*The Annals of Mathematical Statistics*
1969, Vol. 40, No. 2, 381–392

# DISTINGUISHABILITY OF PROBABILITY MEASURES[1]

BY LLOYD FISHER AND JOHN W. VAN NESS

*University of Washington*

**0. Summary.** Independent identically distributed observations, $X_1, X_2, \cdots$, are taken sequentially. All that is known a priori about their common probability measure, $P$, is that it is a member of a given (at most countable) family, $\pi = \{P_n\}_{n=1}^{\infty}$, of such measures. At some time, depending only on the observed data and the tolerable probability of error, one wants to stop and decide which $P_k$ nature has chosen.

Two sampling situations are considered, with and without error, as well as two stopping time requirements, uniformly (over $\pi$) bounded and $P_k$-dependent.

Necessary and/or sufficient conditions for the distinguishability of the measures in $\pi$ in terms of a variety of measure metrics are obtained. The Lévy-Prokhorov metric proves to be particularly relevant.

**1. Distinguishability without observational error.** Let $S$ be a complete separable metric space with metric $\rho(\cdot, \cdot)$ and let $\mathfrak{A}$ be its Borel field. Let $\pi = \{P_n\}_{n=1}^{\infty}$ be a countable family of Borel probability measures on the measurable space $(S, \mathfrak{A})$. In Sections 3 and 4 $S$ is required to be a locally compact metrizable Abelian group and $\rho$ is taken to be translation invariant. Denote by $S^{\infty}$ the product space $S \times S \times S \times \cdots$ and by $\mathfrak{A}^{\infty}$ the product sigma field on $S^{\infty}$ generated by $\mathfrak{A}$. Denote by $\mathfrak{A}^n$ the subsigma field of $\mathfrak{A}^{\infty}$ generated by events of the form $A_1 \times \cdots \times A_n \times S \times S \times \cdots$ where $A_i \varepsilon \mathfrak{A}$, $i = 1, 2, \cdots, n$. For any measure $P$ on $(S, \mathfrak{A})$, let $\mathbf{P}$ be its product measure on $(S^{\infty}, \mathfrak{A}^{\infty})$.

A sequential test $(N, d)$ consists of a stopping time $N$ and a decision function $d$. Here $N$ is a measurable function on $(S^{\infty}, \mathfrak{A}^{\infty})$ taking on positive integer values and $\infty$ and is such that the set $\{w: w \varepsilon S^{\infty}, N(w) = n\} \varepsilon \mathfrak{A}^n$. The decision function $d$ may be randomized so that in general $d$ takes on values which are probability measures on the integers; that is, for $w \varepsilon S^{\infty}$,

$$d(w) = (d^{(1)}(w), d^{(2)}(w), \cdots),$$

where $d^{(j)}(w) \geqq 0$ and $\sum_{j=1}^{\infty} d^{(j)}(w) = 1$. For each positive integer $j$ we require that $d^{(j)}(\cdot)$ is a measurable function on $(S^{\infty}, \mathfrak{A}^{\infty})$. Further, $d(w)$ depends only on the first $N(w)$ coordinates of $w$, i.e., if $w$ and $w_1$ have the same first $N(w)$ coordinates then $d(w) = d(w_1)$. We interpret $d^{(j)}(w)$ as the probability given the observation $w = (s_1, \cdots, s_{N(w)}, \cdots)$ that we choose $P_j$ as the underlying member of $\pi$. Define the *distinguishability* of the family $\pi$ to be

$$\mathfrak{D}_{\pi} = \sup_{(N,d)} \inf_{P_k \varepsilon \pi} \mathbf{P}_k\{(N, d) \quad \text{is correct and} \quad N < \infty\}$$
$$= \sup_{(N,d)} \inf_{P_k \varepsilon \pi} E_{\mathbf{P}_k}(d^{(k)} I_{\{N<\infty\}})$$

Received 6 February 1968.

[1] This work was supported by the National Science Foundation under Grant GP-7519.

where $I_{\{N<\infty\}}$ is the indicator function of the event $\{N < \infty\}$ and $E_{\mathbf{P}_k}$ is the expectation with respect to the measure $\mathbf{P}_k$. If $\mathfrak{D}_\pi = 1$ we say the family $\pi$ is *distinguishable* and write $\mathfrak{D}$ for the collection of all distinguishable families.

A *nonrandomized* decision rule $d$ is one whose range is contained in the degenerate probability measures on $\{1, 2, 3, \cdots\}$. It is sufficient to consider only such rules, as is shown by the following result which is proved in Section 5.

PROPOSITION 1.1. *We may get arbitrarily close to* $\mathfrak{D}_\pi$ *using only sequential tests with non-randomized decision rules.*

Earlier works [2] and [4] on sequential tests $(N, d)$ assume that $\mathbf{P}_k\{N < \infty\} = 1$ for each $P_k \varepsilon \pi$. This assumption is not necessary as shown by the next statement which follows easily from [2] (see Section 5).

PROPOSITION 1.2. *If* $\pi$ *is distinguishable, then for any* $\epsilon > 0$ *we may choose a sequential test* $(N_\epsilon, d_\epsilon)$ *such that for all* $P_n \varepsilon \pi$, $E_{\mathbf{P}_n}(d_\epsilon^{(n)}) \geqq 1 - \epsilon$ *and* $\mathbf{P}_n\{N_\epsilon < \infty\} = 1$.

In looking for necessary and/or sufficient conditions for the various kinds of distinguishability there are several measures of distance between probability distributions which can be used. Those used here are given below.

DEFINITION. (a) The Lévy-Prokhorov [8] metric $L(\cdot, \cdot)$ is defined as follows. For any closed set $A \subseteq S$ define the open set $A^\epsilon$ by

$$A^\epsilon = \{s: \rho(s, A) < \epsilon\}.$$

Then $L(P_1, P_2)$ is the greatest lower bound of the $\epsilon > 0$ such that $P_1\{A\} < P_2\{A^\epsilon\} + \epsilon$ for all closed $A \subseteq S$.

(b) The total variation metric

$$V(P_1, P_2) = \sup_{A\varepsilon\mathfrak{A}} |P_1\{A\} - P_2\{A\}|.$$

(c) If $S = R_n$ ($n$-dimensional Euclidean space), then we define the distance

$$D(P_1, P_2) = \sup_{x\varepsilon R_n} |F_1(x) - F_2(x)|$$

where $F_j(\cdot)$ is the distribution function of the measure $P_j$.

(d) Finally we denote the ordinary Lévy metric for Borel probability measures on $R_1$ by $L'$. The distance $L'(P, Q)$ between two probability measures $P$ and $Q$ is defined as the infimum of all $k$ such that for all $x$

$$P\{(-\infty, x - k]\} - k \leqq Q\{(-\infty, x]\} \leqq P\{(-\infty, x + k]\} + k.$$

The Lévy-Prokhorov metric $L$ is more natural for our purposes than $L'$ (see Section 4). The distinction is illustrated by the following example.

EXAMPLE 1. For $n = 1, 2, \cdots$, let $P_{2n+1}$ put mass $n^{-1}$ at the points $n^2 + 1$, $n^2 + 3, \cdots, n^2 + 2n - 1$ and $P_{2n}$ put mass $n^{-1}$ at the points $n^2 + 2, n^2 + 4, \cdots, n^2 + 2n$. Then $L(P_{2n+1}, P_{2n}) = 1$ whereas $L'(P_{2n+1}, P_{2n}) = 2^{\frac{1}{2}}/n$. Thus the family $\pi = \{P_k\}$ is uniformly separated in $L$ but not in $L'$.

We wish to compare $\mathfrak{D}$ and several other collections of families $\pi$. Let $K$ be the set of all the subscripts $k$ of the measures $P_k \varepsilon \pi$.

DEFINITION. (a) $\pi \varepsilon \mathfrak{D}_1$ if for any $\epsilon > 0$ and any $n \varepsilon K$ there exists an integer

$m(n, \epsilon)$ and a set $A(n, m, \epsilon) \in \mathcal{A}^m$ such that $\mathbf{P}_n\{A\} \geqq 1 - \epsilon$ and $\mathbf{P}_k\{A\} \leqq \epsilon$ for all $k \in K - \{n\}$.

(b) $\pi \in \mathfrak{D}_2$ if for any $n \in K$ there exists an $\epsilon(n) > 0$, an integer $m(n)$, and a collection of sets $A_1^{(n)}, \cdots, A_{m(n)}^{(n)} \in \mathcal{A}$ such that

$$\max_{j=1,\cdots,m(n)} |P_k\{A_j^{(n)}\} - P_n\{A_j^{(n)}\}| > \epsilon(n), \qquad k \in K - \{n\}.$$

(c) $\pi \in \mathfrak{V}$ if for all $n \in K$, $\inf_{k\in K-\{n\}} V(P_k, P_n) > 0$.

(d) $\pi \in \mathfrak{L}$ if for all $n \in K$, $\inf_{k\in K-\{n\}} L(P_k, P_n) > 0$.

(e) $\pi \in \mathfrak{F}(S \subset R_n)$ if for all $n \in K$, $\inf_{k\in K-\{n\}} D(P_k, P_n) > 0$.

The following diagram summarizes the relationships between these collections for the case of observations without error.

Of course the relationships involving $\mathfrak{F}$ only make sense if $S \subset R_n$.

$\mathfrak{D}_1 = \mathfrak{D}$ and $\mathfrak{D} \supset \mathfrak{D}_2$ are proved in Freedman [2]. The fact that $\mathfrak{V} \subset \mathfrak{D}$ and

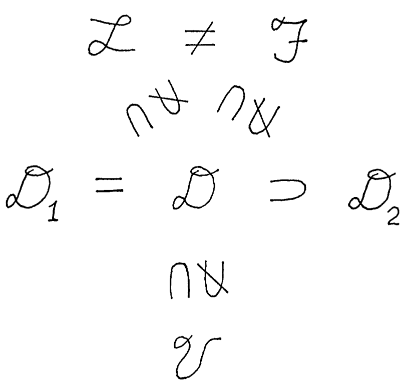

FIG. 1

$\mathfrak{F} \subset \mathfrak{D}$ follows from Hoeffding and Wolfowitz [4], Sections 3 and 4, and the fact that $\pi \in \mathfrak{D}$ iff for all $n \in K$, the pairs of families $\pi_n = \{P_n\}$ and $\pi - \pi_n$ are distinguishable in the following sense.

DEFINITION. We say a probability measure $P$ is distinguishable from a family $\pi$ if $\forall \epsilon > 0$ there is an $n(\epsilon) < \infty$ and an $A \in \mathcal{A}^{n(\epsilon)}$ such that $\mathbf{P}\{A\} \geqq 1 - \epsilon$ and $\mathbf{Q}\{A\} \leqq \epsilon$ for all $Q \in \pi$.

Kraft [6], p. 132, gives an example which shows that $\mathfrak{V} \not\subset \mathfrak{D}$. The following example shows that $\mathfrak{D} \not\subset \mathfrak{F}$ and $\mathfrak{D} \not\subset \mathfrak{L}$.

EXAMPLE 2. Let $K = \{0, 1, 2, \cdots\}$ and $P_0$ be the uniform distribution on [0, 1]. For each $n > 0$, partition [0, 1] into $n$ half open intervals of equal length and select one point from each interval. Define $P_n$ to be the uniform distribution on the $n$ points selected. If the points are chosen so that no two are the same, then $P_1, P_2, \cdots$ have disjoint supports and the family is distinguishable with probability one in one observation.

Finally in Section 5 we prove the following.

PROPOSITION 1.3. *$\pi \in \mathfrak{L}$ implies that $\pi \in \mathfrak{D}$.*

If we restrict ourselves to $\pi$'s which are classes of *discrete* probabilities then all the classes mentioned are equivalent except $\mathfrak{L}$.

PROPOSITION 1.4. *If for all $k \varepsilon K$, $P_k$ is a discrete probability measure on $R_n$ then $\pi \varepsilon \mathfrak{V}$ iff $\pi \varepsilon \mathfrak{F}$.*

EXAMPLE 3. Let $P_0$ put unit mass at 0 and $P_n$ put mass $\frac{1}{2}$ on the points 0 and $n^{-1}$. Then $\pi = \{P_n, n = 0, 1, 2, \cdots\} \varepsilon \mathfrak{D}$ is a collection of discrete probabilities but $\pi \varepsilon \mathfrak{D}$ and $\pi \not\varepsilon \mathfrak{L}$.

It is interesting to look at the case $\mathfrak{D}_\pi \neq 1$. We still may get a partial degree of distinguishability. In fact, if given $\pi$ we define $\pi_0 = \pi$ and $\pi_n$ for $n \geqq 1$ inductively by $P \varepsilon \pi_{n+1}$ iff $P \varepsilon \pi_n$ with $P$ not distinguishable from the set $\pi_n - \{P\}$, then in Section 5 we will prove

THEOREM 1.5. *For any $\pi$: Either $\mathfrak{D}_\pi = 0$, or $\mathfrak{D}_\pi = n^{-1}$ for some $n = 1, 2, \cdots$. Furthermore $\mathfrak{D}_\pi = n^{-1}$* iff *$\pi_n$ is empty but $\pi_{n-1}$ is not.*

The collection $\pi_n$ can be thought of as all the "limit points" of $\pi_{n-1}$.

PROPOSITION 1.6. *If $\mathfrak{D}_\pi = n^{-1}$, $n > 1$ and $(N, d)$ is such that $\forall k \varepsilon K$, $\mathbf{P}_k\{(N, d)$ is correct, $N < \infty\} > \epsilon > 0$, then for at least one $j \varepsilon K$, $\mathbf{P}_j\{N = \infty\} > 0$.*

## 2. Bounded distinguishability without observational error.

DEFINITION. A sequential test $(N, d)$ is said to be bounded (relative to $\pi$) if there exists a constant $b$ such that $\mathbf{P}_k\{N \leqq b\} = 1$ for all $P_k \varepsilon \pi$. A family $\pi$ is said to be finitely distinguishable, written $\pi \varepsilon \mathfrak{B}$, if for any $\epsilon > 0$, there exists a $b(\epsilon)$ and a test $(N, d)$ with $N$ bounded by $b$ such that

$$\inf_{P_k \varepsilon \pi} \mathbf{P}_k\{(N, d) \text{ is correct}\} \geqq 1 - \epsilon.$$

The following theorem gives a sufficient condition for finite distinguishability. Whether or not this condition is necessary remains an open problem.

THEOREM 2.1. *If there exists a disjoint sequence, $\{B_i\}_{i=1}^{\infty}$ of subsets of $S$ and an $\epsilon > 0$ such that $\forall n \varepsilon K$*

$$\inf_{k \varepsilon K - \{n\}} \sup_{1 \leqq m < \infty} |P_n\{B_m\} - P_k\{B_m\}| > \epsilon$$

*then $\pi$ is finitely distinguishable.*

PROOF. Let $p_j = P_i\{B_j\}$. After $n$ observations define $q_j = n^{-1} \cdot$ (number of observations in $B_j$). Kiefer and Wolfowitz [5] prove that in $R_k$ any sample distribution function $F_n$, constructed from $n$ samples from a distribution function $F$ satisfies

$$\mathbf{P}\{\sup_x |F_n(x) - F(x)| \geqq c\} \leqq ae^{-bc^2 n} \tag{2.1}$$

for some numbers $a$ and $b$ (independent of $F$) and for any $c > 0$. If we let $c = \epsilon/2$ then (2.1) says that there is an $n(\epsilon, \delta)$ such that the decision rule, $d$, which chooses that $k$ such that

$$\min_{1 \leqq j \leqq \infty} |p_j - q_j| = |p_k - q_k|$$

is correct with probability greater than $1 - \delta$. Q.E.D.

**3. Distinguishability with observational error.** In this section we assume that we cannot observe the actual outcomes $s_1, s_2, \cdots$ being sequentially generated according to the unknown member of the countable family, $\pi$. Instead we experience observational errors causing us to observe $s_1 + e_1, s_2 + e_2, \cdots$ where the $e_j$'s are outcomes of independent identically distributed random experiments which are also independent of all the experiments generating $s_1, s_2, \cdots$.

Let $Q$ defined on $(S, \mathfrak{A})$ be the probability measure of the error. Then the problem discussed here is equivalent to the afore-mentioned distinguishability of $\pi * Q = \{P_k * Q; k \varepsilon K\}$.

It is clear that if $P_1 \neq P_2$ but $P_1 * Q = P_2 * Q$ then we cannot distinguish $\pi * Q$ even for $\pi = \{P_1, P_2\}$. To avoid this make the

ASSUMPTION. $Q$ is (i) absolutely continuous with respect to Haar measure, and such that (ii) $\mathcal{P} \rightarrow \mathcal{P} * Q$ is a 1-1 map where $\mathcal{P}$ is the family of all Borel probability measures.

EXAMPLE 4. Let $S = R_1$, then $\mathcal{P} \rightarrow \mathcal{P} * Q$ is 1-1 iff $Q$'s characteristic function, $\varphi(t)$, is not zero for all $t$ in any nonempty interval $(a, b)$. If $\mathcal{P} \rightarrow \mathcal{P} * Q$ is 1-1, so is $\mathcal{P} \rightarrow \mathcal{P} * Q * (Q^-)$ where $Q^-$ is the reflection of $Q$ about zero. The characteristic function of $Q * Q^-$, $\psi(t) = |\varphi(t)|^2$ is zero in $(a, b) \cup (-b, -a)$ if $\varphi(t)$ is zero in $(a, b)$. Any positive even function, $g$, concave on $(0, \infty)$, with $g(0) = 1$ is a characteristic function. Thus we can find two characteristic functions, $\varphi_1$ and $\varphi_2$, which differ only on $(a, b) \cup (-b, -a)$. Thus $\varphi$ cannot be zero on any interval. If $\varphi$ is not zero or any interval then $\varphi_{P*Q} = \varphi_P \varphi_Q$ and we can find $\varphi_P$ on a dense set by dividing by $\varphi_Q$. This determines $\varphi_P$ since $\varphi_P$ is continuous.

THEOREM 3.1. $\pi * Q \varepsilon \mathfrak{D}$ iff $\pi \varepsilon \mathfrak{L}$.

LEMMA 3.2. *If* $L(P_n * Q, P * Q) \rightarrow 0$ *as* $n \rightarrow \infty$ *then* $L(P_n, P) \rightarrow 0$ *as* $n \rightarrow \infty$.

PROOF. The hypothesis implies that $\pi * Q = \{P_n * Q; n = 1, 2, \cdots\} \cup \{P * Q\}$ is a tight family. By the shift-compactness theorem of Parthasarathy, Ranga Rao and Varadhan [7], $\{P_n\}$ is tight. Choose a convergent subsequence $\{P_{n_k}\}$ such that $L(P_{n_k}, P') \rightarrow 0$ for some $P'$. Then $L(P_n * Q, P' * Q) \rightarrow 0$ so that $L(P * Q, P' * Q) = 0$. But $Q$ gives a 1-1 mapping and $P = P'$. Q.E.D.

PROOF OF THEOREM 3.1. By the lemma, if $\pi \varepsilon \mathfrak{L}$ then $\pi * Q \varepsilon \mathfrak{L}$ implying $\pi * Q \varepsilon \mathfrak{D}$ by Proposition 1.3.

We will prove the reverse by showing that $L(P_n, P) \rightarrow 0$ implies $V(P_n * Q, P * Q) \rightarrow 0$.

Let $Q\{A\} = \int_A f(x)\, dx$, then

$$V(P_n * Q, P * Q) = \int_S \left|\int_S f(x - y)(P_n(dy) - P(dy))\right| dx.$$

Since continuous functions with compact support are dense in $L_1(S)$ (Hewitt and Ross [3], p. 140) we may by Fubini's theorem assume $f$ is continuous and has compact support, $J$. We may also pick a compact set, $C$, such that $P_n\{C\} \geqq 1 - \epsilon$ for all $n \varepsilon K$, and $C$ is a continuity set for $P$. Then for every $x \varepsilon S$

$$\int_C f(x - y) P_n(dy) \rightarrow \int_C f(x - y) P(dy).$$

Furthermore

$$\int_S \left|\int_S f(x-y)(P_n(dy) - P(dy))\right| dx$$
$$\leqq \int_S \left|\int_C f(x-y)(P_n(dy) - P(dy))\right| dx + \int_S \left|\int_{C'} f(x-y)(P_n(dy) - P(dy))\right| dx$$
$$\leqq \int_{C+J} \left|\int_C f(x-y)(P_n(dy) - P(dy))\right| dx + \epsilon$$

and the first integral approaches zero by dominated convergence. Thus if $\pi \not\in \mathfrak{L}$, $\pi * Q \not\in \mathfrak{V}$ and hence $\pi * Q \not\in \mathfrak{D}$ by Section 1. Q.E.D.

Recall that in Section 1 we mentioned an example of Kraft which shows that $\pi \in \mathfrak{V}$ does not imply that $\pi \in \mathfrak{D}$. This example depends on rapidly oscillating density functions approaching the uniform density on any fixed measurable set. If we convolute with absolutely continuous error this damps out such oscillation so that $\pi * Q \in \mathfrak{V}$ implies $\pi * Q \in \mathfrak{L}$. In Hoeffding and Wolfowitz [4], p. 713, restrictions on a function they call $J$ serves to eliminate the same type of osculatory behavior.

**4. Bounded distinguishability with observational error.** The Lévy-Prokhorov metric proves natural for the study of $\pi * Q \in \mathfrak{B}$.

Definition. (a) For any measure, $P$, and any $x \in S$ we define $P^x$ to be the translation of $P$ by $x$:

$$P^x\{A\} = P\{A - x\}.$$

(b) We call a family $\pi$ shift-compact if there exists a mapping $g: \pi \to S$ such that

$$\pi_g = \{P^{g(P)}: P \in \pi\} \quad \text{is tight.}$$

(c) $\pi$ is uniformly $L$-isolated, written $\pi \in \mathfrak{L}_u$, if there exists an $\epsilon > 0$ such that for all $P \in \pi$

$$\inf_{P' \in \pi, P \neq P'} L(P', P) > \epsilon.$$

Theorem 4.1. (a) $\pi * Q \in \mathfrak{B}$ *implies* $\pi \in \mathfrak{L}_u$.

(b) *If* $\pi$ *is shift-compact,* $\pi * Q \in \mathfrak{B}$ iff $\pi \in \mathfrak{L}_u$.

The proof is below. If $S = R_1$, $\pi \in \mathfrak{L}_u$ is not the same as $\pi$ being uniformly isolated in the ordinary Lévy metric as is shown by Example 1.

Example 5. Let $S = R_1$, then $\pi$ is shift compact if the first absolute moments or the variances exist and are uniformly bounded.

Corollary. *Let* $\pi = \{P_k, k = 1, 2, \cdots\}$ *where* $P_k$ *is* $N(m_k, 1)$. (a) $\pi \in \mathfrak{D}$ iff *no limit points of* $\{m_n\}$ *are in* $\{m_n\}$. (b) $\pi \in \mathfrak{B}$ iff *there is an* $\epsilon > 0$ *such that for all* $n \in K$, $\inf_{k \neq n} |m_k - m_n| > \epsilon$.

Proof. Let $\pi' = \{Q_k\}$ where $Q_k$ puts unit mass at $m_k$. Let $Q$ be $N(0, 1)$, then $\pi = \pi' * Q$ and (a) follows from Theorem 3.1 and the fact that $L(Q_n, Q_k) = \min[1, |m_n - m_k|]$. Similarly (b) follows from Theorem 4.1 since $\pi'$ is clearly shift-compact.

We break up the proof of Theorem 4.1 into several lemmas. Let $\mathfrak{M}_n$ consist of measures which put mass $1/n$ on $n$ points $x_1, \cdots, x_n$ (some points may be identical, but we distinguish the different copies). The next lemma follows from the results of Strassen [9] and Dudley [1].

LEMMA 4.2. *Let $P$ and $P_0 \varepsilon \mathfrak{M}_n$ be such that $L(P, P_0) < \delta$, $P$ concentrates its mass on $x_1, \cdots, x_n$ and $P_0$ concentrates its mass on $y_1, \cdots, y_n$. Then there exist distinct $i_1, \cdots, i_k$ and distinct $j_1, \cdots, j_k$ such that $\rho(x_{i_m}, y_{j_m}) \leqq \delta$, $m = 1, 2, \cdots, k$ where $(n - k)n^{-1} \leqq \delta$.*

LEMMA 4.3. *If $f$ is the density of $Q$, then*

$$V(P * Q, P_0 * Q) \leqq \sup_{\{x:\rho(0,x) \leqq L(P,P_0)\}} \int |f(y + x) - f(y)|\, dy + 2L(P, P_0).$$

PROOF. (a) If $P$ and $P_0 \varepsilon \mathfrak{M}_n$, choose $k$ as in Lemma 4.2, then

$$\begin{aligned} V(P * Q, P_0 * Q) & \\ &= \int_S |\int_S f(x - y)(P(dy) - P_0(dy))|\, dx \\ &= n^{-1} \int_S |\textstyle\sum_1^n f(x - x_i) - f(x - y_i)|\, dx \\ &\leqq n^{-1} \textstyle\sum_{i=1}^k \int_S |f(x - x_i + y_i) - f(x)|\, dx \\ &\quad + n^{-1} \textstyle\sum_{i=k+1}^n \int_S (|f(x - x_i)| + |f(x - y_i)|)\, dx \\ &\leqq k/n \sup_{\{y:\rho(y,0) \leqq L(P,P_0)\}} \int_S |f(x - y) - f(x)|\, dx + 2(n - k)n^{-1}. \end{aligned}$$

(b) If $P^{(n)}$ is the sample measure of $P$ after $n$ observations then $P^{(n)} \varepsilon \mathfrak{M}_n$ and $\mathbf{P}\{L(P^{(n)}, P) \to 0\} = 1$. Thus choose a fixed sequence $\{P^{(n)}\}$ with $P^{(n)} \varepsilon \mathfrak{M}_n$ and $L(P^{(n)}, P) \to 0$ as $n \to \infty$. By the proof of Theorem 3.1 this implies $V(P^{(n)} * Q, P * Q) \to 0$. Similarly choose $\{P_0^{(n)}\}$ such that $P_0^{(n)} \varepsilon \mathfrak{M}_n$ and $L(P_p^{(n)}, P_0) \to 0$. Then for any $\epsilon_1 > 0$ and $n > n_0(\epsilon_1)$,

$$\begin{aligned} L(P^{(n)}, P_0^{(n)}) &\leqq L(P^{(n)}, P) + L(P, P_0) + L(P_0^{(n)}, P_0) \\ &\leqq L(P, P_0) + \epsilon_1 \end{aligned}$$

and

$$\begin{aligned} V(P * Q, P_0 * Q) &\leqq V(P * Q, P^{(n)} * Q) + V(P^{(n)} * Q, P_0^{(n)} * Q) \\ &\quad + V(P_0^{(n)} * Q, P_0 * Q) \\ &\leqq V(P^{(n)} * Q, P_0^{(n)} * Q) + \epsilon_1 . \end{aligned}$$

By part (a) we are done. Q.E.D.

LEMMA 4.4. *If $\pi$ is shift-compact, then for each $\epsilon > 0$ we may find an $N(\epsilon)$ such that for $n \geqq N(\epsilon)$ and for any $P \varepsilon \pi$, $\mathbf{P}\{L(P^{(n)}, P) > \epsilon\} \leqq \epsilon$.*

PROOF. Let $\pi^s$ be a shift of $\pi$ that is tight. Choose a compact $J$ such that $P \varepsilon \pi \Rightarrow P^x(J) > 1 - \epsilon/2$ where $P^x$ is the shift of $P$ which is in $\pi^s$. Cover $J$ with a finite number of disjoint sets $\{A_i : i = 1, \cdots, m\}$ of diameter $<\epsilon/2$. Define $A_{m+1} = S - \bigcup_1^m A_i$. Choose $N(\epsilon)$ independent of $P$ so large that

$$\mathbf{P}\{|P^{(n)x}\{A_i\} - P^x\{A_i\}| < \epsilon/2(m + 1) \quad \text{for} \quad i = 1, 2, \cdots, m + 1\} > 1 - \epsilon.$$

Let $G = F + x$ be a closed set and $\sum_j'$ denote $\sum_{j \ni F \cap A_j \neq \emptyset}$. If $F \cap A_j \neq \emptyset$ then $F^\epsilon \supseteq A_j$ so that for $n > N(\epsilon)$,

$$\begin{aligned} P\{G\} &\leqq P^x\{A_{m+1}\} + \textstyle\sum_{i=1}^n P^x\{F \cap A_i\} \\ &< \epsilon/2 + \textstyle\sum_j' P^x\{F \cap A_j\} \\ &\leqq \epsilon/2 + \textstyle\sum_j' P^x\{F^\epsilon \cap A_j\} \\ &\leqq \epsilon/2 + [\epsilon/2(m+1)] \text{ (no. of } j\text{'s such that } F \cap A_j \neq \emptyset) \\ &\quad + \textstyle\sum_j' P^{(n)^x}\{F^\epsilon \cap A_j\} \\ &\leqq \epsilon + P^{(n)^x}\{F^\epsilon \cap (\bigcup_j' A_j)\} \leqq \epsilon + P^{(n)}\{G^\epsilon\} \end{aligned}$$

where these inequalities hold with probability $\geqq 1 - \epsilon$. A symmetric argument shows that

$$P^{(n)}\{F\} \leqq P\{F^\epsilon\} + \epsilon. \qquad \text{Q.E.D.}$$

LEMMA 4.5. *If $\pi \varepsilon \mathfrak{L}_u$ is shift-compact, then $\pi * Q \varepsilon \mathfrak{L}_u$.*

PROOF. Suppose not, then we have $L(P_j, P_k) > \epsilon$ for all $j, k \varepsilon K$ but there is a sequence of pairs $\{P_j, P_j'\}$, $P_j$, $P_j' \varepsilon \pi$, $j = 1, 2, \cdots$, such that $L(P_j * Q, P_j' * Q) \to 0$ as $n \to \infty$. Let $x_n = x_n(P_n)$ be the shift to get $P_n$ in $\pi^s$. Then since $\rho$ is translation invariant,

$$\begin{aligned} L(P_n * Q, P_n' * Q) &= L((P_n * Q)^{x_n}, (P_n' * Q)^{x_n}) \\ &= L(P_n^{x_n} * Q, P_n'^{x_n} * Q) \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Due to the shift-compactness, we may, without loss of generality, assume that $L(P_n^{x_n}, P) \to 0$. Thus,

$$L(P * Q, P_n'^{x_n} * Q) \leqq L(P * Q, P_n^{x_n} * Q) + L(P_n^{x_n} * Q, P'^{x_n} * Q) \to 0$$

and by Lemma 3.2 $L(P, P_n'^{x_n}) \to 0$. Therefore

$$\begin{aligned} L(P_n, P_n') &\leqq L(P_n, P^{-x_n}) + L(P^{-x_n}, P_n') \\ &= L(P_n^{x_n}, P) + L(P, P_n'^{x_n}) \to 0 \end{aligned}$$

which is a contradiction. Q.E.D.

PROOF OF THEOREM 4.1. (a) If $\pi \not\varepsilon \mathfrak{L}_u$ then $\pi * Q$ is not $V$-uniformly isolated by Lemma 4.3 and the $L_1$ continuity of $f(y) \to f(y + x)$. Since there are measures arbitrarily close in $V$, we can, for any $n$, get the $n$th order product measures arbitrarily close so we cannot distinguish in a finite time.

(b) Let $\pi \varepsilon \mathfrak{L}_u$ be shift-compact, then $\pi * Q$ is shift-compact and uniformly $L$-isolated by Lemma 4.5. Using Lemma 4.4 we may use the $L$-metric to give a rule for bounded distinguishability. Q.E.D.

REMARK. Using the measures of Example 1 and letting $e$ be uniform on [0, 1] we see that $\pi * Q \varepsilon \mathfrak{B}$ does not imply that each measure of $\pi$ is uniformly isolated in the ordinary Levy metric. (Compare to Theorem 4.1 (a)).

**5. Proofs of results in Section 1.**

LEMMA 5.1. *If $\pi$ contains only a finite number of measures (all different), then for any $\epsilon > 0$ we may select a finite number $m(\epsilon)$ such that if we sample m times we may correctly identify each measure with probability greater than $1 - \epsilon$.*

PROOF. If $P$ is the underlying measure and $P^{(n)}$ the sample measure after $n$ independent observations then it is well known that $L(P^{(n)}, P) \to 0$ a.s. as $n \to \infty$.

PROOF OF PROPOSITION 1.1. Let $(N', d')$ be any test with $d'$ possibly randomized. We will construct a new nonrandomized rule, $d$. Define $A = \{w: N(w) < \infty\}$ and choose any $\epsilon > 0$. For any $w \varepsilon A$ we sample $w$ until $N(w)$ and define $\pi(w, \epsilon) = \{P_k : d^{(k)}(w) > \epsilon, k \varepsilon K\}$. By Lemma 5.1 there exists a finite number $m(w, \epsilon)$ such that we can distinguish $\pi(w, \epsilon)$ by sampling $m(w, \epsilon)$ more times. Let $d$ be such a rule, then

$$E_{\mathbf{P}_n}(d^{(n)}) \geqq E_{\mathbf{P}_n}(d^{(n)} I_{AB_n})$$

where $B_n = \{w: d'^{(n)} > \epsilon\}$. But

$$E_{\mathbf{P}_n}(d^{(n)} I_{AB_n}) \geqq (1 - \epsilon) E_{\mathbf{P}_n}(I_{AB_n}) \geqq (1 - \epsilon)(E_{\mathbf{P}_n}(d'^{(n)} - \epsilon))$$

since $A \subseteq B_n$ and $E_{\mathbf{P}_n}(d'^{(n)}) \leqq \mathbf{P}_n\{B_n\} + \epsilon$. Q.E.D.

PROOF OF PROPOSITION 1.2. If $\mathfrak{D}_\pi = 1$, then for any $\epsilon > 0$ there is a test $(N, d)$ such that $E_{\mathbf{P}_k}(d^{(k)}) > 1 - \frac{1}{2}\epsilon$ for all $k \varepsilon K$. However, since

$$1 - \tfrac{1}{2}\epsilon < E_{\mathbf{P}_k}(d^{(k)}) = \lim_{n\to\infty} E_{\mathbf{P}_k}(d^{(k)} I_{\{N<n\}})$$

there exists an $n_0(k, \epsilon)$ such that

$$E_{\mathbf{P}_k}(d^{(k)} I_{\{N<n_0\}}) > 1 - \epsilon.$$

By Proposition 1.1 we may assume $(N, d)$ is nonrandom. Therefore let $A(k, n_0, \epsilon) = \{w: N(w) \leqq n_0, d^{(k)} = 1\}$ then

$$\mathbf{P}_k\{A(k, n_0, \epsilon)\} > 1 - \epsilon$$

and for all $k \neq j \varepsilon K$,

$$\mathbf{P}_j\{A(k, n_0, \epsilon)\} = E_{\mathbf{P}_j}(d^{(k)} I\{_{N \leqq n_0}\}) \leqq E_{\mathbf{P}_j}(d^{(k)}) < \epsilon.$$

The sets $A(k, n_0, \epsilon)$ satisfy the requirements of Condition II of Freedman [2] and our definition of $\mathfrak{D}_1$. In Theorem *I* of [2] it is shown that Condition II leads to a stopping time which is finite with probability one. Q.E.D.

LEMMA 5.2. *If $L(P, Q) > \epsilon > 0$, then for any $\delta > 0$ there is an $N(\delta)$, independent of $P$ and $Q$, such that $\mathbf{P}\{L(P^{(n)}, Q) > \epsilon/2\} \geqq 1 - \delta$ for all $n > N$.*

PROOF. $L(P, Q) > \epsilon$ implies that there exists a closed set $A$ such that $P\{A\} > Q\{A^\epsilon\} + \epsilon$. By Chebyshev's inequality we may pick $N(\delta)$ independent of $P\{A\}$ such that for all $n > N$,

$$\mathbf{P}\{|P\{A\} - P^{(n)}\{A\}| < \tfrac{1}{2}\epsilon\} > 1 - \delta.$$

Thus with probability greater than $1 - \delta$ we have

$$P^{(n)}\{A\} > P\{A\} - \tfrac{1}{2}\epsilon > Q\{A^{\epsilon}\} + \tfrac{1}{2}\epsilon \geqq Q\{A^{\epsilon/2}\} + \tfrac{1}{2}\epsilon. \qquad \text{Q.E.D.}$$

PROOF OF PROPOSITION 1.3. For any $k \in K$ let $\epsilon_k > 0$ be such that $\inf_{j \in K - \{k\}} L(P_j, P_k) > \epsilon_k$. Choose $N_1(k)$ (see Example 6) such that for all $n > N_1$, and $k \neq j \in K$,

$$\mathbf{P}_k\{L(P^{(n)}, P_k) < \epsilon_k/2\} > 1 - \epsilon$$

and

$$\mathbf{P}_k\{L(P^{(n)}, P_j) > \epsilon_k/2\} > 1 - \epsilon$$

(using Lemma 5.2). The sets $A(k, \epsilon) = \{w: L(P^{(N_1)}, P_k) < \epsilon_k/2\}$, where $P^{(n)}$ is the sample measure, satisfy Condition I of [2]. Q.E.D.

EXAMPLE 6. Recalling (2.1), we see that in one dimension, if $L$ were $L'$ in the last proof we could choose $N_1$ independent of $k$. However with $L$ this cannot be done. Let $P_m$ put mass $1/m$ on the integers $1, \cdots, m$. Let $n < m$ then $\min_{\text{possible } P^{(n)}} (L(P^{(n)}, P_m)$ occurs when all $n$ samples are distinct so that $L(P^{(n)}, P) \geqq (m - n)m^{-1}$. Letting $A$ be the $m - n$ points not in the sample we see that $L(P^{(n)}, P_m) \to 1$ as $m \to \infty$.

PROOF OF PROPOSITION 1.4. Obviously if $\pi \in \mathfrak{F}$ then $\pi \in \mathfrak{V}$. Let $\{x_j\}_{j=1}^{\infty}$ be all the possible values attainable under $\pi$. Fix any $k \in K$ and let

$$\inf_{j \in K - \{k\}} V(P_k, P_j) = \epsilon(k) \tag{5.1}$$

where $\epsilon(k) > 0$ if $\pi \in \mathfrak{V}$. Next choose a finite $N(k, \epsilon(k))$ such that $P_k\{x_1, \cdots, x_N\} \geqq 1 - \epsilon(k)/4$. If there exists an $i(j)$ such that $|P_k\{x_i\} - P_j\{x_i\}| \geqq \epsilon(k)/8N$ then $D(P_k, P_j) \geqq \epsilon(k)/16N$. So suppose for some $j$, $|P_k\{x_i\} - P_j\{x_i\}| < \epsilon(k)/8N$ for all $i$. Then

$$\left|\sum_{i=1}^{N} P_j\{x_i\} - \sum_{i=1}^{N} P_k\{x_i\}\right| \leqq \epsilon(k)/8$$

and

$$\sum_{i=N+1}^{\infty} P_j\{x_i\} \leqq 3\,\epsilon(k)/8.$$

For any $B \in \mathfrak{A}$,

$$\begin{aligned} &|P_j\{B\} - P_k\{B\}| \\ &\quad \leqq \left|\sum_{1 \leqq i \leqq N x_i \in B} P_k\{x_i\} - P_j\{x_i\}\right| + \sum_{i=N+1}^{\infty} P_k\{x_i\} + \sum_{i=N+1}^{\infty} P_k\{x_i\} \\ &\quad \leqq (\epsilon(k)/8) + (3\,\epsilon(k)/8) + (\epsilon(k)/8) < \epsilon(k) \end{aligned}$$

which contradicts (5.1). Q.E.D.

LEMMA 5.3. *If $P_k \in \pi$ and $\pi - \{P_k\}$ are not distinguishable, then for every $\epsilon > 0$ and countable collection of disjoint measurable sets $\{A_i\}$, $A_i \in S^n$ there exists a $P_j \in \pi - \{P_k\}$ such that $|\mathbf{P}_j\{A_i\} - \mathbf{P}_k\{A_i\}| < \epsilon$, $i = 1, 2, \cdots$.*

PROOF. If not we may easily distinguish by looking at the fraction of sample points (in blocks of $n$) that fall into the $A_j$'s. (Compare to the proof of Theorem 2.1.) Q.E.D.

PROOF OF THEOREM 1.5. (if) For $n = 1$ the "if" assertion of the second statement is true. Let $\pi_n$ be empty and $\pi_{n-1}$ nonempty. To show $\mathfrak{D}_\pi \leqq n^{-1}$ select any nonrandom test $(N, d)$ and use the following inductive procedure. Take $Q_{n-1}$ in $\pi_{n-1}$ and let $\mathbf{Q}_{n-1}\{(N, d) \text{ is correct}\} = q_{n-1}$. Choose $m_{n-1}$ and $A_{n-1} \varepsilon S^{m_{n-1}}$ such that

$$E_{\mathbf{Q}_{n-1}}(d^{(n-1)} I_{\{N \leqq m_{n-1}\}} I_{A_{n-1}}) = \mathbf{Q}_{n-1}\{A_{n-1}\} \geqq q_{n-1} - \epsilon_{n-1}.$$

Use Lemma 5.3 to select $Q_{n-2} \varepsilon \pi_{n-2} - \{Q_{n-1}\}$ such that

$$|\mathbf{Q}_{n-1}\{A_{n-1}\} - \mathbf{Q}_{n-2}\{A_{n-1}\}| < \epsilon(n-1).$$

For general $0 \leqq k < n-1$ select $Q_k \varepsilon \pi_k - \pi_{k+1}$ such that

$$|\mathbf{Q}_k\{A_r\} - \mathbf{Q}_{k+1}\{A_r\}| < \epsilon(k+1)$$

for $r = k+1, k+2, \cdots, n-1$ and a new set $A_k \varepsilon S^{m_k}$

$$E_{\mathbf{Q}_k}(d^{(k)} I_{\{N \leqq m_k\}} I_{A_k}) = \mathbf{Q}_k\{A_k\} \geqq q_k - \epsilon_k.$$

Now

$$\begin{aligned} \mathbf{Q}_0\{A_j\} &= \mathbf{Q}_j\{A_j\} + \textstyle\sum_{i=0}^{j-1} (\mathbf{Q}_i\{A_j\} - \mathbf{Q}_{i+1}\{A_j\}) \\ &\geqq -|\mathbf{Q}_0\{A_j\} - \mathbf{Q}_1\{A_j\}| - \cdots - |\mathbf{Q}_{j-1}\{A_j\} - \mathbf{Q}_j\{A_j\}| + \mathbf{Q}_j\{A_j\} \\ &\geqq q_j - (\epsilon_j + \textstyle\sum_{t=0}^{j-1} \epsilon(t)). \end{aligned}$$

Choose the $\epsilon$'s so that

$$\epsilon_j + \textstyle\sum_{t=0}^{j-1} \epsilon(t) \leqq \epsilon/2^j.$$

Since the $A_k$'s are disjoint,

$$\begin{aligned} \mathbf{Q}_0\{(N, d) \text{ is correct}\} &\leqq 1 - \mathbf{Q}_0\{\textstyle\bigcup_1^{n-1} A_j\} \leqq 1 + \sum (\epsilon/2^i) - \sum_1^{n-1} q_j \\ &< 1 + \epsilon - \textstyle\sum_{j=1}^{n-1} q_j. \end{aligned}$$

If $q_j \geqq n^{-1}$ for each $j = 1, 2, \cdots, n-1$; then $\mathbf{Q}_0\{(N, d) \text{ is correct}\} \leqq n^{-1}$ so that $\mathfrak{D}_\pi \leqq n^{-1}$.

To obtain the inequality the other way is much easier. By definition $\pi_j - \pi_{j-1}$ is a distinguishable family. Therefore, we can distinguish with probability arbitrarily close to $n^{-1}$ by first randomly, with probability $n^{-1}$, selecting among the $n$ classes, $\pi_{n-1}, \pi_{n-2} - \pi_{n-1}, \cdots, \pi_0 - \pi_1$. We can distinguish within these classes with probability arbitrarily close to one. Q.E.D.

PROOF OF PROPOSITION 1.6. Suppose that $E_{\mathbf{P}_k}(d^{(k)}) > 1 - \epsilon$ and $\mathbf{P}_k\{N < \infty\} = 1$ for $k \varepsilon K$. By the construction of the proof of Proposition 1.1 we may assume that $(N, d)$ is nonrandom. Theorem 1.5 allows us to select a $j$ such that $P_j \varepsilon \pi_1$. We may find a set $A$ in $S^n$ for some finite $n$, such that $N \leqq n$ on $A$ and $\mathbf{P}_j\{A\} > 1 - \epsilon/4$. Let $\pi' = \{P_k : P_k \neq P_j \text{ and } \mathbf{P}_k\{A\} > 1 - \epsilon/2\}$. Then $P_j$ and $\pi'$ are not distinguishable (since $P_j$ and $\pi - \{P_j\}$ are not distinguishable). Let $A_k = \{d^{(k)} = 1\} \cap A$. By Lemma 5.3 there exist an infinite number of $P_k \varepsilon \pi'$ such that $|\mathbf{P}_k\{A_m\} - \mathbf{P}_j\{A_m\}| < \epsilon/4$ for $m = 1, 2, 3, \cdots$.

Since $\sum_m \mathbf{P}_j\{A_m\} \leqq 1$ all but a finite number of $A_m$ have $\mathbf{P}_j\{A_m\} < \epsilon/4$. Choose an $n$ such that $\mathbf{P}_j\{A_n\} < \epsilon/4$, $P_n \varepsilon \pi'$ and $|\mathbf{P}_n\{A_m\} - \mathbf{P}_j\{A_m\}| < \epsilon/4$ for $m = 1, 2, 3 \cdots$. Then $E_{\mathbf{P}_n}(d^{(n)}I_{\{N<\infty\}}) \leqq \mathbf{P}_n\{A^c\} + \mathbf{P}_n\{A_n\} \leqq \epsilon/2 + \mathbf{P}_j\{A_n\} + \epsilon/4 < \epsilon$ which gives a contradiction. Q.E.D.

**6. Acknowledgment.** The authors would like to thank the referee for vastly improving the quality of the paper.

## REFERENCES

[1] DUDLEY, R. M. (1968). Distances of probability measures and random variables. To be published in the *Ann. Math. Statist.* 1563–1572.

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

[3] HEWITT, E. and ROSS, E. A. (1963). *Abstract Harmonic Analysis.* Springer-Verlag.

[4] HOEFFDING, W. and WOLFOWITZ, J. (1958). Distinguishability of sets of distributions. *Ann. Math. Statist.* **29** 700–718.

[5] KIEFER, J. and WOLFOWITZ, J. (1958). On the deviations of the emperic distribution of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

[6] KRAFT, C. (1955). Some conditions for the consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** No. 6, 125–142.

[7] PARTHASARATHY, K. R., RANGA RAO, R. VARADHAN, S. R. S. (1963). Probability distributions on locally compact Abelian groups. *Illinois J. Math.* **7** 337–369.

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

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