ON THE TOPOLOGICAL STRUCTURE OF SOME ORDERED FAMILIES OF DISTRIBUTIONS[1]

BY J. PFANZAGL

University of Cologne

1. Introduction. In many statistical contexts one studies a family $\mathfrak{F}$ of distribution functions F on the real line with the following properties. (a) $\mathfrak{F}$ is a dominated family; (b) $\mathfrak{F} = \{F_\theta\}$ where θ is a real parameter taking values in an interval; (c) convergence of a sequence in $\mathfrak{F}$ to an element of $\mathfrak{F}$ in the usual (weak) sense is represented by the convergence of the corresponding parameter points; (d) the natural order of θ values corresponds to a simple ordering of the elements of $\mathfrak{F}$ in terms of probability.

An example of such an ordering in terms of probability is $\theta_1 \geqq \theta_2 \Leftrightarrow F_{\theta_1}(t) \leqq F_{\theta_2}(t)$ for all real t, this ordering of $\mathfrak{F}$ being typical of cases where θ is a location parameter. An example of (d), with θ a scale parameter is $F_\theta =$ the normal distribution function with mean zero and variance θ; here $\theta_1 \geqq \theta_2 \Leftrightarrow G_1(s) \leqq G_2(s)$ *for all real* s, where G_θ is the distribution function of the sufficient statistic $s = t^2$; in view of the one-one correspondence through θ, between the families $\{F_\theta\}$ and $\{G_\theta\}$, the probabilistic ordering of $\{G_\theta\}$ corresponds to such an ordering for $\{F_\theta\}$.

The main object of this paper is to study an arbitrary set $\mathfrak{F}$ of distribution functions which is simply ordered according to the following slightly stronger order relationship: $\theta_1 \geqq \theta_2 \Leftrightarrow F_{\theta_1}(t)/F_{\theta_2}(t)$ and $[1 - F_{\theta_1}(t)]/[1 - F_{\theta_2}(t)]$ are nondecreasing functions of t. This order relation is still weaker than the natural order relation of monotone likelihood ratio families. It is shown, that for such ordered families $\mathfrak{F}$, convergence in the weak sense is equivalent to convergence in the strong sense (see Theorem 1). Thus, we obtain for ordered families much stronger results than can be obtained in the general case. In general, conclusions from weak convergence to a stronger type of convergence can be drawn only if the limit function is continuous (see R. R. Rao (1962) p. 662, Theorem 3.1).

Furthermore it is shown that (except for certain pathological situations) the order relationship defined above implies that $\mathfrak{F}$ is a one-parameter family for which (a), (b) and (c) hold (see Theorems 2, 3 and 4).

In Section 2 the relevant notions of order and distances are defined and discussed. The main theorems are stated in Section 3 and proved in Section 5 using the lemmata of Section 4. In Section 6 it is pointed out that the present results extend to probability measures on arbitrary spaces provided there exists a

Received 5 August 1963; revised 24 February 1964.

[1] This paper was written while the author was visiting professor at the Mathematics Research Center, United States Army, University of Wisconsin. The visit was sponsored under Contract No. DA-11-022-ORD-2059. An earlier version of this paper appeared as MRC-Report No. 424 in 1963.

pairwise sufficient statistic, and an application is given to the problem of existence of perfect tests.

2. Some basic definitions.

2.1. *Orders.* Let $\mathfrak{F}$ be a family of distribution functions F on the real line. We assume that $\mathfrak{F}$ is totally ordered, i.e. that there exists an asymmetric, irreflexive and transitive relation "$<$" such that for arbitrary $F_1, F_2 \in \mathfrak{F}$ with $F_1 \neq F_2$ either $F_1 < F_2$ or $F_2 < F_1$. We shall consider the following three different types of order relations:

$$(1) \qquad F_1 \leqq F_2 \quad \text{if} \quad F_2(t) \leqq F_1(t) \quad \text{for all } t;$$

$$(2) \qquad F_1 \leqq F_2 \quad \text{if both}$$

$$(2') \qquad F_2(t)/F_1(t) \quad \text{is nondecreasing in } t, \quad \text{and}$$

$$(2'') \qquad (1 - F_2(t))/(1 - F_1(t)) \quad \text{is nondecreasing in } t;$$

$$F_1 \leqq F_2 \quad \text{if} \quad F_1(t) = 0 \quad \text{implies} \quad F_2(t) = 0 \quad \text{and if}$$

$$(3) \quad [F_2(t'') - F_2(t')]/[F_1(t'') - F_1(t')] \quad \text{is nondecreasing in both variables } t' \text{ and } t'' \text{ whenever } F_1(t') < F_1(t'').$$

If $F_1 \leqq F_2$ and $F_1 \neq F_2$ we shall write $F_1 < F_2$. It is easily seen that the relations $<$ defined by (1), (2′), (2″) or (3) are asymmetric, irreflexive and transitive. The first two properties are obvious in all cases and the transitivity except in case (3), too.

In case (3) the relations $F_1 < F_2$ and $F_2 < F_3$ imply: If $F_1(t) = 0$ then $F_2(t) = 0$ and therefore also $F_3(t) = 0$. Now, we assume $F_1(t') < F_1(t'')$. If $F_2(t'') = 0$ then $F_3(t'') = 0$ and therefore also $F_3)t') = 0$. Hence $[F_3(t'') - F_3(t')]/[F_1(t'') - F_1(t')] = 0$ and therefore this ratio is trivially nondecreasing. If $F_2(t'') > 0$, then $0 < F_2(t'')/F_1(t'') = \lim_{t \to -\infty} [F_2(t'') - F_2(t)]/[F_1(t'') - F_1(t)] \leqq [F_2(t'') - F_2(t')]/[F_1(t'') - F_1(t')]$, whence $F_2(t') < F_2(t'')$. Therefore the desired result follows from

$$\frac{F_3(t'') - F_3(t')}{F_1(t'') - F_1(t')} = \frac{F_2(t'') - F_2(t')}{F_1(t'') - F_1(t')} \cdot \frac{F_3(t'') - F_3(t')}{F_2(t'') - F_2(t')}.$$

Order (3) is equivalent to "monotony of likelihood ratios" or—more correctly—"monotony of relative densities."

We say that F_2 has a monotone relative density with respect to F_1 if there exists a function $H_{21}(t)$ which is nondecreasing for all t such that $F_2(t) = \int_{-\infty}^{t} H_{21}\, dF_1$ for all t with $F_1(t - 0) < 1$. (We choose $H_{21}(t) = +\infty$ for t with $F_1(t - 0) = 1$.)

That monotony of the relative density implies (3) is obvious. That (3) in turn implies the existence of a monotone relative density can be seen by taking (see e.g. Doob (1953), pp. 611–612):

$$(4) \qquad H_{21}(t) = \lim_{h \to \infty} [F_2(t) - F_2(t - h)]/[F_1(t) - F_1(t - h)]$$

for all $t \varepsilon S$ where $S = \{t : F_1(t) > F_1(t - h) \text{ for all } h > 0\}$. Obviously $\int_S dF_1 = 1$, and therefore H_{21} is defined F_1—a.e. We can complete the definition of H_{21}, preserving monotony, by $H_{21}(t) = \inf \{H_{21}(s) : t \leqq s \varepsilon S\}$ for $t \not\varepsilon S$ with $F_1(t - 0) < 1$ and $H_{21}(t) = +\infty$ for t with $F_1(t - 0) = 1$.

Orders (1) and (3) correspond to orders A and C considered by Lehmann (1955). The order relations (1), (2), (3) are of increasing stringency. It is straightforward to show that order (3) implies order (2) and order (2) implies order (1). (More accurately: each of (2′) or (2″) implies (1), which can be seen by considering the limit as $t \to \infty$ or $t \to -\infty$, respectively.) That (2) is actually stronger than (1) follows, for example, from the fact that any location parameter family is ordered (1), but not all of them are ordered (2), (e.g. the location parameter family of Cauchy distributions is not ordered (2).) That (3) is actually stronger than (2) can be seen from the following example: Let P_1 and P_2 be two probability distributions having their probabilities concentrated on the points 1, 2, 3, 4. Let $P_1(i) = 1/4$, for $i = 1, \cdots, 4$ and $P_2(1) = 1/14$, $P_2(2) = 4/14$, $P_2(3) = 3/14$, $P_2(4) = 6/14$. Then, $F_1 < F_2$ in the sense of order (2). They are, however, not ordered (3).

2.2. *Distances; convex support.* In the following, we will consider the family $\mathfrak{F}$ of distribution functions as a metric space with two different distance functions d' and d'':

$$d'(F_1, F_2) = \sup \{|F_1(t) - F_2(t)| : -\infty < t < +\infty\} \tag{5}$$

$$d''(F_1, F_2) = \sup_{B \varepsilon \mathfrak{B}} |P_1(B) - P_2(B)| \tag{6}$$

where

$$P_i(B) = \int_B dF_i \tag{7}$$

$\mathfrak{B}$ being the field of all Borel sets of the real line. We remark that from $P_2(B) - P_1(B) = P_1(\bar{B}) - P_2(\bar{B})$ we have

$$d''(F_1, F_2) = \sup_{B \varepsilon \mathfrak{B}} (P_1(B) - P_2(B)). \tag{8}$$

Furthermore, $d'(F_1, F_2) = 0$ is equivalent to $d''(F_1, F_2) = 0$.

Let μ be a measure on the real line, such that P_1 and P_2 are absolutely continuous with respect to μ. Such a measure always exists, e.g. $\mu = P_1 + P_2$. In applications, usually the Lebesgue measure or the counting measure has this property. Let p_i be the Radon-Nikodym derivative of P_i with respect to μ, i.e. $P_i(B) = \int_B p_i \, d\mu$. Then

$$d''(F_1, F_2) = P_1(B_{12}) - P_2(B_{12}) \tag{9}$$

where

$$B_{12} = \{t : p_1(t) > p_2(t)\}. \tag{10}$$

It is obvious from the definitions (5) and (6) that

$$d'(F_1, F_2) \leqq d''(F_1, F_2). \tag{11}$$

Therefore, if $\mathfrak{F}$ is ordered (2′) or (2″), according to Theorem 1 both distance functions induce the same topology, say $\mathfrak{D}$. The important consequence of this fact is that d'-convergence is equivalent to d''-convergence which in turn is equivalent to convergence in the mean.

In the case of order (3), both distance functions are identical. For, $p_2(t)/p_1(t) = H_{21}(t)$ $(F_1 + F_2)$-a.e. with monotone H_{21} and hence B_{12} is a semi-infinite interval.

Both distance functions assume values in [0, 1] only. The value 1 plays the role of an "infinite" distance. This can be made more intuitive by introducing the concept of a "*convex support*". Let $s(F)$ be the smallest convex set which has probability one. $s(F)$ is the intersection of all convex sets with probability 1, whence

$$s(F) = \bigcap\{(t_1, t_2) : F(t_2 - 0) - F(t_1) = 1\} \tag{12}$$

where (t_1, t_2) denotes the open interval with end points t_1, t_2. This concept of "convex support" is not identical to the usual concept of support (see e.g. Bourbaki (1952) p. 67 ff.).

The convex support is an interval except in the case of degenerate distributions (having their probability concentrated in a single point), in which case the convex support reduces to a point. The boundary points of the interval belong to the convex support iff they are of positive probability.

It can be easily seen (Lemma 1) that $d'(F_1, F_2) = 1$ if and only if $s(F_1) \cap s(F_2) = \varnothing$.

If F_1 and F_2 are ordered (2′) or (2″), then $d''(F_1, F_2) = 1$ implies $d'(F_1, F_2) = 1$ (Lemma 4).

If $\mathfrak{F}$ is ordered (1), then d' is monotone in the sense that $F_0 \leqq F_1 \leqq F_2$ implies $d'(F_0, F_1) \leqq d'(F_0, F_2)$. If $\mathfrak{F}$ is ordered (2) and $s(F_0) = s(F_1) = s(F_2)$ we have $d'(F_0, F_1) < d'(F_0, F_2)$ if $F_1 < F_2$ (see Lemma 6 and its Corollary).

Let Y be a set in a metric space $(\mathcal{Y}, d)$. As usual, we define the diameter $D(Y)$ by

$$D(Y) = \sup \{d(y_1, y_2) : y_1, y_2 \varepsilon Y\}, \tag{13}$$

and the distance of two sets Y_1, Y_2 by $d(Y_1, Y_2) = \inf \{(d(y_1, y_2) : y_1 \varepsilon Y_1, y_2 \varepsilon Y_2\}$. We call a set Y *linked*, if it is not the union of two sets of distance 1.

We denote the union of the convex supports of a class of probability measures $\mathfrak{F}$ by

$$s(\mathfrak{F}) = \bigcup\{s(F) : F \varepsilon \mathfrak{F}\}. \tag{14}$$

Using this concept, we also can say: A class of distribution functions is linked if it is not the union of two subsets $\mathfrak{F}_1, \mathfrak{F}_2$ which are separated in the sense that $s(\mathfrak{F}_1) \cap s(\mathfrak{F}_2) = \varnothing$. Statistically this means that it is impossible to partition the family into two subfamilies between which we can distinguish deterministically.

If $\mathfrak{F}$ is linked, by Lemma 8 and 1 to each pair F_0, $F_1 \varepsilon \mathfrak{F}$ there exists a chain $F_1, F_2, \cdots F_n, F_{n+1} = F_0$ such that $s(F_i) \cap s(F_{i+1}) \neq \emptyset$ for $i = 1, 2, \cdots n$.

From the corollary to Lemma 9 we have that any linked set is a countable union of sets of diameter less than one.

2.3. *Parameter space.* We call a topological space $(\Theta, \mathfrak{J})$ a parameter space of the family $\mathfrak{F}$, if there exists a homeomorphic (i.e. 1-1 and bi-continuous) mapping $(\mathfrak{F}, \mathfrak{D}) \leftrightarrow (\Theta, \mathfrak{J})$. If $\mathfrak{F}$ is ordered, it is natural to take for Θ a simply ordered space and to require in addition the mapping $\mathfrak{F} \leftrightarrow \Theta$ to be order isomorphic.

A topological space with an order relation is called an "ordered topological space" if the given topology is finer (contains) the order topology (based on the open intervals). An important example of an ordered topological space is the Euclidean real line. More generally, any metric space with monotone distance functions is an ordered topological space. Therefore, $(\mathfrak{F}, \mathfrak{D})$ is an ordered topological space.

We remark, that for two connected ordered topological spaces, every homeomorphic mapping either preserves the order or reverses it (Eilenberg (1941), Theorem 4.1, p. 42).

In the applications, the parameter space usually is an interval on the real line. Therefore, if $\mathfrak{F}$ is ordered (1), for any homeomorphic mapping of $(\mathfrak{F}, \mathfrak{D})$ on the real line the order of $\mathfrak{F}$ is represented by the order of the parameter values.

3. The main theorems.

Theorem 1. *If $\mathfrak{F}$ is ordered* (2′) *or* (2″), *both distance functions d' and d'' induce the same topology, say $\mathfrak{D}$.*

The proof is obvious from (11) and Lemma 5.

Let $\mathfrak{F}^-$ be the $\mathfrak{D}$-completion of $\mathfrak{F}$. Then, we have

Theorem 2. *If $\mathfrak{F}$ is ordered* (2) *and of diameter less than* 1, *then $\mathfrak{F}^-$ is compact.*

The order of $\mathfrak{F}$ can be extended to $\mathfrak{F}^-$ in an obvious way: Let F', F'' be two elements of $\mathfrak{F}^-$ and $\{F_n'\}$, $\{F_n''\}$ two fundamental sequences d-converging to F' and F'', respectively. Then, if $F' \neq F''$, it is easily seen from Lemma 6 that there exists n_0, such that either $F_n' < F_n''$ or $F_n' > F_n''$ for all $n \geqq n_0$. Then, we define the order between F' and F'' as $F' < F''$ or $F' > F''$, respectively. As d-convergence implies pointwise convergence of distribution functions, the order relations (1)–(3) defined in terms of distribution functions immediately extend to $\mathfrak{F}^-$. This means: if $[\mathfrak{F}, >]$ is ordered (k), $\mathfrak{F}^-$ is the $\mathfrak{D}$-*completion* of $\mathfrak{F}$ and $[\mathfrak{F}^-, >]$ is the ordered space obtained from $[\mathfrak{F}, >]$ by extension of the order as outlined above, then $[\mathfrak{F}^-, >]$ is ordered (k) also.

If $(\Theta, \mathfrak{J})$ is a parameter space of the family $(\mathfrak{F}, \mathfrak{D})$ the completion $(\Theta^-, \mathfrak{J}^-)$ is a parameter space of $(\mathfrak{F}^-, \mathfrak{D}^-)$, where $\mathfrak{J}^-$ and $\mathfrak{D}^-$ are the topologies corresponding to $\mathfrak{J}$ and $\mathfrak{D}$, respectively.

Theorem 3. *If $\mathfrak{F}$ is ordered* (2) *and contains at most a countable number of degenerate distributions, then $\mathfrak{F}$ is $\mathfrak{D}$-separable and hence dominated.*

The proof of Theorem 3 is given in Section 5.

The assumption that $\mathfrak{F}$ contains at most a countable number of degenerate distributions is essential for separability (and hence domination) as the following trivial example shows: Let $\mathfrak{F}$ be the family of all degenerate distribution functions $F_\theta(t) = 0$ or 1 according as $t < \theta$ or $t \geqq \theta$ with $\theta \varepsilon [0, 1]$. Then $\mathfrak{F}$ is ordered by $F_{\theta'} < F_{\theta''}$ if $\theta' < \theta''$ in the sense of order (3). $\mathfrak{F}$ is, however not separable. (It contains a nondenumerable number of degenerate distributions.)

As every ordered separable space is homeomorphic to a subset of the real line, we have:

THEOREM 4. *If $\mathfrak{F}$ fulfills the assumptions of Theorem* 3, *then it has a parameter space which is a subset of the real line.*

We say that a distribution F_0 "fits" into an ordered family of distributions, if it is in an order relation (according to one of the definitions (1), (2), (3)) to each element of $\mathfrak{F}$ and if there exist $F', F'' \varepsilon \mathfrak{F}$ such that $F' < F_0 < F''$.

THEOREM 5. *Assume that $\mathfrak{F}$ is ordered* (2), *order dense, linked and $\mathfrak{D}$-complete. Then $\mathfrak{F}$ is complete in the sense that each F_0 which fits into $\mathfrak{F}$ is actually contained in $\mathfrak{F}$.*

The proof of Theorem 5 is given in Section 5.

4. A few lemmata.

LEMMA 1. *$d'(F_1, F_2) = 1$ if and only if $s(F_1) \cap s(F_2) = \emptyset$.*

PROOF. Assume that there exists $t_0 \varepsilon s(F_i)$, $i = 1, 2$. Then, $F_i(t_0) > 0$, $i = 1, 2$. Without loss of generality: $0 < F_2(t_0) \leqq F_1(t_0) \leqq 1$. If $F_1(t_0) < 1$, then $d'(F_1, F_2) \leqq \max(F_1(t_0), F_2(t_0 - 0) - F_1(t_0 - 0), 1 - F_2(t_0)) < 1$. Therefore, $F_1(t_0) = 1$. As $t_0 \varepsilon s(F_1)$, we have $F_1(t_0 - 0) < 1$. This implies $d'(F_1, F_2) \leqq \max(F_1(t_0 - 0), 1 - F_2(t_0)) < 1$. This establishes that $d'(F_1, F_2) = 1$ implies $s(F_1) \cap s(F_2) = \emptyset$. The converse follows easily from the fact that $s(F)$ is an interval.

LEMMA 2. *Let F^- be the distribution function defined by $F^-(t) = 1 - F(-t - 0)$. Then $F_1 < F_2$ in the sense of order $(2'')$ is equivalent to $F_2^- < F_1^-$ in the sense of order $(2')$. Furthermore, $d(F_1, F_2) = d(F_1^-, F_2^-)$ for both distances $d = d', d''$.*

PROOF. Trivial. We remark further that $s(F^-) = -s(F)$; $F^{--}(t) = F(t)$. Let $P(B)$ and $P^-(B)$ be the probability measures defined by F and F^-, respectively. Then, we have $P^-(B) = P(-B)$, where $-B = \{x : -x \varepsilon B\}$.

LEMMA 3. *Let $t' \leqq \inf B$ and $t'' \geqq \sup B$ and assume that $P_1(B) + P_2(B) > 0$. If $F_1 < F_2$ according to order $(2')$, we have*

$$F_2(t')/F_1(t') \leqq P_2(B)/P_1(B).$$

If $F_1 < F_2$ according to order $(2'')$, we have

$$P_2(B)/P_1(B) \leqq [1 - F_2(t'' - 0)]/[1 - F_1(t'' - 0)].$$

PROOF: We give the proof for the first assertion. The proof of the second assertion follows from Lemma 2. To each $\epsilon > 0$, there exists a finite union of disjoint intervals $B_n = \bigcup_1^n I_j$ such that $P_i(B \Delta B_n) < \epsilon$, $i = 1, 2$ (see Halmos (1950), p. 58). Without loss of generality, we can assume that for each $j = 1, \cdots, n$ at least one of the values $P_1(I_j)$, $P_2(I_j)$ is positive. For, if there exists

j_0 such that $P_1(I_{j_0}) = P_2(I_{j_0}) = 0$, then $B'_{n-1} = \bigcup_{j=1, j\neq j_0}^{n} I_j$ satisfies $P_i(B \Delta B'_{n-1}) < \epsilon$, $i = 1, 2$. Therefore, we can assume that none of the expressions $P_2(I_j)/P_1(I_j)$ is of the indeterminate form 0/0. As $\tau \geqq t'$ for all $\tau \varepsilon B$, we can assume without loss of generality that $\tau \geqq t'$ for all $\tau \varepsilon B_n$, as $P_i(B \Delta B_n) < \epsilon$, $i = 1, 2$ implies $P_i(B \Delta (B_n \cap \{\tau : \tau \geqq t'\})) < \epsilon$, $i = 1, 2$. Therefore, if $F_1 < F_2$ we have from $(2')$: $F_2(t')/F_1(t') \leqq P_2(I_j)/P_1(I_j)$ for $j = 1, \cdots, n$, which implies $F_2(t')/F_1(t') \leqq P_2(B_n)/P_1(B_n)$. Since $\epsilon > 0$ is arbitrary, this implies $F_2(t')/F_1(t') \leqq P_2(B)/P_1(B)$.

COROLLARY. *If $F_1 < F_2$ according to order $(2')$, then F_1 is absolutely continuous with respect to F_2 in $s(F_2)$. If $F_1 < F_2$ according to order $(2'')$, then F_2 is absolutely continuous with respect to F_1 in $s(F_1)$.* (This corollary is due to R. Borges.)

PROOF. The first assertion implies the second one by Lemma 2. To prove the first assertion, we have to show that for any $B \subset s(F_2)$, $P_1(B) > 0$ implies $P_2(B) > 0$.

At first let $B_i \subset s(F_2)$ with $P_1(B_i) > 0$ and $t'_i \varepsilon s(F_2)$ for $t'_i = \inf B_i$. Lemma 3 implies $0 < F_2(t'_i)/F_1(t'_i) \leqq P_2(B_i)/P_1(B_i)$ whence $P_2(B_i) > 0$. Since any $B \subset s(F_2)$ with $P_1(B) > 0$ can be written as a countable union of B_1, the assertion follows.

LEMMA 4. *If $(2')$ or $(2'')$ is satisfied, then $d''(F_1, F_2) = 1$ implies $d'(F_1, F_2) = 1$.*

PROOF. Let $d''(F_1, F_2) = 1$. Then (9) and (10) imply $P_1(B_{12}) = 1$ and $P_2(B_{12}) = 0$. Define $B_1 = B_{12} \cap \{\tau : F_1(\tau - 0) < 1\}$. Let $t \varepsilon B_1$. Then $P_1\{\tau : \tau \geqq t\} = 1 - F_1(t - 0) > 0$. Hence from $P_1(B_{12}) = 1$, we have $P_1(B_1 \cap \{\tau : \tau \geqq t\}) > 0$. If $F_1 < F_2$ and $(2')$ is satisfied, then Lemma 3 implies

$$F_2(t)/F_1(t) \leqq P_2(B_1 \cap \{\tau : \tau \geqq t\})/P_1(B_1 \cap \{\tau : \tau \geqq t\}) = 0.$$

Therefore, $F_2(t) = 0$. Since $\sup\{F_1(t) : t \varepsilon B_1\} = 1$, we have $d'(F_1, F_2) \geqq \sup\{F_1(t) - F_2(t) : t \varepsilon B_1\} = 1$. The proof for $(2'')$ is analogous, q.e.d.

That order (1) is not sufficient to establish Lemma 4 can be seen from the following example: Let $P_1(1) = P_1(3) = \frac{1}{2}$ and $P_2(2) = P_2(4) = \frac{1}{2}$. Then for P_1, P_2 relation (1) is fulfilled. We have $d''(P_1, P_2) = 1$, but $d'(P_1, P_2) = \frac{1}{2}$.

LEMMA 5. *If F_1 and F_2 are ordered $(2')$ or $(2'')$ we have*

$$d''(F_1, F_2) \leqq 2d'(F_1, F_2)^{\frac{1}{2}}.$$

PROOF. Let $F_1 < F_2$ and put $\delta = d'(F_1, F_2)$. Inasmuch as the lemma is trivial for $\delta = 0$ and $\delta = 1$, we can assume that $\delta \varepsilon (0, 1)$.

(a) Assume that F_1 and F_2 are ordered $(2')$. Let $t_1 = \sup\{t : F_1(t) \leqq \delta^{\frac{1}{2}}\}$. At first we prove

$$P_1(B) - P_2(B) \leqq \delta^{\frac{1}{2}} \tag{15}$$

for $B \subset \{\tau : \tau \geqq t_1\}$. For $P_1(B) = 0$, (15) is trivial. For $P_1(B) > 0$, order $(2')$ implies by Lemma 3:

$$1 - P_2(B)/P_1(B) \leqq 1 - F_2(t_1)/F_1(t_1) \leqq \delta^{\frac{1}{2}} \leqq \delta^{\frac{1}{2}}/P_1(B).$$

The second inequality follows from $F_1(t_1) - F_2(t_1) \leqq \delta$ and $F_1(t_1) \geqq \delta^{\frac{1}{2}}$.

On the other hand $P_1(B) \leqq F_1(t_1 - 0) \leqq \delta^{\frac{1}{2}}$ for B with $B \subset \{\tau : \tau < t_1\}$. Hence (15) holds also in this case. From

$$B = [B \cap \{\tau : \tau < t_1\}] \cup [B \cap \{\tau : \tau \geqq t_1\}]$$

we obtain for arbitrary $B \varepsilon \mathfrak{B}$ that $P_1(B)P_2(B) \leqq 2\,\delta^{\frac{1}{2}}$. Hence it follows from (8) that $d''(F_1, F_2) \leqq 2\delta^{\frac{1}{2}}$.

(b) If F_1 and F_2 are ordered (2″) the assertion follows by Lemma 2.

LEMMA 6. *If* $\mathfrak{F}$ *is ordered* (1), *we have* $d'(F_0, F_1) \leqq d'(F_0, F_2)$ *whenever* $F_0 \leqq F_1 \leqq F_2$.

PROOF. $F_0 \leqq F_1 \leqq F_2$ implies $F_0(t) \geqq F_1(t) \geqq F_2(t)$ for all t, whence

$$0 \leqq F_0(t) - F_1(t) \leqq F_0(t) - F_2(t). \tag{16}$$

Hence

$$d'(F_0, F_1) = \sup_t(F_0(t) - F_1(t)) \leqq \sup_t(F_0(t) - F_2(t)) = d'(F_0, F_2). \tag{17}$$

COROLLARY. *Let* $s(F_0) = s(F_1) = s(F_2)$ *and* $F_0 \leqq F_1 \leqq F_2$ *in the sense of order* (2), *then* $d'(F_0, F_1) = d'(F_0, F_2)$ *implies* $F_1 = F_2$. *Therefore, if* $\mathfrak{F}$ *is ordered* (2), d' *is* strictly *monotone increasing.*

PROOF. Since $F_i(t)$ is continuous from the right, there exists a t_0 such that $d'(F_0, F_1) = F_0(t_0) - F_1(t_0)$ or $d'(F_0, F_1) = F_0(t_0 - 0) - F_1(t_0 - 0)$. Hence, by (16) and (17) $d'(F_0, F_1) = d'(F_0, F_2)$ implies

$$F_1(t_0) = F_2(t_0) \tag{18}$$

or

$$F_1(t_0 - 0) = F_2(t_0 - 0). \tag{19}$$

Since $s(F_i)$ does not depend on i, both quantities are in (0, 1).

From (2′) we obtain for $t \geqq t_0$

$$F_2(t_0 - 0)/F_1(t_0 - 0) \leqq F_2(t_0)/F_1(t_0) \leqq F_2(t)/F_1(t) \leqq 1,$$

and from (2″) for $t < t_0$

$$\frac{1 - F_2(t_0)}{1 - F_1(t_0)} \geqq \frac{1 - F_2(t_0 - 0)}{1 - F_1(t_0 - 0)} \geqq \frac{1 - F_2(t)}{1 - F_1(t)} \geqq 1.$$

Both relations together yield $F_1(t) = F_2(t)$ for all t, regardless of whether (18) or (19) are true.

That both (2′) and (2″) are necessary to establish strict monotonicity can be seen by supplementing the example at the end of Section 2.1 with $P_3(1) = P_3(2) = P_3(3) = 4/21$, $P_3(4) = 9/21$. Then, the order $F_1 < F_3 < F_2$ fulfills (2′) and we have $d'(F_1, F_2) = d'(F_1, F_3) = 5/28$, though $P_2 \neq P_3$.

Another example due to R. Borges shows the necessity of the assumption that $s(F_i)$ is independent of i. Let $P_1(0) = P_1(1) = 1/2$, $P_1(2) = 0$; $P_2(0) = P_2(2) = 0$, $P_2(1) = 1$; $P_3(0) = 0$, $P_3(1) = P_3(2) = 1/2$. Then $F_1 < F_2 < F_3$ in the sense of order (2), but $d'(F_1, F_2) = d'(F_1, F_3)$.

According to Helly's weak compactness theorem (see e.g. Loève (1960) p. 179) any sequence of distribution functions $\{F_n\}$ contains a subsequence, converging to a limit function F at all continuity points of F. This limit function is nondecreasing. It is, however, not necessarily a distribution function.

LEMMA 7. *If $\mathfrak{F}$ is ordered* (2), *then any sequence $\{F_n\}$ with $D(\{F_n\}) < 1$ contains a d-convergent subsequence.*

If, in a metric space, each bounded sequence has a Cauchy subsequence, then this metric space is totally bounded (precompact). Thus, we have the following:

COROLLARY. *If $\mathfrak{F}$ is ordered* (2), *then any subset of $\mathfrak{F}$ with diameter less than* 1 *is totally bounded* (*precompact*).

PROOF OF LEMMA 7. From the sequence $\{F_n\}$ we select a monotone subsequence, which will be assumed nonincreasing. In order to simplify our notation, we also denote this subsequence by $\{F_n\}$. As $F_n(t) \leqq F_{n+1}(t)$, $\lim_{n\to\infty} F_n(t)$ exists for each t. It will be denoted by $F(t)$. Obviously, $F(t)$ is nondecreasing, $F(-\infty) = 0$ and $F(+\infty) = 1$.

At first, suppose that there exists t_0 such that $0 < F(t_0) < 1$. Since the order (2′) is transitive $F_n(t)/F_m(t)$ is nondecreasing for $m \geqq n$. Hence letting $m \to \infty$, $F_n(t)/F(t)$ is nondecreasing. By $0 < F(t_0) < 1$ this implies that $|F(t) - F_n(t)| = F(t)(1 - F_n(t)/F(t)) \leqq 1 - F_n(t_0)/F(t_0)$ for $t \geqq t_0$. Similarly order (2″) implies $|F(t) - F_n(t)| \leqq (1 - F_n(t_0))/(1 - F(t_0)) - 1$ for $t \leqq t_0$. Letting $n \to \infty$, the last two inequalities yield that $F_n(t) \to F(t)$ as $n \to \infty$ uniformly on the extended real line. Hence $F(t)$ is a distribution. For, the uniform limit of right continuous functions is right continuous. Thus $d'(F_n, F) \to 0$ as $n \to \infty$.

If there exists no t_0 with $0 < F(t_0) < 1$ it follows that $F(t) = 0$ or 1 according as $t < t_0$ or $t > t_0$ for some t_0. We can assume that t_0 is finite, since the order (2) and the distance function d' is invariant against a monotone transformation of the extended real line on the interval [0, 1]. By choice of the sequence F_n we have $|F_m(t) - F_n(t)| \leqq \delta < 1$ for all n, m and t. Hence, letting $m \to \infty$ we obtain $|F(t) - F_n(t)| \leqq \delta$ for all n and t. Letting $t \to t_0 + 0$ we obtain that $0 < 1 - \delta \leqq F_n(t_0)$ for all n. Therefore by (2′) $F_n(t_0)/F_m(t_0) \leqq F_n(t)/F_m(t)$ for all $t \geqq t_0$ and all $m \geqq n$. Letting $m \to \infty$ we obtain that $F_n(t_0)/F(t_0) \leqq F_n(t)$; letting $t \to t_0 + 0$ we obtain that $F(t_0) = 1$. Since $|F(t) - F_n(t)| = 1 - F_n(t) \leqq 1 - F_n(t_0)$ for $t \geqq t_0$ and $0 = F(t) \geqq F_n(t)$ otherwise, it follows that $d'(F_n, F) \to 0$ which completes the proof for monotone nonincreasing sequences.

The proof for monotone nondecreasing sequences follows from this result and Lemma 2.

LEMMA 8. *Let $(\mathcal{Y}, d)$ be a metric space with a distance function $d(y_1, y_2)$ bounded by one and Y a linked set in $\mathcal{Y}$. Then for each $y_0, y_1 \in Y$ there exists a chain $y_1, y_2, \cdots, y_n, y_{n+1}$ such that $y_{n+1} = y_0$, $y_i \in Y$ and $d(y_i, y_{i+1}) < 1$ for $i = 1, 2, \cdots, n$, n being some integer depending on y_0, y_1.* (This lemma due to R. Borges.)

PROOF. Let $y_1 \in Y$ be given. Denote by Y_1 the set of all $y \in Y$ such that a chain $y_1, y_2, \cdots, y_n, y_{n+1} = y$ with the properties stated in the lemma exists.

Assume that $Y - Y_1$ is not empty. Then, $d(Y_1, Y - Y_1) = 1$, as $d(y', y'') = 1$ for all $y' \in Y_1$ and $y'' \in Y - Y_1$. This, however, contradicts the assumption that Y is linked.

Lemma 9. *Let* $(\mathcal{Y}, d)$ *be a totally ordered metric space with a monotone distance function bounded by* 1. *If* $\mathcal{Y}$ *is linked, then any order bounded set is a finite union of sets of diameter less than* 1.

Corollary. *If* $\mathcal{Y}$ *is linked, then* $\mathcal{Y}$ *is a countable union of sets of diameter less than* 1.

Proof. Starting from an arbitrary value y_1, we construct a sequence $\{y_n\}$ according to the following procedure:

(a) If $\sup\{d(y_1, y) : y \geqq y_1\} < 1$ then the assertion is trivial. If $\sup\{d(y_1, y) : y \geqq y_1\} = 1$, we have to distinguish two cases:

(b) First, we assume that $d(y_1, y) < 1$ for all $y \geqq y_1$. Then, we construct a sequence $\{y_n\}$ such that $d(y_1, y_n) \geqq \sum_{i=1}^{n} 2^{-i}$, terminating the sequence in case $\sup\{d(y_n, y) : y \geqq y_n\} < 1$ should ever occur. As $d(y_1, y_n) < 1$, each distance $d(y_{n-1}, y_n)$ is also less than 1. Whatever the value $y > y_1$ might be, there exists an n, such that $\sum_{i=1}^{n} 2^{-i} \geqq d(y_1, y)$, which implies $y \leqq y_n$.

(c) If $Y_2 = \{y : d(y_1, y) = 1, y \geqq y_1\}$ is not empty, there exists a value $y_2 < Y_2$, such that $d(y_2, Y_2) < 1$. (Otherwise, $\mathcal{Y}$ would not be linked.) From the definition of Y_2, we have $y_1 < y_2$ and $d(y_1, y_2) < 1$. Let $Y_3 = \{y : d(y_2, y) = 1, y \geqq y_2\}$. Obviously, $Y_3 \subset Y_2$. If Y_3 is empty, we proceed as in (a) or (b) for the case of y_1. If Y_3 is not empty, there exists $y_3 \varepsilon Y_2 - Y_3$ such that $d(y_3, Y_3) < 1$. (Otherwise, $\mathcal{Y}$ would not be linked.) From the definition of Y_2 and Y_3, we have $y_2 < y_3$ and $d(y_2, y_3) < 1$. Proceeding in this way, we might either arrive at case (a) or (b) or remain in case (c). For the latter case, we have to show that to each $y > y_1$ there exists an n, such that $y \leqq y_n$. Assume, that this is not the case. Then, the set U of all y which are upper bounds of $\{y_n\}$ without belonging to $\{y_n\}$ is not empty. As $\mathcal{Y}$ is linked, there exists $y' < y''$, $y' \not\varepsilon U$, $y'' \varepsilon U$, such that $d(y', y'') < 1$. Since $y' \not\varepsilon U$, there exists n, such that $y' \leqq y_n$. Therefore, $d(y_n, y'') < 1$, which implies that $y'' < Y_{n+1}$. Since $y_{n+2} \varepsilon Y_{n+1}$, we have $y'' < y_{n+2}$, which is a contradiction.

Therefore, the sequence constructed above reaches every value $y > y_1$ within a finite number of steps. As the same procedure can be applied for $y < y_1$, this proves the Lemma as well as the Corollary.

Lemma 10. *Let* $(\mathcal{Y}, d)$ *be a metric space with a distance function bounded by* 1. *Then,* $\mathcal{Y}$ *is the union of linked components of distance* 1.

Proof. Let $\mathcal{C}_{y_0}$ the class of all linked sets containing y_0. Then, $C_0 = \bigcup \{C : C \varepsilon \mathcal{C}_{y_0}\}$ is called the linked component of y_0. The linked component C_0 itself is linked, as any union of linked sets with nonempty intersection is linked. Furthermore, if C_0 and C_1 are two linked components with distance less than 1, then $C_0 \cup C_1$ is linked. As $y_0 \varepsilon C_0 \cup C_1$ and as C_0 is the largest linked set containing y_0, we have $C_1 \subset C_0$. By the same reasoning, we obtain $C_1 \supset C_0$ whence $C_1 = C_0$.

5. Proof of the main theorems.

Proof of Theorem 3. By Lemma 6 $(\mathfrak{F}, d')$ is a totally ordered space with monotone distance functions. By Lemma 10 $\mathfrak{F}$ is the union of linked components of distance 1. Each of these linked components—except those consisting of a single degenerate distribution—has a support which is an interval. As the linked components have distance 1, the supports have pairwise empty intersections. As the number of intervals with pairwise empty intersections on the real line is countable, and since furthermore the number of isolated degenerate distributions is countable by assumption the number of linked components of $\mathfrak{F}$ is countable.

By the Corollary to Lemma 9 each linked component is a countable union of sets of diameter less than 1 and hence $\mathfrak{F}$ itself is a countable union of sets of diameter less than 1. Therefore by Theorem 2 $\mathfrak{F}$ is separable.

As $\mathfrak{F}$ is $\mathfrak{D}$-separable, more specific: separable under the metric d'', it is dominated according to a well-known theorem (see e.g. Lehmann (1959) p. 352, Theorem 1).

PROOF OF THEOREM 5. Let $\mathfrak{F}_1 = \{F : F \varepsilon \mathfrak{F}, F < F_0\}$ and $\mathfrak{F}_2 = \{F : F \varepsilon \mathfrak{F}, F > F_0\}$. Since F_0 fits into $\mathfrak{F}$, none of these sets is empty. Since $\mathfrak{F}$ is linked, there exists $F_i \varepsilon \mathfrak{F}_i$, such that $d(F_1, F_2) < 1$. Therefore, $\sup\{d(F_1, F) : F \varepsilon \mathfrak{F}_1, F > F_1\} \leqq d(F_1, F_2) < 1$. From Lemma 7 there exists a least upper bound $(\mathfrak{D})$, say F_1^*. As $\mathfrak{F}$ is $\mathfrak{D}$-complete, we have $F_1^* \varepsilon \mathfrak{F}$. The same reasoning shows that $\mathfrak{F}_2$ has a greatest lower bound, say $F_2^* \varepsilon \mathfrak{F}_2$. From the definition of $\mathfrak{F}_i$, we have $F_1^* \leqq F_0 \leqq F_2^*$. However, $F_1^* < F_2^*$ would contradict the assumption that $\mathfrak{F}$ is order dense in itself. Therefore, $F_1^* = F_2^* = F_0 \varepsilon \mathfrak{F}$. q.e.d.

6. Abstract sample spaces. Statistical applications. Let $\mathfrak{Q}$ be a class of probability measures Q on a measurable space $(\mathfrak{X}, \mathfrak{A})$.

We assume throughout this section that there exists a real valued statistic $T(x)$ which is pairwise sufficient of $\mathfrak{Q}$. (For the concept of pairwise sufficiency see Lehmann (1959) p. 56, Problem 9 or Halmos and Savage (1949) p. 234 ff). It will turn out that all of our considerations are immediately concerned with a finite or countably infinite number of probability measures only. The fact, that a pairwise sufficient statistic is sufficient for a finite or countably infinite number of probability measures explains why it is sufficient for our purpose to consider the family $\mathfrak{F}$ of distribution functions induced by $T(x) \rightarrow t$ on the real line: $F_Q(t) = Q\{x : T(x) \leqq t\}$. We shall say $Q_1 < Q_2$ in the sense of order (k), $k = 1, 2$ or 3, if $F_1 < F_2$ (we write F_n instead of F_{Q_n}) in the sense of order (k). We assume that $\mathfrak{Q}$ is ordered in the sense of order (k).

Furthermore, let $d'(Q_1, Q_2) = d'(F_1, F_2)$ and $d''(Q_1, Q_2) = \sup_{A \varepsilon \mathfrak{A}} |Q_1(A) - Q_2(A)|$. Since there exists a pairwise sufficient statistic $T(x)$ we have

$$d''(Q_1, Q_2) = d''(F_1, F_2). \tag{20}$$

For, let $Q_0 = Q_1 + Q_2$ and let $q_i(x)$ be the $\mathfrak{A}$-measurable density of Q_i with respect to Q_0. Then,

$$\sup_{A \varepsilon \mathfrak{A}} |Q_1(A) - Q_2(A)| = \int_{q_1(x) > q_2(x)} (q_1 - q_2)\, dQ_0 .$$

If T is sufficient for $\{Q_1, Q_2\}$ then the densities $q_i(x)$ can be taken to be functions of $T(x)$, i.e., $q_i(x) = p_i(T(x))$. Therefore, using (7),

$$\int_{q_1(x)>q_2(x)} (q_1(x) - q_2(x))\, dQ_0(x) = \int_{p_1(t)>p_2(t)} (p_1(t) - p_2(t))\, dP_0(t)$$

$$= \sup_{B \varepsilon \mathfrak{B}} |P_1(B) - P_2(B)|.$$

This proves (20).

Therefore all theorems and lemmata concerning the order (k) and the distance function $d''(F_1, F_2)$ in the real case extend immediately to the present case of a family of abstract probability measures, possessing a pairwise sufficient statistic and ordered (k).

We obtain as a corollary to Theorem 3:

THEOREM 6. *If $\mathfrak{Q}$ is ordered* (2) *and contains at most a countable number of degenerate distributions, then T is sufficient for $\mathfrak{Q}$.*

Finally, Theorem 4 shows that a conjecture of R. A. Fisher (1934) pp. 294–296 is essentially true. At least it is possible to interpret his statement such that it is true. Fisher's statement was, that a family of probability measures possessing an uniformly most powerful test for each level of significance must have a sufficient statistic and has only one parameter. Some phrases in Fisher's paper suggest[2] that in talking about uniformly most powerful tests he actually had in mind tests which are most powerful for testing any hypotheses ϑ_0 against any alternative $\vartheta_1 > \vartheta_0$. If tests of this kind exist for arbitrary levels of significance, then the family of distributions has monotone likelihood ratios, i.e. it is ordered (3), as was shown in [12]. If, in addition, this family contains at most a countable number of degenerate distributions, we obtain from Theorem 4 that it also has 1-dimensional real parameter space. As in [12], we call a test function φ *everywhere most powerful* in an ordered family $\mathfrak{Q}$ of probability measures, if φ is most powerful for Q_0 against Q_1 with $Q_0 < Q_1$ as long as $E_0\varphi > 0$ and $E_1\varphi < 1$. A family Φ of test functions is called *perfect* with respect to $\mathfrak{Q}$, if the following three conditions are satisfied: (i) all test functions contained in Φ are everywhere most powerful, (ii) for each $Q_0 \varepsilon \mathfrak{Q}$ and for each $\alpha \varepsilon [0, 1]$ there exists a test function $\varphi \varepsilon \Phi$ with $E_0\varphi = \alpha$, (iii) for $\alpha = 0(1)$ there exists a test function $\varphi \varepsilon \Phi$ which is most powerful for Q_0 against any Q with $Q > Q_0$ (for any Q against Q_0 with $Q < Q_0$). Therefore, we have proved the following

THEOREM 7. *A family of probability measures which has a perfect family of tests and contains at most a countable number of degenerate distributions is $\mathfrak{D}$-separable and therefore dominated. Furthermore, it has a parameter space which is a subset of the real line.*

The counterexample, given by Neyman and Pearson (1936), pp. 122–123 (see also Lehmann (1959) p. 110, Problem 2 (ii)) showing the existence of a family of distributions with two real parameters having uniformly most powerful tests, is of no relevance as it fails to meet the requirement that each of these tests is most powerful for every pair of probability measures of this family. In

[2] See e.g. p. 295: . . . the contours defined by the ratio of the likelihood of H_1 and H_0 shall be the same as those defined by the ratios of the likelihood of any two hypotheses in the class.

the example by Neyman and Pearson, to each hypothesis specified by the parameter value (a_0, b_0), there exists a test most powerful against the class of alternatives $\{(a, b) : a \geqq a_0, b \leqq b_0\}$. This test has, however, no optimal properties against alternatives (a, b) with $a > a_0, b > b_0$ or $a < a_0, b < b_0$.

Acknowledgment. I wish to thank Mr. R. Borges for his critical comments and valuable suggestions. Furthermore, I wish to thank the referee for his suggestions which contributed much to make the paper more readable.

REFERENCES

[1] Bourbaki, N. (1952). Integration. *Actualités Sci. Ind.* 1175. Hermann, Paris.
[2] Doob, L. J. (1953). *Stochastic Processes*. Wiley, New York.
[3] Eilenberg, S. (1941). Ordered topological spaces. *Amer. J. Math.* **63** 39–45.
[4] Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. London Ser. A* **144** 285–307.
[5] Halmos, P. R. (1950). *Measure Theory*. Van Nostrand, New York.
[6] Halmos, P. R. and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.
[7] Kelley, J. L. (1955). *General Topology*. Van Nostrand, New York.
[8] Lehmann, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26** 399–419.
[9] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
[10] Loève, M. (1960). *Probability Theory*, (2nd. ed.). Van Nostrand, New York.
[11] Neyman, J. and Pearson, E. S. (1936). Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Statist. Res. Mem. London* **1** 113–137.
[12] Pfanzagl, J. (1962). Überall trennscharfe Tests und monotone Dichtequotienten. *Z. Wahrscheinlichkeitstheorie* **1** 109–115.
[13] Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.