The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1262–1272

THE TOPOLOGY OF DISTINGUISHABILITY[1]

BY LLOYD FISHER AND JOHN W. VAN NESS
University of Washington

1. Introduction. Hoeffding and Wolfowitz [6] consider the following problem. Independent identically distributed observations are sequentially taken on a random vector X with distribution function F. All that is known is that F belongs to a given family $\mathscr{J}$. It is desired to eventually decide either $F \in \mathscr{G}$ or $F \in \mathscr{H}$, where $\mathscr{G}$ and $\mathscr{H}$ are disjoint subsets of $\mathscr{J}$, in such a way that if $f \in \mathscr{G} \cup \mathscr{H}$ the probability of error is less than any preassigned number greater than zero.

Freedman [4] studies a modification of this problem which supposes that F is known only to be a member of a countable family $\mathscr{J}$ and it is desired to eventually decide with prescribed accuracy which member of $\mathscr{J}$ is F. This same problem is also considered in [2] and [3].

In this paper the framework is such that one would like to distinguish between a countable number of families of probability distributions, thus including both of the approaches mentioned above. Section 2 shows that under certain restrictions there is an appropriate topology associated with this question. The results of Section 2, although posed in a more general framework than Freedman's work, are essentially an easy extension of his paper. LeCam and Schwartz [8] consider estimation problems which are relevant to distinguishability by appropriate modifications.

Section 3 considers the problem of the metrizability of the topology considered. Necessary and sufficient conditions (that are somewhat difficult to apply) are given. An important example is given to show that often the topology may be nonmetrizable. The remainders of Section 3 and Section 4 give cases where the topology is metrizable and describe appropriate metrics.

2. The topology of distinguishability. Let Ω be a set and $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \mathscr{A}_3 \subset \cdots$ be a non-decreasing sequence of σ-fields on Ω. Let $\mathscr{A} = \bigcup_{n=1}^{\infty} \mathscr{A}_n \subseteq \mathscr{B}$ where $\mathscr{B}$ is a σ-field. Take $\mathscr{M}$ to be a set of probability measures on $(\Omega, \mathscr{B})$.

Let $\mathscr{C} = \{C_\alpha, \alpha \in A\}$ be a family of disjoint subsets of $\mathscr{M}$. A decision rule for $\mathscr{C}$ written Φ, is a collection of functions ϕ_n^{α}, $\alpha \in A$, $n = 1, 2, \cdots$ on Ω with the following properties:

(a) each $\phi_n^{\alpha}(\cdot)$ is a real measurable function on $(\Omega, \mathscr{A}_n)$,
(b) $\phi_n^{\alpha} \geqq 0$,
(c) $\phi_{n+1}^{\alpha} \geqq \phi_n^{\alpha}$, and
(d) $\sum_{\alpha \in A} \phi_n^{\alpha} \leqq 1, n = 1, 2, \cdots$.

Received February 24, 1969; revised December 4, 1970.
[1] Research supported by NSF Grants GP-7519 and GP-13163 at the University of Washington.

The quantity $\phi_n^{\alpha}(\omega)$ is interpreted as the probability given ω that one decides on or before the nth trial that the underlying measure is a point in C_α. Thus if A is denumerable $1-\sum_{\alpha\in A}\phi_n^{\alpha}(\omega)$ is the probability given ω that we will have to continue observing beyond time n. Condition (c) is a consistency condition. Denote by $R(\mathscr{C})$ the collection of all decision rules for $\mathscr{C}$.

The *distinguishability of* $\mathscr{C}$ *under* Φ is defined to be

$$D(\Phi,\mathscr{C}) = \inf_{\alpha\in A}\inf_{\mu\in C_\alpha}\lim_{n\to\infty}\int_\Omega \phi_n^{\alpha}\,d\mu.$$

The *distinguishability* of $\mathscr{C}$ is defined to be:

$$D(\mathscr{C}) = \sup_{\Phi\in R(\mathscr{C})} D(\Phi,\mathscr{C}).$$

We say $\mathscr{C}$ is *distinguishable* if $D(\mathscr{C}) = 1$.

THEOREM 2.1. *Let* $\mathscr{C} = \{C_\alpha, \alpha\in A\}$ *be such that A is countable and each C_α is a countable subset of* $\mathscr{M}$. *Then the following are equivalent*:

(i) $\mathscr{C}$ *is distinguishable.*

(ii) *For every* $\alpha\in A$ *and every* $\mu\in C_\alpha$, $\mathscr{D} = \{\{\mu\}, \bigcup_{\beta\neq\alpha}C_\beta\}$ *is distinguishable.*

(iii) *For every* $\alpha\in A$, $\mathscr{D} = \{C_\alpha, \bigcup_{\beta\neq\alpha}C_\beta\}$ *is distinguishable.*

(iv) *For every* $\alpha\in A$, *every* $\mu\in C_\alpha$, *and every* $0<\varepsilon<1$ *there exists a* $G\in\mathscr{A}$ *such that* $\{\lambda: |\lambda(G)-\mu(G)| < 1-\varepsilon\}\cap\bigcup_{\beta\neq\alpha}C_\beta = \varnothing$.

PROOF. (i) implies (iii). Let $\varepsilon>0$ be given. Choose Φ such that $D(\Phi,\mathscr{C}) > 1-\varepsilon$. Let $\Phi^*(\mathscr{D}) = \{\phi_n^{\alpha}, \phi_n^{U}\}_{n=1}^{\infty}$ where $\phi_n^{U} = \sum_{\beta\neq\alpha}\phi_n^{\beta}$ is measurable since the sum is countable. Then

$$D(\Phi^*,\mathscr{D}) \geqq \min\left[\inf_{\mu\in C_\alpha}\lim_{n\to\infty}\int\phi_n^{\alpha}\,d\mu, \inf_{\mu\in\cup_{\beta\neq\alpha}C_\beta}\lim_{n\to\infty}\int\sum_{\beta\neq\alpha}\phi_n^{\beta}\,d\mu\right]$$
$$\geqq D(\Phi,\mathscr{C}) > 1-\varepsilon$$

since

$$\inf_{\mu\in\cup_{\beta\neq\alpha}C_\beta}\lim_{n\to\infty}\int\sum_{\beta\neq\alpha}\phi_n^{\beta}\,d\mu \geqq \inf_{\beta_0\neq\alpha}\inf_{\mu\in C_{\beta_0}}\lim_{n\to\infty}\int\phi_n^{\beta_0}\,d\mu.$$

(iii) implies (ii). Note that $\mu\in C_\alpha$ implies $\lim_{n\to\infty}\int\phi_n^{\alpha}d\mu \geqq \inf_{\gamma\in C_\alpha}\lim_{n\to\infty}\int\phi_n^{\alpha}d\lambda$.

(ii) implies (iv). By (ii), choose ϕ_n^{1} and ϕ_n^{2}, $n = 1, 2, \cdots$ such that

$$\lim_{n\to\infty}\int\phi_n^{1}\,d\mu > 1-\varepsilon/2.$$
$$\inf_{\mu\in\cup_{\beta\neq\alpha}C_\beta}\lim_{n\to\infty}\int\phi_n^{2}\,d\mu > 1-\varepsilon/2.$$

Pick m_0 such that if $m \geqq m_0 \int\phi_m^{1}\,d\mu > 1-\varepsilon/2$ and let $G = \{\omega: \phi_m^{1}(\omega) \geqq \frac{1}{2}\}$, then $1-\varepsilon/2 < \int\phi_m^{1}\,d\mu \leqq \mu(G^c)/2+\mu(G) = \frac{1}{2}+\mu(G)/2$ (where G^c is the complement of G) or $\mu(G) > 1-\varepsilon$. Since $\phi_n^{1}+\phi_n^{2} \leqq 1$, and since there exists an $n(\alpha)$ such that for $\lambda\in\bigcup_{\beta\neq\alpha}C_\beta$, $n\geqq n(\alpha)$ and $1-\varepsilon/2 < \int\phi_n^{2}\,d\lambda$, we have

$$1-\varepsilon/2 < \int\phi_n^{2}\,d\lambda < \lambda(G)/2+\lambda(G^c) = 1-\lambda(G)/2$$

or $\lambda(G) < \varepsilon$ if $m = n$, which can be done by taking m sufficiently large.

(iv) implies (i). Order the measures in $\bigcup_{\alpha\in A}C_\alpha$ to form the sequence $\{\mu_n\}_{n=1}^{\infty}$. Given $\varepsilon > 0$, choose $A_i \in \mathscr{A}_{n_i}$ such that $\mu_i(A_i) > 1-\varepsilon/2^{i+1}$ and $\mu_j(A_i) < \varepsilon/2^{i+1}$, for all $\mu_j \in \bigcup_{\beta\neq\alpha}C_\beta$ where $\mu_i \in C_\alpha$. Since the $\mathscr{A}_n$ are non-decreasing, we may without loss of generality assume that $n_i < n_{i+1}$. Let $B_i = A_i - \bigcup_{j<i}A_j$ and $D(\alpha, n) = \bigcup_{\mu_i\in C_\alpha, n_i \leqq n} B_i$. If $\phi_n{}^\alpha$ is the indicator function of $D(\alpha, n)$ then $0 \leqq \phi_n{}^\alpha$, $\phi_n{}^\alpha \leqq \phi_{n+1}^\alpha$ and $\phi_n{}^\alpha$ is measurable $(\Omega, \mathscr{A}_n)$. Since the sets $D(\alpha, n)$ are disjoint $\sum_{\alpha\in A}\phi_n{}^\alpha \leqq 1$. Take $\Phi = \{\phi_n{}^\alpha, \alpha \in A, n = 1, 2, \cdots\}$. If $\mu_i \in C_\alpha$

$$\lim_{n\to\infty} \int \phi_n{}^\alpha \, d\mu_i \geqq \mu_i(D(\alpha, n_i)).$$

But, $D(\alpha, n_i)$ contains $A_i - \bigcup'_{j<i}A_j$ where $\bigcup'_{j<i}A_j$ is the union of the A_j such that $\mu_j \notin C_\alpha$. Thus,

$$\mu_i(D(\alpha, n_i)) > 1-\varepsilon/2^{i+1} - \textstyle\sum'_{j<i}\mu_i(A_j) > 1-\varepsilon/2^{i+1} - \varepsilon/2 > 1-\varepsilon.$$

Thus $D(\Phi, \mathscr{C}) \geqq 1-\varepsilon$. □

Most of the research in distinguishability published to date is primarily concerned with observing independent identically distributed random variables which are governed by some unknown probability distribution. This paper will also follow this pattern.

Let $\mathscr{M}_1$ be a set of probability measures on the measurable space $(S, \mathscr{F})$ and let $\Omega = S\times S\times S\times\cdots$. $\mathscr{A}_n$ is the σ-field of subsets of Ω generated from $\mathscr{F}$ by the first n-coordinates, and $\mathscr{M}$ consists of the product measures on Ω generated by the product of each of the elements of $\mathscr{M}_1$ with itself. If $\mu \in \mathscr{M}_1$, denote its product measure in $\mathscr{M}$ by $\boldsymbol{\mu}$. Restrict $\mathscr{B}$ to be the smallest σ-field containing $\mathscr{A}$. We will frequently use one of the following conditions.

ASSUMPTION I. The quantities $\mathscr{M}$ and $(\Omega, \mathscr{B})$ are generated in the above manner.

ASSUMPTION II. Assumption I is satisfied, S is a Polish space, and $\mathscr{F}$ is the class of Borel sets.

Let $\mathscr{T}$ be the topology on $\mathscr{M}$ given by pointwise convergence on $\mathscr{A}$. A subbase for $\mathscr{T}$ is given by sets of the form $\{\lambda: \lambda \in \mathscr{M}, |\lambda(G)-\mu(G)| < \varepsilon\}$ where $G \in \mathscr{A}$, $\boldsymbol{\mu} \in \mathscr{M}$ and $\varepsilon > 0$ are fixed.

THEOREM 2.2. *Let Assumption* I *hold and* $\mathscr{C} = \{C_\alpha, \alpha \in A\}$ *have* A *countable and each* C_α *countable, then the following statements are equivalent*:

(i) $\mathscr{C}$ *is distinguishable.*

(ii) *For every* $\alpha \in A$ *and every* $\mu \in C_\alpha$, $\{\boldsymbol{\mu}\}$ *and* $\bigcup_{\beta\neq\alpha}C_\beta$ *are separated* $(\mathscr{T})$, *that is*, $\{\boldsymbol{\mu}\} \cap \overline{\bigcup_{\beta\neq\alpha}C_\beta} = \varnothing$. ($\bar{G}$ *denotes the* $\mathscr{T}$ *closure of* G.)

(iii) *For every* $\alpha \in A$, C_α *and* $\bigcup_{\beta\neq\alpha}C_\beta$ *are separated, that is*

$$(\bar{C}_\alpha \cap \textstyle\bigcup_{\beta\neq\alpha}C_\beta) \cup (C_\alpha \cap \overline{\bigcup_{\beta\neq\alpha}C_\beta}) = \varnothing.$$

PROOF. (ii) iff (iii). This is clear from the equations

$$C_\alpha \cap \overline{\textstyle\bigcup_{\beta\neq\alpha}C_\beta} = \textstyle\bigcup_{\mu\in C_\alpha}(\{\boldsymbol{\mu}\} \cap \overline{\bigcup_{\beta\neq\alpha}C_\beta})$$

and

$$\bar{C}_\alpha \cap \textstyle\bigcup_{\beta\neq\alpha}C_\beta = \textstyle\bigcup_{\lambda\in\cup_{\beta\neq\alpha}C_\beta}(\bar{C}_\alpha \cap \{\boldsymbol{\lambda}\}).$$

(i) implies (ii). Condition (iv) of Theorem 2.1 gives a neighborhood of $\boldsymbol{\mu} \in C_\alpha$ which has empty intersection with $\bigcup_{\beta \neq \alpha} C_\beta$.

(ii) implies (i). If (ii) holds and $\boldsymbol{\mu} \in C_\alpha$, we may find an $\varepsilon > 0$ and sets $A_1, \cdots, A_n$ such that

$$\{\boldsymbol{\lambda}: |\boldsymbol{\lambda}(A_i) - \boldsymbol{\mu}(A_i)| < \varepsilon, i - 1, \cdots, n\} \cap \bigcup_{\beta \neq \alpha} C_\beta = \varnothing$$

since sets of this type form a base for $\mathscr{T}$. Pick an m such that $A_j \in \mathscr{A}_m, j = 1, \cdots, n$. By observing data in blocks of m and observing the proportion of points in each A_i we may find a set $A \in \mathscr{A}_k$ for some k such that $\boldsymbol{\mu}(A) > 1 - \varepsilon$ and $\boldsymbol{\theta}(A) < \varepsilon$ for $\boldsymbol{\theta}$ not in

$$\{\boldsymbol{\lambda}: |\boldsymbol{\lambda}(A_i) - \boldsymbol{\mu}(A_i)| < \varepsilon, i = 1, 2, \cdots, n\}.$$

Thus (ii) implies (iv) of Theorem 2.1.

The proof of the following is contained in the above proof.

COROLLARY 2.3. *Under Assumption* I, *sets of the form* $\{\boldsymbol{\lambda}: |\boldsymbol{\lambda}(A) - \boldsymbol{\mu}(A)| < \varepsilon\}$, $A \in \mathscr{A}$, $\varepsilon > 0$, $\boldsymbol{\mu} \in \mathscr{M}$ *form a base for* $\mathscr{T}$.

3. Metrizability of $\mathscr{T}$. In this section we examine the question of metrizing the topology $\mathscr{T}$. First, it is shown that $\mathscr{T}$ is not metrizable in many cases of interest. Secondly, necessary and sufficient conditions on $\mathscr{M}$ and $\mathscr{C}$ are given for metrizability. Under I, let $\mathscr{T}_1$ be the topology on $\mathscr{M}_1$ induced by the map $\mu \to \boldsymbol{\mu}$. It is then seen that the topology $\mathscr{T}_1$ lies between the weak and strong topologies, and conditions are considered which imply that the two topologies are the same, thus giving the metrizability.

To show the nonmetrizability, we will use the following:

LEMMA 3.1. *Let μ be a probability measure on $(S, \mathscr{F})$. Let μ_n be the random sample measure obtained by sampling according to μ independently n times. For any fixed positive integer k and measure λ on $(S, \mathscr{F})$ let λ^k be the* kth *product measure generated by λ on the product measurable space $(S^k, \mathscr{F}^k)$. Then for any $A \in \mathscr{F}^k$*

$$P\{(\mu_n)^k(A) \to \mu^k(A)\} = 1.$$

PROOF. The random set function $\mu_n{}^k(\cdot)$ is a function of $X_1, \cdots, X_n$ (i.i.d. $(S, \mathscr{F}, \mu)$ random variables); it puts mass $1/n^k$ at each of the n^k k-tuples with entries from $X_1, \cdots, X_n$. (Note: If $X_i = X_j$, $i \neq j$ we "bookkeep" these as separate entries.) Single point sets may not be in $\mathscr{F}^k$, so strictly speaking we do not put mass $1/n^k$ at a point in general, but add $1/n^k$ to the measure of each set in $\mathscr{F}^k$ containing the point.

We form a new sequence of random variables $Y_1, Y_2, \cdots$ as follows. As we consider the $n+1$st sample point $X_{n+1} (n+1 \geqq k)$ we form the $k!\binom{n}{k-1}$ k-tuples containing X_{n+1} that may be formed from $X_1, \cdots, X_{n+1}$ with no X_i repeated in the k-tuple. Order these k-tuples and add a new Y_i for each one. For Y_l corresponding to $(X_{n_1}, \cdots, X_{n_k})$,

$$\begin{aligned} Y_l &= 1 \quad \text{if} \quad (X_{n_1}, \cdots, X_{n_k}) \in A, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Let $Z_i = Y_i - \mu^k(A)$. Then $E(Z_i) = 0$, and the Z_i's have uniformly bounded variances, since $|Z_i| \leqq 1$. Set $S_n = \sum_{i=1}^n Z_i/n$. We estimate $|E(Z_n S_n)|$. $E(Z_n S_n) = \sum_{i=1}^n E(Z_i Z_n)/n$. Let Z_n occur in the block when the jth sample point was considered so that

$$k!\binom{j-1}{k} < n \leqq k!\binom{j}{k}.$$

Now $k!\binom{j-k}{k}$ of the Z_i, $i < n$ correspond to k-tuples whose entries do not occur in the k-tuple corresponding to Z_n. For such Z_i, $E(Z_i Z_n) = E(Z_i)E(Z_n) = 0$. Thus,

$$|E(Z_n S_n)| \leqq k![\binom{j}{k} - \binom{j-k}{k}]/n < [\binom{j}{k} - \binom{j-k}{k}]/\binom{j-1}{k} \leqq c/j$$

for some $c > 0$. Since $n \leqq k!\binom{j}{k} < j^k$

$$E(Z_n S_n) \leqq c/n^{1/k}.$$

Therefore the strong law of large numbers holds for S_n (see Parzen [11], Theorem 2B, page 420), i.e. $P(S_n \to 0) = 1$.

The k-tuples which have different entries coming from the same sample point can be neglected since $k!\binom{n}{k}/n^k \to 1$ as $n \to \infty$. The proof is completed by noting that

$$1 - k!\binom{n}{k}/n^k + (k!\binom{n}{k}/n^k)(S_{k!\binom{n}{k}} + \mu^k(A))$$

$$\geqq \mu_n{}^k(A) \geqq (k!\binom{n}{k}/n^k)(S_{k!\binom{n}{k}} + \mu^k(A)). \quad \square$$

The following shows that $\mathscr{T}$ is not metrizable in many cases of interest.

THEOREM 3.2. *Assume that* II *holds and that* $\mathscr{M}_1$ *is the set of all Borel probability measures on* S, *then* $\mathscr{T}$ *is metrizable* iff S *is countable.*

PROOF. The if statement follows from the subsequent portions of this section.

Suppose $\mathscr{T}$ is metrizable and S is uncountable. Let μ be a nonatomic measure on $(S, \mathscr{F})$ (such a μ exists, see Parthasarathy [10], Theorem 8.1, page 53). Since $\mathscr{T}$ is metrizable, it is first countable and by Corollary 2.3 we may find a sequence of sets $A_1, A_2, \cdots$ in $\mathscr{A}$ such that if

$$D_n = \{\lambda : |\mu(A_n) - \lambda(A_n)| < 1/n\}$$

then $\{D_n\}$ is a base for the neighborhood system of μ with respect to $\mathscr{T}_1$. By Lemma 3.1 we may find a discrete measure $\mu_n \in \mathscr{M}_1$ such that $|\mu_n(A_n) - \mu(A_n)| < 1/n$. Let s_n be the support of μ_n (a finite set) and let $A = \bigcup_{n=1}^\infty s_n$. Then A is countable and can be considered as a subset of $\mathscr{A}_1$. But $|\mu(A) - \mu_n(A)| = 1$ so that μ is not a limit point of $\{\mu_n\}_{n=1}^\infty$, while each member of the local base contains a member of $\{\mu_n\}_{n=1}^\infty$, giving a contradiction. $\square$

THEOREM 3.3. *Let* $\mathscr{B}$ *be the* σ*-field generated by* $\mathscr{A}$, *then* $(\mathscr{M}, \mathscr{T})$ *is a Tychonoff space and the following statements are equivalent*:

(i) $\mathscr{T}$ *is metrizable.*

(ii) $\mathscr{T}$ *has a* σ*-locally finite base.*

(iii) $\mathscr{T}$ *has a* σ*-discrete base.*

Furthermore, the following two statements are equivalent:

(i) $\mathscr{T}$ *is second countable.*

(ii) $\mathscr{T}$ *is separable and metrizable.*

PROOF. $\mathscr{M}$ is given the topology of pointwise convergence as a subset of the unit cube $[0, 1]^{\mathscr{A}}$, that is each coordinate corresponds to a set in $\mathscr{A}$. The map is one-to-one due to the assumption on $\mathscr{B}$. By Kelley [7], Theorem 7, page 118, $(\mathscr{M}, \mathscr{T})$ is a Tychonoff space. The space is thus T_1 and completely regular and hence regular. The first and second sets of equivalences thus follow respectively from Theorems 18 and 17, Kelley [7] pages 125–127. ▯

COROLLARY 3.4. *Under* I *if S is countable and every single point set is in $\mathscr{F}$ then $\mathscr{T}$ is metrizable and separable.*

PROOF. The space obviously has a countable base.

THEOREM 3.4. *Let* I *hold and $\mathscr{T}$ be second countable, then the following are equivalent:*

(i) $\mathscr{C} = \{C_\alpha, \alpha \in A\}$ *is distinguishable.*

(ii) *For every $\alpha \in A$ and every $\mu \in C_\alpha$, $\{\mu\}$ and $\bigcup_{\beta \neq \alpha} C_\beta$ are separated.*

(iii) *For every $\alpha \in A$, C_α and $\bigcup_{\beta \neq \alpha} C_\beta$ are separated.*

PROOF. The equivalence of (ii) and (iii) follows as in the proof of Theorem 2.2.

To show (ii) implies (i), we first note that $\mathscr{T}$ is second countable so that any open cover of a set has a countable subcover.

Next we will show that A must be a countable set. Define $D(\alpha) = \bigcup_{\beta \neq \alpha} C_\beta$. Let d be a metric for $(\mathscr{M}, \mathscr{T})$ and $H = \{\boldsymbol{\mu}_i; i = 1, 2, \cdots\}$ be a dense set. Choose $\boldsymbol{\mu}_\alpha \in C_\alpha$, then $d(\boldsymbol{\mu}_\alpha, D(\alpha)) > 0$. With each $\boldsymbol{\mu}_\alpha$ associate a $\boldsymbol{\mu}_{i(\alpha)} \in H$ such that $d(\boldsymbol{\mu}_\alpha, \boldsymbol{\mu}_i) < d(\boldsymbol{\mu}_\alpha, D(\alpha))/2$. Note that if $\beta \neq \alpha$, then $\boldsymbol{\mu}_{i(\alpha)} \neq \boldsymbol{\mu}_{i(\beta)}$ and A is in a one-to-one correspondence with a countable set.

For fixed $\alpha \in A$, cover C_α in the following manner. If $\boldsymbol{\mu} \in C_\alpha$ then $d(\boldsymbol{\mu}, D(\alpha)) > 0$ and by Corollary 2.3 we may find a $B \in \mathscr{A}_m$ (for some m) and an $\varepsilon > 0$ such that

$$\{\boldsymbol{\lambda}: |\boldsymbol{\lambda}(B) - \boldsymbol{\mu}(B)| < \varepsilon\} \subseteq \{\boldsymbol{\lambda}: d(\boldsymbol{\mu}, \boldsymbol{\lambda}) < d(\boldsymbol{\mu}, D(\alpha))/2\}.$$

If $A_\mu = \{\boldsymbol{\lambda}: |\boldsymbol{\lambda}(B) - \boldsymbol{\mu}(B)| < \varepsilon/2\}$ then $\bigcup_{\mu \in C_\alpha} A_\mu$ covers C_α and we pick a countable subcovering. Do this for each C_α and note that the coverings of C_α and C_β, $\beta \neq \alpha$, are disjoint.

Order these countable subcoverings of a countable number of sets as $A_1, A_2, \cdots$, where $A_i = \{\boldsymbol{\lambda}: |\boldsymbol{\lambda}(B_i) - \boldsymbol{\mu}_i(B_i)| < \varepsilon_i/2\}$. By observing in blocks of m_i and noting the number of the resulting m_i-tuples which fall into B_i we may find for $\varepsilon > 0$ a $G_i \in \mathscr{A}$ such that if

$$|\boldsymbol{\lambda}(B_i) - \boldsymbol{\mu}_i(B_i)| < \varepsilon_i/2, \qquad \text{then} \quad \boldsymbol{\lambda}(G_i) > 1 - \varepsilon/2^{i+1},$$

and if

$$|\boldsymbol{\lambda}(B_i) - \boldsymbol{\mu}_i(B_i)| > \varepsilon_i, \qquad \text{then} \quad \boldsymbol{\lambda}(G_i) < \varepsilon/2^{i+1}.$$

The proof is completed similarly to the end of the proof of Theorem 2.1.

To show that (i) implies (iii) we use the countability of A and proceed as in the proof of Theorem 2.1. ▯

COROLLARY. In Theorem 3.4, the statements imply A is a countable set.

Note that each C_α may contain an uncountable number of elements.

DEFINITION. Under II let $\mathscr{V}$ be the total variation topology on $\mathscr{M}_1$ and $\mathscr{L}$ be the weak topology on $\mathscr{M}_1$.

More explicitly, $\mathscr{V}$ is generated by metric $v(\mu, \lambda) = \sup_{A \in \mathscr{F}} |\mu(A) - \lambda(A)|$. The weak topology is generated by the Lévy–Prokhorov metric on $\mathscr{M}_1$. Let d be a metric which makes $(S, \mathscr{F})$ complete and separable. Let μ and λ be Borel measures on $(S, \mathscr{F})$, then $L(\mu, \lambda)$ is the infimum of those $\varepsilon > 0$ such that for each closed set A,

$$\mu(A) \leqq \lambda(A^\varepsilon) + \varepsilon$$

where $A^\varepsilon = \{y\colon \text{there is an } x \in A \text{ with } d(x, y) < \varepsilon\}$.

PROPOSITION 3.5. *Under* II, $\mathscr{L} \subseteq \mathscr{T}_1 \subseteq \mathscr{V}$.

PROOF. This result is known. See for example, LeCam and Schwartz [8].

COROLLARY 3.6. *Under* II *if* $\mathscr{L} = \mathscr{V}$ *then* $\mathscr{T}_1$ ($\mathscr{T}$) *is metrizable and the Lévy–Prokhorov or total variation metric may be used to give the topology.*

The approach of Corollary 3.6 has been used in Fisher and Van Ness [2] to metrize the topology of distinguishability. Another example is given by the following theorem.

THEOREM 3.7. *Assume that* II *holds with S a locally compact Abelian group having a translation invariant metric ρ. If each $\mu \in \mathscr{M}_1$ has a density f_μ with respect to Haar measure and if there exists a constant* $k > 0$ *such that for all $\mu \in \mathscr{M}_1$ and $x, y \in S$,* $|f_\mu(x) - f_\mu(y)| < k\rho(x, y)$ *then* $\mathscr{L} = \mathscr{V}$.

PROOF. To show that $\mathscr{L} = \mathscr{V}$ it is sufficient to show that under the uniform Lipschitz condition $L(\mu_n, \mu) \to 0$ iff $v(\mu_n, \mu) \to 0$ for sequences $\{\mu_n\}_{n=1}^\infty \subset \mathscr{M}_1$ and $\mu \in \mathscr{M}_1$. Since $v(\mu_n, \mu) \to 0$ always implies $L(\mu_n, \mu) \to 0$ we need only show the reverse implication.

Let $F = \sup \{f(x)\colon x \in S, f \text{ is a density of some } \mu \in \mathscr{M}_1\}$. We first show that $F < \infty$. Define $S(x, \delta) = \{y\colon \rho(x, y) < \delta\}$ and choose a δ such that $c = \int_{S(0,\delta)} dx < \infty$ (where dx is Haar measure and 0 is the identity of the group). Then if $\mu \in \mathscr{M}_1$

$$1 \geqq \mu(S(x, \delta)) = \int_{S(x,\delta)} f(y)\, dy \geqq (f(x) - k\delta) \int_{S(x,\delta)} dy = (f(x) - k\delta)c$$

so that $f(x) \leqq k\delta + 1/c$ and $F \leqq k\delta + 1/c$.

Let f and g be densities of measures in $\mathscr{M}_1$. Suppose $f(x) > g(x) + \varepsilon$ for some x. Choose δ and γ such that

(i) $0 < \delta < \gamma$,

(ii) $\delta < \varepsilon/8k$,

(iii) $\int_{S(0,\delta)} dx < \infty$,

(iv) $\int_{S(0,\gamma) - S(0,\delta)} dx < (\varepsilon/4F) \int_{S(0,\delta)} dx$.

By the translation invariance of Haar measure, (i), (ii), (iii), (iv), and the Lipschitz condition we see that if μ has density f and λ has density g,

$$\begin{aligned}
&\mu(S(x,\delta))-\lambda(S(x,\gamma)) \\
&\quad = \int_{S(x,\delta)}(f(y)-g(y))\,dy-\int_{S(x,\delta)-S(x,\gamma)}g(y)\,dy \\
&\quad \geqq \int_{S(x,\delta)}(f(x)-\delta k-(g(x)+\delta k))\,dy-F\int_{S(x,\delta)-S(x,\gamma)}dy \\
&\quad \geqq (\varepsilon-2\delta k)\int_{S(0,\delta)}dy-F\int_{S(0,\gamma)-S(0,\delta)}dy \\
&\quad \geqq (\varepsilon/2)\int_{S(0,\delta)}dy > 0.
\end{aligned}$$

Thus $L(\mu, \lambda) \geqq \min(\gamma-\delta, (\varepsilon/2)\int_{S(0,\delta)}dy) > 0$ for any pair $\mu, \lambda \in \mathscr{M}_1$ whose densities differ by more than ε, so that $L(\mu_n, \mu) \to 0$ implies that the densities converge everywhere which in turn implies $v(\mu_n, \mu) \to 0$ by Scheffé's Theorem (Parthasarathy [10], page 206). □

COROLLARY 3.8. *Under* II *if S is R^n and the measures in $\mathscr{M}_1$ have densities with first partial derivatives which are uniformly bounded then $\mathscr{L} = \mathscr{V}$.*

It should be noted that under II the set of all Borel probability measures on S is a complete separable metric space under the weak topology and thus if $\mathscr{M}_1$ is a subset such that $\mathscr{L} = \mathscr{V}$, then $\mathscr{T}$ is separable.

The following is useful in proving the metrizability of in specific cases.

PROPOSITION 3.9. *Under* II, *let $\mathscr{M}_1 = \bigcup_{n=1}^{\infty}\mathscr{N}_n$. Let $\mathscr{L}$ restricted to η_n be $\mathscr{L}_n$, similarly for $\mathscr{V}$ and $\mathscr{V}_n$. If $\mathscr{L}_n = \mathscr{V}_n$ for each n and if $\{\mu_n\}_{n=1}^{\infty} \cup \{\mu\} \subseteq \mathscr{M}_1$ and $L(\mu_n, \mu) \to 0$ implies there exists an m and N such that for all $n \geqq N$, $\mu_n \in \eta_m$ and $\mu \in \eta_m$; then $\mathscr{L} = \mathscr{V}$.*

PROOF. Let $L(\mu_n, \mu) \to 0$ with $\mu_n; n \geqq N$ and μ in η_m where $\mathscr{L}_m = \mathscr{V}_m$ then $v(\mu_n, \mu) \to 0$. □

4. Examples for which $\mathscr{T}$ is metrizable. The preceding results describe special cases for which $\mathscr{T}$ is metrizable. It is of interest to exhibit some possible "metrics for distinguishability" in other cases involving common families of probability measures.

4.1. *Stable distributions.* Suppose $\mathscr{M}_1$ consists of all stable distributions on the real line. Then the distributions in $\mathscr{M}_1$ have characteristic functions Ψ which can be parametrized by four parameters $(\alpha, \beta, \gamma, \sigma^2)$ with parameter space

$$Q \equiv \{(\alpha,\beta,\gamma,\sigma^2): 0 < \alpha \leqq 2, |\beta| \leqq 2-\alpha, -\infty < \gamma < \infty, \sigma^2 > 0\}$$

and with $\log \Psi$ written for $0 < \alpha < 1$ and $1 < \alpha < 2$,

$$\log\Psi(t) = it\left(\gamma+\frac{\sigma^2\beta\alpha}{\pi(1-\alpha)}\right)-\frac{\sigma^2}{2}|t|^{\alpha}\left\{1+i\frac{t}{|t|}\frac{\beta}{(2-\alpha)}\tan\frac{\pi}{2}\alpha\right\};$$

$$\log\Psi(t) = it\gamma-\frac{\sigma^2}{2}|t|\left\{1+i\frac{t}{|t|}\frac{2\beta}{\pi}\log|t|\right\} \quad \text{for} \quad \alpha = 1; \qquad \text{and}$$

$$\log \Psi(t) = it\gamma - \frac{\sigma^2}{2} t^2 \quad \text{for} \quad \alpha = 2.$$

This parametrization has the useful property that the weak topology on $\mathscr{M}_1$ is equivalent to the Euclidean topology $\mathscr{R}$ on Q.

Let r correspond to the ordinary Euclidean metric on R_4, i.e.

$$r(\mu_1, \mu_2) = [(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2 + (\gamma_1 - \gamma_2)^2 + (\sigma_1{}^2 - \sigma_2{}^2)^2]^{\frac{1}{2}}$$

where μ_j has parameters $(\alpha_j, \beta_j, \gamma_j, \sigma_j{}^2)$.

THEOREM 4.1. *Let* I *hold with* $\mathscr{M}_1$ *a collection of stable distributions. Then* $\mathscr{L} = \mathscr{R} = \mathscr{T}_1 = \mathscr{V}$.

PROOF. Since $\mathscr{L} = \mathscr{R}$, all that need be shown is that $r(\mu_n, \mu) \to 0$ implies $v(\mu_n, \mu) \to 0$. If $r(\mu_n, \mu) \to 0$ then $\Psi_n(t) \to \Psi(t)$ uniformly on any finite interval. If F_n' is the density function of μ_n then

$$F_n'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Psi_n(t)\, dt$$
$$= \frac{1}{2\pi} \int_{-k}^{k} e^{-itx} \Psi_n(t)\, dt + \frac{1}{2\pi} \int_{|t|>k} e^{-itx} \Psi_n(t)\, dt.$$

However

$$\left|\int_{|t|>k} e^{-itx} \Psi_n(t)\, dt\right| \leqq \int_{|t|>k} |\Psi_n(t)|\, dt \leqq 2\int_{t>k} \exp(-\sigma^2 |t|^\alpha/2)\, dt.$$

There is an $\varepsilon > 0$ and an integer $N(\varepsilon)$ such that for all $m > N(\varepsilon)$, $\alpha_m > \alpha - \varepsilon > \varepsilon$ and $\sigma_n{}^2/2 > \sigma^2/2 - \varepsilon > 0$. Thus for $m > N(\varepsilon)$ the above is

$$\leqq 2\int_{t>k} \exp[-(\sigma^2/2 - \varepsilon) t^{\alpha - \varepsilon}]\, dt \to 0 \quad \text{as} \quad k \to \infty.$$

Hence for any $\varepsilon > 0$ there is a number $M(\varepsilon)$ such that for $m > M(\varepsilon)$

$$|F_m'(x) - F'(x)| < \varepsilon.$$

By Scheffé's theorem [10]

$$\int_{-\infty}^{+\infty} |F_n'(x) - F'(x)|\, dx \to 0 \quad \text{as} \quad n \to \infty,$$

which implies $v(\mu_n, \mu) \to 0$ as $n \to \infty$. □

4.2. *Exponential families.* For a background on exponential families refer to Ferguson [1] and Lehmann [9]. We will consider exponential families with the natural parametrization and in as low a dimensional parameter set as possible. That is, we will work with the following assumption.

ASSUMPTION III. Let $\mathscr{C} = \{\mu_\theta\}$ be a family of probability measures on the Borel sigma field in Euclidean l-space, R^l. Let each probability measure have a density with respect to a fixed sigma-finite measure λ. The densities are indexed by $\boldsymbol{\theta}$, a k-dimensional real vector where the densities are

$$f(x \mid \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left(\textstyle\sum_{i=1}^k \theta_i t_i(x)\right).$$

Let $\Theta = \{\boldsymbol{\theta}: \int \exp[\sum_{i=1}^{k} \theta_i t_i(x)]\lambda(dx) < \infty\}$. Θ is necessarily a convex subset of R^k. We assume that $\boldsymbol{\theta} \to \mu_\theta$ is a one-to-one map from Θ onto $\mathscr{C}$ and that Θ has non-empty interior. We also assume that each $t_i(x)$ is continuous, except possibly on a set of λ measure zero.

Let $r''(\mu_{\theta_1}, \mu_{\theta_2})$ be the Euclidean distance between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ (for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \in \Theta$). Let $\mathscr{R}''$ be the topology induced by r'' on $\mathscr{C}$.

THEOREM 4.2. *Under Assumption* III, $\mathscr{L} = \mathscr{R}'' = \mathscr{V} = \mathscr{T}_1$.

PROOF. Suppose that $L(\mu_{\theta_n}, \mu_\theta) \to 0$, we will show that $r''(\mu_{\theta_n}, \mu_\theta) \to 0$. Suppose, to the contrary, that $\boldsymbol{\theta}_n$ does not converge to $\boldsymbol{\theta}$. No subsequence can converge to a point $\boldsymbol{\theta}' \in \Theta$ where $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$. (Since this would imply that $L(\mu_{\theta'}, \mu_\theta) = 0$, which would imply that $\mu_\theta = \mu_{\theta'}$, contradicting Assumption III.) By taking subsequences of subsequences we may assume that $\theta_i^{(n)}$ is monotone for each $i = 1, \cdots, k$. In fact we may assume that each $\theta_i^{(n)}$ is monotone non-decreasing (by redefining t_i as $-t_i$ if necessary.)

Define $f_i(x) = e^{t_i(x)}$ and $\pi(h, \theta) = \int_{-\infty}^{+\infty} h(x) \prod_{i=1}^{k} f_i^{\theta_i}(x)\lambda(dx)$, where $h(x)$ is any continuous, bounded function. Let $\boldsymbol{\theta}' = (\lim_n \theta_1^{(n)}, \cdots, \lim_n \theta_k^{(n)})$. We will divide the proof into two separate cases.

(a) Suppose that $\boldsymbol{\theta}'$ has finite entries. Since $\boldsymbol{\theta}' \notin \Theta$, $\pi(1, \boldsymbol{\theta}') = +\infty$. By Fatou's lemma, $\pi(1, \boldsymbol{\theta}_n) \to +\infty$ as $n \to \infty$.

Choose x a continuity point of all the f_i. Then choose $\mathscr{O}$ an open set containing x such that all the f_i are bounded in $\mathscr{O}$ then

$$\mu_{\theta_n}(\mathscr{O}) = \frac{\int_{\mathscr{O}} \prod_{i=1}^{k} f_i^{\theta_i(n)}(x)\lambda(dx)}{\pi(1, \boldsymbol{\theta}_n)} \to 0 \quad \text{contradicting}$$

$\lim_n \inf \mu_{\theta_n}(\mathscr{O}) \geqq \mu_\theta(\mathscr{O})$ which is required for weak convergence.

(b) Suppose that $\theta_i' = +\infty$ for one or more i. We may add or subtract constants to the θ_i's (by redefining λ). Thus without loss of generality we may assume that Θ contains the origin as an interior point.

By taking subsequences of subsequences and possibly relabeling the index i we may assume without loss of generality that $\theta_1^{(n)} \geqq \theta_2^{(n)} \geqq \cdots \geqq \theta_k^{(n)}$ for all n and that $\theta_i^{(n)}/\theta_1^{(n)} \to l_i$, $0 \leqq l_i \leqq 1$, $i = 1, 2, \cdots, k$. Since Θ contains the origin we may choose $\boldsymbol{\theta}^* \in \Theta$ such that $\theta_i^*/\theta_1^* = l_i$. We may find points x_0 and x_1 in the support of λ and also continuity points of all the f_i such that

$$\prod_{i=1}^{k} f_i^{\theta_i^*}(x_0) > \prod_{i=1}^{k} f_i^{\theta_i^*}(x_1).$$

(If this were not so, then both μ_{θ^*} and μ_0' would be proportional to λ and hence equal, since both are probability measures.)

Now we note that

$$\left(\frac{\prod_{i=1}^{k} f_i^{\theta_i^{(n)}}(x_0)}{\prod_{i=1}^{k} f_i^{\theta_i^{(n)}}(x_1)}\right)^{1/\theta_1^{(n)}} \to \left(\frac{\prod_{i=1}^{k} f_i^{\theta_i^*}(x_0)}{\prod_{i=1}^{k} f_i^{\theta_i^*}(x_1)}\right)^{1/\theta_1^*} > 1.$$

Thus, $[\prod_{i=1}^{k} f_i^{\theta_i^{(n)}}(x_0')]/[\prod_{i=1}^{k} f_i^{\theta_i^{(n)}}(x_1')] \to +\infty$ uniformly for x_0' and x_1' in some neighborhoods N_0 of x_0 and N_1 of x_1 of finite λ measure. Taking a function g_i nonzero at x_i and zero outside N_i it follows that

$$\frac{\pi(g_0, \boldsymbol{\theta}_n)}{\pi(1, \boldsymbol{\theta}_n)} \to \frac{\pi(g_0, \boldsymbol{\theta})}{\pi(1, \boldsymbol{\theta})},$$

implying $\pi(g_1, \boldsymbol{\theta}_n)/\pi(1, \boldsymbol{\theta}_n) \to 0$, which contradicts $\pi(g_1, \boldsymbol{\theta})/\pi(1, \boldsymbol{\theta}) > 0$. Thus $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$.

But $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ implies $f(x \mid \boldsymbol{\theta}_n') \to f(x \mid \boldsymbol{\theta})$ for all x which implies $v(\mu_{\theta_n}, \mu_\theta) \to 0$ by Scheffé's Theorem [10].

The proof is complete since $v(\mu_{\theta_n}, \mu_\theta) \to 0$ always implies $L(\mu_{\theta_n}, \mu_\theta] \to 0$.

5. Acknowledgment. The authors wish to thank the referee for several corrections and suggestions and for pointing out the existence of the parametrization used in Section 4.1.

REFERENCES

[1] Ferguson, Thomas S. (1967). *Mathematical Statistics.* Academic Press, New York.

[2] Fisher, Lloyd and Van Ness, John W. (1969). Distinguishability of probability measures. *Ann. Math. Statist.* **40** 381–392.

[3] Fisher, Lloyd and Walkup, D. W. (1969). An example of the difference between the Lévy and Levy–Prokhorov metrics. *Ann. Math. Statist.* **40** 322–324.

[4] Freedman, D. A. (1967). A remark on sequential discrimination. *Ann. Math. Statist.* **38** 1666–1670.

[5] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison–Wesley, Reading, Mass.

[6] Hoeffding, W. and Wolfowitz, J. (1958). Distinguishability of set of distributions. *Ann. Math. Statist.* **29** 700–718.

[7] Kelley, John L. (1955). *General Topology.* Van Nostrand, Princeton, N.J.

[8] LeCam, L. and Schwartz, L. (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* **31** 140–150.

[9] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[10] Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

[11] Parzen, Emanuel (1960). *Modern Probability Theory and Its Applications.* Wiley, New York.