

## SEPARATION METRICS FOR REAL-VALUED RANDOM VARIABLES

MICHAEL D. TAYLOR

Department of Mathematics  
University of Central Florida  
Orlando, Florida 32816

(Received April 3, 1984)

ABSTRACT. If  $W$  is a fixed, real-valued random variable, then there are simple and easily satisfied conditions under which the function  $d_W$ , where  $d_W(X, Y) =$  the probability that  $W$  "separates" the real-valued random variables  $X$  and  $Y$ , turns out to be a metric. The observation was suggested by work done in [1].

KEY WORDS AND PHRASES. *Random variables, probability spaces, distribution functions metrics, metrics on random variables*

1980 MATHEMATICS SUBJECT CLASSIFICATION CODE. 60 A 99

THEOREM 1. Let  $\Omega$  be a probability space with probability measure  $P$ , and let  $W$  be a fixed, real-valued random variable on  $\Omega$ . Then the function  $d_W$  defined by

$$d_W(X, Y) = P[X \leq W < Y \text{ or } Y \leq W < X]$$

is a pseudo metric on the space of real-valued random variables defined on  $\Omega$ .

PROOF. We need check only the triangle inequality. Let  $X, Y$ , and  $Z$  be real-valued random variables on  $\Omega$  and set

$$A = \{\omega \in \Omega : X(\omega) \leq W(\omega) < Y(\omega) \text{ or}$$

$$Y(\omega) \leq W(\omega) < X(\omega)\},$$

$$B = \{\omega \in \Omega : Y(\omega) \leq W(\omega) < Z(\omega) \text{ or}$$

$$Z(\omega) \leq W(\omega) < Y(\omega)\}, \text{ and}$$

$$C = \{\omega \in \Omega : X(\omega) \leq W(\omega) < Z(\omega) \text{ or}$$

$$Z(\omega) \leq W(\omega) < X(\omega)\}$$

Let  $\omega \in C$ . If  $X(\omega) \leq W(\omega) < Z(\omega)$ , then  $Y(\omega) \leq W(\omega)$  implies  $W(\omega)$  "separates"  $Y(\omega)$  and  $Z(\omega)$  in such a fashion that  $\omega \in B$ , and  $W(\omega) < Y(\omega)$  implies  $W(\omega)$  "separates"  $X(\omega)$  and  $Y(\omega)$  in such a fashion that  $\omega \in A$ . A similar conclusion holds in the case  $Z(\omega) \leq W(\omega) < X(\omega)$ . Thus  $C \subset A \cup B$  and  $P(C) \leq P(A) + P(B)$  which is the triangle inequality.

REMARK. Random variables which differ only on a set of probability measure 0 will be considered to be identical.

THEOREM 2. Let  $\Omega$ ,  $P$ , and  $W$  be as in Theorem 1. Let  $R$  be some given collection of real-valued random variables on  $\Omega$ , and suppose that (1)  $W$  is independent of every pair of members of  $R$  in the sense that if  $X, Y \in R$  and  $A, B$ , and  $C$  are intervals in  $\mathbb{R}$ , then

$$P [X \in A, Y \in B, W \in C] = P [X \in A, Y \in B] \cdot P [W \in C] \quad \text{and}$$

(2) for every open interval  $J$  in  $\mathbb{R}$  we have

$$P [W \in J] > 0 .$$

Then  $d_W$ , as defined in Theorem 1, is a metric on  $R$ .

PROOF. Let  $X, Y \in R$  such that  $X \neq Y$ . We have only to show  $d_W(X, Y) > 0$ . We may, without loss of generality, suppose that the set

$$A = \{\omega \in \Omega : X(\omega) < Y(\omega)\}$$

has positive  $P$ -measure. Then there must be rational numbers  $p$  and  $q$  such that

$$B = \{\omega \in \Omega : X(\omega) < p < q < Y(\omega)\}$$

has positive  $P$ -measure. It follows that

$$\begin{aligned} d_W(X, Y) &\geq P [X < W < Y] \\ &\geq P [X < p < W < q < Y] \\ &= p(B) \cdot P [p < W < q] > 0. \end{aligned}$$

REMARK 1. In connection with this last theorem, it can be shown that if  $F_W$ , the cumulative distribution function of  $W$ , is continuous, then

$$d_W(X, Y) = E(|F_W(X) - F_W(Y)|)$$

where  $E$  means expected value.

REMARK 2. Again in connection with the last theorem, it might be objected that for some  $R$  no  $W$  exists with the desired properties; this would be the case, for example, if  $R$  was the set of all real-valued random variables on  $\Omega$ . However, one can always "embed"  $R$  in a larger space of real-valued random variables containing something suitable for use as  $W$ . Simply take another probability space  $\Omega'$ , let  $W$  be a real-valued random variable on  $\Omega'$  taking on values in every open interval of  $\mathbb{R}$  with positive probability, let  $\Omega^*$  be the product space  $\Omega \times \Omega'$ , let each  $X$  in  $R$  be replaced  $X^*$  where  $X^*(\omega, \omega') = X(\omega)$ , and let  $W$  be replaced by  $W^*$  where  $W^*(\omega, \omega') = W(\omega')$ . It follows that  $X_1^*, \dots, X_n^*$  must have the same joint distribution function as  $X_1, \dots, X_n$  when  $X_1, \dots, X_n$  are members of  $R$  and that  $W^*$  is independent of all  $X^*$  such that  $X \in R$  in the desired fashion. So one may construct the metric  $d$  on  $R$  defined by

$$d(X, Y) = d_{W^*}(X^*, Y^*).$$

REFERENCE

1. TAYLOR, Michael D., New metrics for weak convergence of distribution functions. To appear.