributed according to μ and ν, and their (Ky Fan) distance equals the Prokhorov distance of their distributions.

Strassen's proof, however, is a non-constructive one. In 1968 R. M. Dudley [1] improved Strassen's result, and gave a construction for the approximation of random variables with minimum distance. His proof is based on some clever and involved mappings and combinatorial arguments.

In this paper we give a simple construction based on approximating μ and ν by measures on finite sets and applying the duality theorem of linear programming to them.

As in Dudley's paper one can obtain the minimum distance between random variables X and Y distributed respectively according to μ and ν in two steps: by finding first

$$\beta(\alpha) = \inf\{\beta : \beta = \Pr(d(X, Y) > \alpha)\} \qquad \text{for} \quad 0 \leqq \alpha \leqq 1,$$

where the infimum is taken over all X and Y distributed according to μ and ν; and then finding $\inf\{\alpha : \beta(\alpha) \leqq \alpha\}$. Now $\beta(\alpha)$ can also be obtained by finding a joint distribution for X and Y with the given marginals, which maximize the probability of (X, Y) falling into the diagonal strip given by $d(x, y) \leqq \alpha$. If one writes this out for X and Y distributed on a finite set, then this becomes a linear programming problem, the dual of which happens to be closely related to the Prokhorov distance of the distributions of X and Y, as will be seen below.

Let x_i, $i = 1, 2, \cdots$ be a dense sequence in a separable metric space S, let n be any positive integer, and P and Q probability measures on the finite set

Received January 26, 1973; revised May 14, 1973.

AMS 1970 *subject classifications*. Primary 60B10; Secondary 60B99.

Key words and phrases. Prokhorov distance, Ky Fan distance, minimum distance between random variables, duality theorem of linear programming.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability*. www.jstor.org

$S_n = \{x_i : i = 1, 2, \cdots, n\}$. We shall write p_i for $P\{x_i\}$ and q_i for $Q\{x_i\}$, and seek appropriate joint probabilities x_{ij} with these as marginals. Also, for any $F \subset S_n$, we shall use the notation

$$F_n^{\alpha]} = \{x_i \in S_n : d(x_i, x_j) \leqq \alpha \text{ for some } x_j \in F\} .$$

Then we have the following

THEOREM 1. *Given any* $\alpha \geqq 0$. *Let* $\{1, 2, \cdots, n\}$ *be denoted by* I, *and let*

$$\beta_n(\alpha) = 1 - \max \sum\sum_{i,j \in I} d_{ij} x_{ij} , \tag{1}$$

where $d_{ij} = 1$ *if* $d(x_i, x_j) \leqq \alpha$, *and is* 0 *otherwise*; *and the maximum is taken over all* x_{ij} *subject to the constraints*

$$\sum_{j \in I} d_{ij} x_{ij} \leqq p_i \quad \textit{for each} \quad i \in I , \quad \textit{and} \tag{2}$$
$$\sum_{i \in I} d_{ij} x_{ij} \leqq q_j \quad \textit{for each} \quad j \in I .$$

Then

$$\begin{aligned} \beta_n(\alpha) &= \max_{F \subset S_n} [P(F) - Q(F_n^{\alpha]})] \\ &= \inf \{\beta : P(F) \leqq Q(F_n^{\alpha]}) + \beta \ \textit{for all} \ F \subset S_n\} . \end{aligned} \tag{3}$$

PROOF. By the duality theorem of linear programming (see e.g., [2] page 320, or for a summary [3] pages 70–73), the value of the maximum in (1) equals the value of the minimum in a dual problem, which in our case is easily seen to be that of finding u_i and v_i for $i = 1, 2, \cdots, n$, that solve

$$\min \sum_{i \in I} (p_i u_i + q_i v_i) \tag{4}$$

subject to the constraints

$$u_i, v_j \geqq 0, u_i + v_j \geqq d_{ij} \qquad \text{for all} \quad i, j \in I . \tag{5}$$

Let us substitute $w_i = 1 - u_i$. Then the minimum in (4) becomes

$$1 - \max \sum_{i \in I} (p_i w_i - q_i v_i) , \tag{6}$$

with

$$v_j \geqq 0, v_j \geqq (d_{ij} - 1) + w_i, w_i \leqq 1 \qquad \text{for all} \quad i, j \in I . \tag{7}$$

We may assume $w_i \geqq 0$, since if any one of the w_i were negative, then it could be replaced by 0 without violating the constraints, and that would increase the sum whose maximum we seek. Similarly, we may assume $v_j \leqq 1$. These constraints describe half-spaces in $2n$ dimensions, which intersect in a convex polyhedron contained in the "unit cube." It is well known that a solution to such an extremum problem is always given by the coordinates of a vertex of this polyhedron. (Indeed, the level surfaces of a linear function are parallel planes, and the one with the "highest level" among those that intersect the polyhedron must obviously contain a vertex.) Now the coordinates of the vertices of our polyhedron are solutions of sets of $2n$ simultaneous equations obtained by replacing the inequalities by equalities in (7) and in $w_i \geqq 0$ and $v_j \leqq 1$. It is easy to see that for all such solutions, that is, for all vertices, each v_j and w_i is either 0 or 1, Now if $w_i = 1$ for all i with x_i in some $F \subset S_n$, and $w_i = 0$ otherwise, then the second set of inequalities in (7) shows that $v_j = 1$ for all $j \in I$ for which $x_j \in F_n^{\alpha]}$. Hence the maximum in (6) equals the maximum in (3). The second equality in (3) is trivial, so the theorem is proved.

Now we want to consider arbitrary Borel probability measures μ and ν on S. We write, for any $F \subset S$, $F^{\alpha]} = \{x \in S : d(x, F) \leqq \alpha\}$, and we change our notations P and Q to P_n and Q_n. Also, we let P_n^* and Q_n^* denote the Borel probability measures on S whose restrictions to S_n are P_n and Q_n, and let λ_n denote the Borel probability measure on S, whose restriction to $S_n \times S_n$ is defined by the optimal x_{ij}, that is, for any Borel set $T \subset S \times S$, $\lambda_n(T) = \sum\sum_T x_{ij}$.

With these notations we have

THEOREM 2. *If the separable metric space S is complete,*[1] *then for any Borel probability measures μ and ν on S, we can find P_n^* and Q_n^* for $n = 1, 2, \cdots$, converging weakly to μ and ν. If the corresponding $\beta_n(\alpha)$ and λ_n are constructed as above, then there exists a subsequence λ_{n_k} of the λ_n which converges weakly to some Borel probability measure λ on $S \times S$, and we have*

$$(8) \qquad \begin{aligned}\beta(\alpha) &\equiv \lim_{k\to\infty} \beta_{n_k}(\alpha) = \lambda\{(x, y) : d(x, y) > \alpha\} \\ &= \inf\{\beta : \mu(F) \leqq \nu(F^{\alpha]}) + \beta \text{ for all Borel sets } F \subset S\}.\end{aligned}$$

PROOF. This result and its proof are similar to those of Dudley's Theorem 2. The main difference is that our approximating λ_n are constructed differently, and do not have the same marginals for each n except in trivial cases.

Due to the assumed completeness of S (see [6] page 202) for any $\delta > 0$ there is a compact $K \subset S$ such that

$$(9) \qquad \mu(S - K) < \delta/2, \qquad \nu(S - K) < \delta/2,$$

and we can choose P_n^* and Q_n^* converging to μ and ν such that also

$$(10) \qquad P_n^*(S - K) < \delta/2 \quad \text{and} \quad Q_n^*(S - K) < \delta/2$$

hold for each n. Then obviously

$$(11) \qquad \lambda_n((S \times S) - (K \times K)) < \delta$$

for each n.

Thus the sequence λ_n is uniformly tight, and so it has a subsequence such that $\lambda_{n_k} \to \lambda$ weakly for some Borel probability measure λ on $S \times S$. The relation (8) follows easily from (3) and Theorem 1.2 of [4].

REFERENCES

[1] DUDLEY, R. M. (1968). Distance of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572.

[1] In fact, somewhat weaker conditions than completeness would suffice (see Dudley [1], Theorem 2).

[2] GAL, D., KUHN, H. W. and TUCKER, A. W. (1951). Linear programming and the theory of games, in *Activity Analysis of Production and Allocation*, ed. T. C. Koopmans. Wiley, New York.

[3] HALL, M. (1958). A Survey of combinatorial analysis, in *Some Aspects of Analysis and Probability*. Wiley, New York.

[4] PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–214.

[5] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist* **36** 423–439.

[6] VARADARAJAN, V. S. (1961). Measures on topological spaces. *Amer. Math. Soc. Trans. Ser.* 2 **48** 161–228.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MASSACHUSETTS
100 ARLINGTON STREET
BOSTON, MASSACHUSETTS 02116