The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 322–324

AN EXAMPLE OF THE DIFFERENCE BETWEEN THE LÉVY AND LÉVY-PROKHOROV METRICS

BY LLOYD FISHER[1] AND D. W. WALKUP

The University of Washington and Boeing Scientific Research Laboratories

The notation and terminology of [1] will be used below. The purpose of this note is to prove the following:

PROPOSITION. *Let e be a probability measure on R^1 which is absolutely continuous with respect to Lebesgue measure. Also let $\mu \to \mu * e$ be a 1-1 map of the Borel probability measures into themselves. Let $\mathfrak{F}$ be a countable family of Borel probability measures and*

$$\pi = \{\mu * e : \mu \varepsilon \mathfrak{F}\}. \quad \textit{Then:}$$

(i) *π is finitely distinguishable if the members of $\mathfrak{F}$ are uniformly isolated in the Lévy metric, but the converse does not hold.*

(ii) *If π is finitely distinguishable then the members of $\mathfrak{F}$ are uniformly isolated in the Lévy-Prokhorov metric but the converse does not hold.*

PROOF. Everythring except the converse in (ii) follows immediately from [1]. To construct an example showing the converse does not hold the following combinatorial lemma is needed.

LEMMA. *For any two integers $0 \leqq m < n$ there exists a finite set S and subsets $S_1, \cdots, S_l$ of S such that (letting $|B|$ denote the cardinality of B):*

(1) $|S_i| = n, i = 1, 2, \cdots, l$.

(2) $|S_i \cap S_j| \leqq m$ *if* $i \neq j$.

(3) *If* $C \subseteq S_i$ *and* $|C| = m$, *there exists at least one* S_j *such that* $S_i \cap S_j = C$.

PROOF OF LEMMA. Select an $m + 1$ by n matrix A with entries from the finite field F_p with characteristic p such that any submatrix consisting of $m + 1$ column vectors is nonsingular, i.e., has an inverse matrix over the finite field. This is always possible for given m and n if p is sufficiently large. Let V be the collection of row vectors of length $m + 1$ with entries from F_p and let Z be the corresponding collection of row vectors of length n with entries from F_p which may be written in the form yA, $y \varepsilon V$. Since A has rank $m + 1$, $|Z| = |V| = p^{m+1}$. Now let $S = \{(k, q) : k = 1, 2, \cdots, n; q \varepsilon F_p\}$. Each member z_i of Z determines a unique subset S_i of S where $(k, q) \varepsilon S_i$ iff the kth coordinate of z_i is q. It is clear that $|S_i| = n$. Since every $m + 1$ by $m + 1$ submatrix of A is nonsingular, it follows that for any set of values from F_p for any $m + 1$ coordinates there is one and only one vector z in Z with those coordinate values. From this fact parts (2) and (3) of the lemma follow immediately.

Returning to the construction of an example in (ii) of the proposition we may use the lemma with $n = 2m$, $m = 1, 2, \cdots$, to choose a family $\mathfrak{F}$ of discrete prob-

Received 3 June 1968.

[1] Research supported by NSF Grant GP-7519.

ability measures, where $\mathfrak{F} = \bigcup_{m=1}^{\infty} \mathfrak{F}_m$. Each $\mathfrak{F}_m$ has the following properties.
(a) $\mu \in \mathfrak{F}_j, \lambda \in \mathfrak{F}_k, j \neq k$ implies support $\lambda \cap$ support $\mu = \emptyset$.
(b) $\mu \in \mathfrak{F}_m$ puts mass $1/2m$ at $2m$ different integers, say $\mu_1, \cdots, \mu_{2m}$.
(c) $\mu, \lambda \in \mathfrak{F}_m, \mu \neq \lambda$ implies
$|\{\mu_1, \cdots, \mu_{2m}\} \cap \{\lambda_1, \cdots, \lambda_{2m}\}| \leq m$.
(d) $\mu \in \mathfrak{F}_m, \{x_1, \cdots, x_m\} \subseteq \{\mu_1, \cdots, \mu_{2m}\}$ implies the existence of a $\lambda \in \mathfrak{F}_m$ such that
$\{\mu_1, \cdots, \mu_{2m}\} \cap \{\lambda_1, \cdots, \lambda_{2m}\} = \{x_1, \cdots, x_m\}$.
(e) $|\mathfrak{F}_m| < \infty$.

It is easy to see that $\mu, \lambda \in \mathfrak{F}$ and $\mu \neq \lambda$ implies $L(\mu, \lambda) \geq \frac{1}{2}$ where L is the Lévy-Prokhorov metric.

Let e be uniform on [0, 1]. If we observe a sequence of iid observations from a measure of π since the measures of $\mathfrak{F}$ have support on the integers, the integer part of each observation is a sufficient statistic and we may restrict ourselves to decision rules based on the sequence of integers (integer parts) observed. Suppose that π is finitely distinguishable, then there exists for each $\epsilon > 0$ a number $N = N(\epsilon)$ and a set of functions ϕ_μ, $\mu \in \mathfrak{F}$ defined on N-tuples of integers such that for each N-tuple $\mathbf{x}$, $\sum \phi_\mu(\mathbf{x}) = 1$. ($\phi_\mu(\mathbf{x})$ is probability that we decide we are sampling from μ if we observe $\mathbf{x}$.) Further

$$E(\phi_\mu \mid \mu) \geq 1 - \epsilon \quad \text{for all} \quad \mu \in \mathfrak{F}, \qquad \text{that is,}$$

we make the correct decision with probability $\geq 1 - \epsilon$ regardless of which μ we are observing.

In particular, let $\epsilon = \frac{1}{4}$ and $n = N(\epsilon)$. Choose an element of $\mathfrak{F}_n$ at random (equally likely) and observe n iid observations, giving a vector $\mathbf{x}$. Use the ϕ_μ's to make a decision. Let $P(\cdot)$ be the probability measure referring to this situation. Let C be the event the decision made is correct. We find two estimates for $P(C)$.

$$P(C) = \sum_{\mu \in \mathfrak{F}_n} P(C \mid \mu)P(\mu) \geq \sum_{\mu \in \mathfrak{F}_n} (\tfrac{3}{4})P(\mu) = \tfrac{3}{4}$$

where μ represents the event that we are observing rv's with distribution μ.

For the second estimate let X be the set of observational vectors $\mathbf{x}$ which occur with a positive probability for at least one $\mu \in \mathfrak{F}_n$. Then:

$$P(C) = \sum_{\mathbf{x} \in X} P(C \mid \mathbf{x})P(\mathbf{x})$$

where $\mathbf{x}$ denotes the event that $\mathbf{x}$ has been observed. Further:

$$P(C \mid \mathbf{x}) = \sum_{\mu \in \mathfrak{F}_n} \phi_\mu(\mathbf{x})P(\mu \mid \mathbf{x}).$$

Now $P(\mu \mid \mathbf{x})$ is the same for any $\mu \in \mathfrak{F}_n$ whose support contains all the coordinates of x. By (d) there are at least two such measures. Thus, $P(\mu \mid \mathbf{x}) \leq \frac{1}{2}$ for all $\mu \in \mathfrak{F}_n$, $\mathbf{x} \in X$ and

$$P(C \mid \mathbf{x}) \leq (\tfrac{1}{2}) \sum_{\mu \in \mathfrak{F}_n} \phi_\mu(\mathbf{x}) \leq \tfrac{1}{2}.$$

Finally,

$$P(C) \leqq \sum_{\mathbf{x} \varepsilon X} (\tfrac{1}{2}) P(\mathbf{x}) = \tfrac{1}{2}.$$

The two estimates contradict each other from which we conclude that π is not finitely distinguishable.

REFERENCE

[1] FISHER, L. and VAN NESS, J. W. (1969) Distinguishability of probability measures. *Ann. Math. Statist.* **40** No. 2.