The Annals of Statistics
1994, Vol. 22, No. 1, 106–117

A TOPOLOGICAL CRITERION FOR HYPOTHESIS TESTING

BY AMIR DEMBO[1] AND YUVAL PERES[2]

Stanford University

A simple topological criterion is given for the existence of a sequence of tests for composite hypothesis testing problems, such that almost surely only finitely many errors are made.

1. Introduction. Suppose you are given, sequentially, independent observations $X_1, X_2, X_3, \dots$ from an unknown distribution μ, with the goal of deciding between two mutually exclusive hypotheses: $\mu \in H_0$ or $\mu \in H_1$. For each n, after viewing $X_1, \dots, X_n$ you guess 0 or 1. When can you ensure, with probability 1, that you will guess correctly from some point on? For example, limiting ourselves to distributions supported on a finite interval, we shall see this is possible if H_0 is $\{\mu\colon \mu$ has a rational mean$\}$ and H_1 is $\{\mu\colon \mu$ has mean in $\mathbb{Q} + \sqrt{2}\}$, but it is impossible if H_1 is enlarged to $\{\mu\colon \mu$ has an irrational mean$\}$.

Our interest in this problem was initiated by the results in [7] which, following [2], obtain certain sufficient conditions by explicit constructions. The main results of this work, which are necessary and sufficient conditions under which deciding between H_0 and H_1 is possible, are stated in the next section. We also discuss there the relation with the classical works of [6] and [8]. For other related works, see [10] and [11]. Two lemmas concerning a deterministic metric space setting are established in Section 3. Section 4 is devoted to the proofs of our main results while extensions are discussed in Section 5.

2. Statement of results.

DEFINITION. Let H_0 and H_1 be two ensembles of probability measures on $\mathbb{R}^d$. We say that H_0 and H_1 are *discernible* if there exists a sequence of Borel functions $f_n\colon (\mathbb{R}^d)^n \to \{0, 1\}$ such that for $j = 0, 1$ and each $\mu \in H_j$, if $\{X_k\}_{k \ge 1}$ are independent $\mathbb{R}^d$-valued random variables with distribution μ, then

$$\lim_{n \to \infty} f_n(X_1, X_2, \dots, X_n) = j \quad \text{almost surely (a.s.).}$$

Our first result characterizes exactly the pairs (H_0, H_1) which are discernible when H_j are determined by the location of the mean of the distribution.

Received April 1992; revised March 1993.

[1]Supported in part by NSF Grant DMS 92-09712.

[2]Supported in part by the Weizmann postdoctoral fellowship.

AMS 1991 *subject classifications.* Primary 62F03; secondary 62G10, 60F15.

Keywords and phrases. Hypothesis testing, moment conditions.

THEOREM 1. *Let A_0 and A_1 be subsets of $\mathbb{R}^d$. Denote by $H(A_j,p)$ the set of all distributions with mean in A_j and finite p-th moment.*

(i) *For $p > 1$ the ensembles $H(A_0,p)$ and $H(A_1,p)$ are discernible iff A_0 and A_1 are contained in disjoint F_σ-sets (an F_σ-set is any countable union of closed sets).*
(ii) *The ensembles $H(A_0,1)$ and $H(A_1,1)$ are discernible iff A_0 and A_1 are contained in disjoint open sets.*

REMARK 1. We find it interesting that moment assumptions are reflected in the separation conditions for discernibility.

In fact, any integrability condition stronger than first moment suffices for the equivalence in part (i) to hold. More precisely, let $\lambda(t)$ be any Borel positive function satisfying

$$\lim_{t\to\infty} \frac{\lambda(t)}{t} = \infty,$$

and denote by $H(A_j;\lambda)$ the set of distributions μ with mean in A_j for which $\int_{-\infty}^{\infty} \lambda(|x|)\, d\mu(x) < \infty$. Then $H(A_0;\lambda)$ and $H(A_1;\lambda)$ are discernible iff A_0 and A_1 are separated by F_σ-sets. See the remark following the proof of Theorem 1 in Section 4.

REMARK 2. In a different direction, the equivalence in part (i) of the theorem also holds for the ensembles $H_C(A_j)$ of *compactly supported* measures with mean in A_j, and even for more restricted classes of distributions (see the proof).

EXAMPLES.

(i) If A_0 is a closed set and A_1 its complement in $\mathbb{R}^d$, then $H(A_0,p)$ and $H(A_1,p)$ are discernible iff $p > 1$.
(ii) The same holds for any pair A_0,A_1 of disjoint countable sets such as $A_0 = \mathbb{Q}$ and $A_1 = \mathbb{Q} + \sqrt{2}$.
(iii) For $A_0 = \mathbb{Q}$ and $A_1 = \mathbb{R}\backslash\mathbb{Q}$, the ensembles $H(A_j,p)$ are *not* discernible for any p, since by Baire's theorem (see [13], Theorem 5.6) the irrationals are not an F_σ-set.

COROLLARY 1 (cf. [7]). *Let $A_0 \subset \mathbb{R}^d$ be an F_σ-set. Then there exists a subset $A_1 \subset \mathbb{R}^d - A_0$ such that the complement of $A_0 \cup A_1$ has Lebesgue measure zero, and the hypotheses $H(A_0,p)$ and $H(A_1,p)$, determined as in Theorem 1, are discernible for any $p > 1$.*

PROOF. In fact, Lebesgue measure can be replaced here by any Radon measure ν. Simply find an F_σ-set $A_1 \subset \mathbb{R}^d - A_0$ such that $\nu[\mathbb{R}^d - (A_0 \cup A_1)] = 0$ (this is always possible [13], Theorem 2.18) and apply Theorem 1. □

REMARK. This corollary was established in [2] for A_0 which consists of a countable number of points.

Focusing on the mean of a distribution is somewhat arbitrary. Replacing it by another linear functional is straightforward. However, one may consider families of distributions determined (nonlinearly) by other parameters (e.g., the cumulants). It is natural to assume that the parametrization is continuous; thus we are led to the following statement, in which the topology on probability measures corresponds to convergence in law.

THEOREM 2. *Let H_0 and H_1 be two disjoint families of probability measures on $\mathbb{R}^d$.*

(i) *If H_0 and H_1 are contained in disjoint F_σ-sets of measures, then they are discernible.*
(ii) *The converse holds under the assumption that all measures in $H_0 \cup H_1$ are absolutely continuous with respect to Lebesgue measure and, for each $\mu \in H_0 \cup H_1$, there is some $p > 1$ such that the density of μ is in $L^p(\mathbb{R}^d)$.*

REMARK. Let F be the distribution function of a specific probability measure supported on finitely many rational points. Taking H_0 and H_1 as the ensembles of measures with distribution functions $\{F(x-c)\colon\ c \in \mathbb{Q}\}$ and $\{F(x-c)\colon\ c \notin \mathbb{Q}\}$, respectively, it is clear that H_0 and H_1 are discernible, although they cannot be separated by F_σ-sets. This example illustrates the need for *some* auxiliary condition in part (ii) of Theorem 2, although the particular assumption made there could be weakened.

Theorem 2 extends to the discernibility of any finite or countable number of hypotheses. To that end, $\{H_i\}_{i\in I}$, are *discernible* iff there exist Borel functions $f_n\colon (\mathbb{R}^d)^n \to I$ such that $\lim_{n\to\infty} f_n(X_1,\dots,X_n) = j$ whenever $\{X_k\}_{k\geq 1}$ are i.i.d. random variables with distribution $\mu \in H_j$ and $j \in I$.

COROLLARY 2. *Let I be a finite or a countable set. If $\{H_i\}_{i\in I}$ are contained in disjoint F_σ-sets of measures, then they are discernible. The converse holds under the assumption that all measures in $\bigcup_{i\in I} H_i$ are absolutely continuous with respect to Lebesgue measure, with each density in some L^p, $p > 1$.*

REMARK. Corollary 1 and Theorem 1 extend to a countable number of hypotheses in a similar manner (cf. [2] and [14]).

The notion of distinguishability studied in [6] is the same as our discernibility apart from replacing the a.s. convergence with convergence in probability. It is shown there that a sufficient condition for distinguishability is that H_0 and H_1 are contained in disjoint KS-open sets of probability measures on $\mathbb{R}^d$, and a necessary condition is that they are contained in disjoint open sets with respect to the variational metric on probability measures. Here KS-open refers to the topology induced by the Kolmogorov–Smirnov distance between distribution functions. While useful for the study of parametric ensembles (see the study of multivariate normal distributions in [6]), these conditions yield little information in the context of Theorem 1.

The work of [8] extends [6] in giving a necessary and sufficient condition for the consistent estimation of real-valued parameters. When specialized to the context of Theorem 2 this condition involves all n-fold product measures obtained from the measures in $H_0 \cup H_1$ and is thus difficult to verify. It is considerably simplified once $H_0 \cup H_1$ is a countable union of sets of measures, each set being sequentially compact with respect to the τ-topology. However, the setup of Theorem 1 does not fall into this category.

3. Metric discernibility. In proving Theorem 1, it is natural to use the convergence of the sample means to the expectation. In fact, some (standard) estimate on the *rate* of convergence is needed. We shall find it convenient to consider this in a more general setting.

DEFINITION. Two disjoint sets A_0 and A_1 in a metric space (Ω, ρ) are *metrically discernible* if for every positive sequence $\varepsilon_n \to 0$ there exists a sequence of functions $g_n: \Omega \to \{0,1\}$ such that (for $j \in \{0,1\}$) if $\{y_n\}$ is a sequence in Ω satisfying $\rho(y_n, z) < \varepsilon_n$ for some $z \in A_j$ and all sufficiently large n, then $g_n(y_n) \to j$ as $n \to \infty$.

REMARK. We refer to the sequence $\{g_n\}$ as a *discerning* sequence corresponding to $\{\varepsilon_n\}$.

LEMMA 3. *Two sets in a metric space are metrically discernible iff they are contained in disjoint F_σ-sets. In this case the functions g_n in the definition of metric discernibility can be taken to be Borel measurable.*

Before proving Lemma 3, let us emphasize the role of the (arbitrary) "rate" sequence $\{\varepsilon_n\}$; the required separation of A_0 and A_1 becomes much stronger when this sequence is discarded.

LEMMA 4. *Let A_0 and A_1 be subsets of a metric space Ω. There exist functions $h_n: \Omega \to \{0,1\}$ such that, for $j \in \{0,1\}$, we have $h_n(y_n) \to j$ as $n \to \infty$, for any sequence $\{y_n\}$ which converges to a point in A_j, iff A_0 and A_1 can be separated by open sets in Ω.*

PROOF. If there exists a sequence $\{h_n\}$ with the specified property, then A_0 is in the interior of the set $\{y \in \Omega: \sum_{n=1}^{\infty} h_n(y) < \infty\}$ [or, otherwise, one can find $z_m \to y \in A_0$ with $h_n(z_m) = 1$ for infinitely many values of n, leading to a contradiction]. Likewise, A_1 is in the interior of $\{y \in \Omega: \sum_{n=1}^{\infty}[1 - h_n(y)] < \infty\}$. Since these sets are disjoint, we have separated A_0 and A_1 by open sets. Conversely, if $A_0 \subset G_0$, $A_1 \subset G_1$, where G_0 and G_1 are disjoint open sets, then $h_n(y) = 1_{y \in G_1}$ have the specified property. □

PROOF OF LEMMA 3. (i) By assumption there are closed sets B_n and C_n such that $B = \bigcup_{n=1}^{\infty} B_n$ is disjoint from $C = \bigcup_{n=1}^{\infty} C_n$ and $A_0 \subset B$, $A_1 \subset C$. We are given a positive sequence $\varepsilon_n \to 0$. For $y \in \Omega$ and $n \geq 1$, let

$$K_0(y,n) = \inf\{k \geq 1: \rho(y,B_k) < \varepsilon_n\}$$

and

$$K_1(y,n) = \inf\{k \geq 1: \rho(y,C_k) < \varepsilon_n\},$$

where the infimum of the empty set is $+\infty$. Define $g_n(y) = 0$ if $K_0(y,n) < K_1(y,n)$ and $g_n(y) = 1$ otherwise. Let $\{y_n\}$ be a sequence in Ω satisfying $\rho(y_n,z) < \varepsilon_n$ for $n > n_0$. Suppose first that $z \in A_0$. Then $z \in B_r$ for some r and thus $\rho(z, \bigcup_{j=1}^{r} C_j) > 0$. For any n which is large enough so that $\rho(z,y_n) < \varepsilon_n$ and $2\varepsilon_n < \rho(z, \bigcup_{j=1}^{r} C_j)$, we have $\rho(y_n, \bigcup_{j=1}^{r} C_j) > \varepsilon_n$. This implies that $K_0(y_n,n) \leq r < K_1(y_n,n)$ and therefore $g_n(y_n) = 0$. The case $z \in A_1$ is similar, which proves that $\{g_n\}$ is a discerning sequence of functions corresponding to $\{\varepsilon_n\}$. Clearly these functions g_n are Borel measurable.

(ii) Now assume the existence of discerning functions $g_n: \Omega \to \{0,1\}$ for $A_0, A_1 \subset \Omega$ and *some* positive sequence $\varepsilon_n \to 0$. Let $B_{j,m} = \{y: g_m(y) = j\}$ and, for $j = 0$, 1, consider

$$\Omega_j = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{x \in \Omega: \rho(x, \Omega \backslash B_{j,m}) \geq \varepsilon_m\}.$$

Since $\{x \in \Omega: \rho(x, \Omega\backslash B_{j,m}) \geq \varepsilon_m\}$ are closed (and disjoint for fixed m), the sets Ω_0 and Ω_1 are disjoint F_σ-sets. The definition of metric discernibility ensures that $A_0 \subset \Omega_0$ and $A_1 \subset \Omega_1$. For example, if $x \in A_0$ and $x \notin \Omega_0$, then $\rho(x, B_{1,m_k}) < \varepsilon_{m_k}$ for some subsequence $m_k \to \infty$, contradicting the metric discernibility. □

REMARK. The proof shows that the existence of a discerning sequence of functions corresponding to *one* positive sequence $\varepsilon_n \to 0$ suffices for metric discernibility; this is also easy to verify directly.

4. Probabilistic discernibility: Proofs.

PROOF OF THEOREM 1. (i.1) Assume that $A_0, A_1 \subset \mathbb{R}^d$ are separated by F_σ-sets. Let $\varepsilon_n = 1/\log(n+1)$. Lemma 3 provides us with a discerning sequence of functions $g_n: \mathbb{R}^d \to \{0,1\}$, corresponding to $\{\varepsilon_n\}$. If $X_1, X_2, \ldots$ is a sequence of independent observations from a distribution μ on $\mathbb{R}^d$ with mean $z = \int_{\mathbb{R}^d} x \, d\mu$ and a finite p-th moment, for some $2 > p > 1$, then Marcinkiewicz' theorem asserts that almost surely

$$\lim_{n\to\infty} n^{-1/p} \left\| \sum_{i=1}^{n} (X_i - z) \right\| = 0,$$

where $\|\cdot\|$ denotes the Euclidean norm (see [9], Section 16.4). In particular, almost surely, for sufficiently large n,

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i - z \right\| < \varepsilon_n.$$

(If we restrict attention to distributions with finite second moment, then the law of the iterated logarithm or weaker estimates suffice.) Thus if $\mu \in H(A_j, p)$, that is, if $z \in A_j$, then almost surely

$$g_n \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) \to j,$$

so we define $f_n(x_1, \dots, x_n) = g_n((1/n)\sum_1^n x_i)$.

(i.2) Now we assume that the ensembles $H_C(A_0)$ and $H_C(A_1)$ of compactly supported distributions with mean in A_0 and A_1, respectively, are discernible, and we infer that A_0 and A_1 are separated by F_σ-sets; this clearly suffices. Denote by $f_n: (\mathbb{R}^d)^n \to \{0,1\}$ the functions in the definition of discernibility. Let $\mathcal{L}_{dn}$ be Lebesgue measure on $\mathbb{R}^{dn}$. Lusin's theorem [13] provides continuous functions $\phi_n: (\mathbb{R}^d)^n \to [0,1]$ such that

$$\mathcal{L}_{dn}\{v \in \mathbb{R}^{dn}: \phi_n(v) \neq f_n(v)\} < \frac{1}{n}.$$

Consider i.i.d. random variables $X_1, X_2, X_3, \dots$ which are uniformly distributed on a unit cube centered at the origin in $\mathbb{R}^d$. For $z \in A_j$ (where $j \in \{0,1\}$), we have

$$f_n(X_1 + z, X_2 + z, \dots, X_n + z) \to j \quad \text{a.s. as } n \to \infty.$$

We may take expectations here (bounded convergence) and then use the choice of ϕ_n to infer that

$$E\phi_n(X_1 + z, X_2 + z, \dots, X_n + z) \to j \quad \text{as } n \to \infty.$$

For each n, the set

$$B_n = \bigcap_{k=n}^{\infty} \{y \in \mathbb{R}^d: E\phi_k(X_1 + y, \dots, X_k + y) \leq \tfrac{1}{3}\}$$

is closed by the continuity of ϕ_k. Clearly $A_0 \subset \bigcup_n B_n$. Thus $B = \bigcup_{n=1}^{\infty} B_n$ and

$$C = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{y \in \mathbb{R}^d: E\phi_k(X_1 + y, \dots, X_k + y) \geq \tfrac{2}{3}\}$$

are disjoint F_σ-sets which separate A_0 and A_1.

(ii.1) If A_0 and A_1 are separated by open sets G_0 and G_1, then the strong law of large numbers immediately implies that $H(A_0, 1)$ and $H(A_1, 1)$ are discernible—simply take $f_n(x_1, \ldots, x_n) = 1_{G_1}((1/n)\sum_{i=1}^n x_i)$.

(ii.2) The converse is less straightforward; the following construction is motivated by the construction in [3]. We are given that $H(A_0, 1)$ and $H(A_1, 1)$ are discernible. Let $f_n: \mathbb{R}^{nd} \to \{0, 1\}$ be the sequence of functions which discerns between them. Assume that A_0 and A_1 *cannot* be separated by open sets in $\mathbb{R}^d$. This implies that (at least) one of the sets intersects the closure of the other; without loss of generality, assume that A_0 intersects $\bar{A}_1$. Thus there is a sequence of points $\{y_m\}$ in A_1 which converges to some point $z \in A_0$. Passing to a subsequence if necessary, we may assume that

$$\sum_{m=1}^{\infty} \|y_{m+1} - y_m\| < \infty.$$

Next, we shall define inductively a sequence $\{\mu_m\}$ of probability measures such that $\int x \, d\mu_m = y_m$. The sequence will converge in norm to a measure μ with mean z: independent observations from μ will masquerade for a long time as observations from μ_m, which will yield the desired contradiction. First, let μ_1 be the point mass

$$\mu_1 = \delta(y_1).$$

If μ_m is already defined and has mean $y_m \in A_1$, then by the definition of discernibility there is an integer $N(m)$ such that

$$n \geq N(m) \quad \Longrightarrow \quad \int f_n(x_1, \ldots, x_n) \, d\mu_m^n \geq \tfrac{1}{2},$$

where μ_m^n is the n-fold product measure. We may, and shall, take $N(m) > N(m-1)$ for $m \geq 2$.

Now μ_{m+1} is obtained from μ_m by averaging with a small point mass:

$$\mu_{m+1} = (1 - \alpha_m)\mu_m + \alpha_m \delta\big(\alpha_m^{-1}(y_{m+1} - y_m) + y_m\big),$$

so that indeed $\int x \, d\mu_{m+1} = y_{m+1}$.

Here we take $\alpha_m = 2^{-m} N(m)^{-1}$. Using the total variation norm on measures, $\|\mu_{m+1} - \mu_m\| \leq 2\alpha_m$ so the sequence $\{\mu_m\}$ converges in norm to some probability measure μ. This measure has a finite first moment since

$$\int_{\mathbb{R}^d} \|x\| \, d\mu \leq \|y_1\| + \sum_{m=1}^{\infty} \|y_{m+1} - y_m + \alpha_m y_m\| < \infty.$$

Therefore its mean is

$$\int x \, d\mu = \lim_{m \to \infty} \int x \, d\mu_m = z \in A_0.$$

(All the measures μ_m are bounded by a constant multiple of μ, so the leftmost equality above follows from uniform integrability.)

We have

$$||\mu - \mu_m|| \leq 2\sum_{k=m}^{\infty} \alpha_k \leq N(m)^{-1}2^{2-m}.$$

Now, for every n,

$$||\mu^n - \mu_m^n|| \leq n\,||\mu - \mu_m||$$

and also

$$\int f_n(x_1, \ldots, x_n)\, d\mu^n \geq \int f_n(x_1, \ldots, x_n)\, d\mu_m^n - ||\mu^n - \mu_m^n||.$$

Taking $n = N(m)$ yields

$$\int f_n(x_1, \ldots, x_n)\, d\mu^n \geq \tfrac{1}{2} - n\,||\mu - \mu_m|| \geq \tfrac{1}{2} - 2^{2-m}.$$

Since $\mu \in H(A_0, 1)$, our choice of $\{f_n\}$ forces the left-hand side in this inequality to tend to zero as m and $n = N(m)$ tend to ∞, a contradiction. □

REMARK. A stronger version of Theorem 1(i) was stated immediately after the statement of that theorem; the proof is identical, except that instead of Marcinkiewicz' law of large numbers the following result, proved in 1946 by Feller [4], is invoked.

LEMMA 5. *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. real random variables such that $E[\lambda(|X_n|)] < \infty$, for some Borel positive function $\lambda(t)$ which satisfies*

$$\lim_{t\to\infty} \frac{\lambda(t)}{t} = \infty.$$

Then there exists a sequence $\varepsilon_n \downarrow 0$, depending only on the function λ, such that almost surely

$$\left|\frac{1}{n}\sum_{i=1}^{n} X_i - EX_1\right| < \varepsilon_n,$$

for all sufficiently large n.

Strictly speaking, Feller proves this (with precise estimates on ε_n) assuming that $\lambda(t)/t$ *increases* to ∞ and that $\lambda(t)/t^p$ *decreases* for some $1 < p < 2$. However, for any function λ satisfying $\lambda(t)/t \to \infty$, there exists another function $\tilde{\lambda}$ with these two properties such that $\tilde{\lambda}(t) \leq \lambda(t)$ for all large t, so we can apply Feller's result to λ.

PROOF OF THEOREM 2. (i) The space of probability measures $M_1(\mathbb{R}^d)$ may be equipped with a translation-invariant metric ρ compatible with convergence in law. Let $X_1, \dots, X_n, \dots$ be independent observations from $\mu \in M_1(\mathbb{R}^d)$. Let $L_n = (1/n)\sum_{i=1}^n \delta_{X_i}$ denote the n-th empirical measure.

For any $\varepsilon > 0$, there exist $\delta(\varepsilon) > 0$ and continuous functions f_l: $\mathbb{R}^d \to [-1, 1]$, for $l = 1, \dots, K(\varepsilon)$, such that, for any $\mu \in M_1(\mathbb{R}^d)$,

$$\{\nu\colon \rho(\nu,\mu) < \varepsilon\} \supset \bigcap_{l=1}^{K(\varepsilon)} \left\{\nu\colon \left|\int f_l\, d\mu - \int f_l\, d\nu\right| < \delta(\varepsilon)\right\}.$$

Note that $K(\varepsilon)$, $\delta(\varepsilon)$ and the functions f_l are independent of μ. Therefore,

$$\begin{aligned}\text{Prob}\,\big(\rho(L_n,\mu) \geq \varepsilon\big) &\leq \sum_{l=1}^{K(\varepsilon)} \text{Prob}\left(\left|\frac{1}{n}\sum_{i=1}^n f_l(X_i) - \int_{\mathbb{R}^d} f_l\, d\mu\right| \geq \delta(\varepsilon)\right)\\ &\leq K(\varepsilon)e^{-n\delta^2(\varepsilon)/2}\end{aligned}$$

where the right-hand inequality follows from Hoeffding's bound [5]. Let $N(m)$ increase rapidly enough so that

$$\sum_{n=N(m)}^{\infty} K\left(\frac{1}{m}\right) \exp\left[\frac{-n\delta(1/m)^2}{2}\right] \leq \frac{1}{m^2}.$$

Choosing $\varepsilon_n = 1/m$ for $N(m) \leq n < N(m+1)$ we find that

$$\sum_{n=1}^{\infty} \text{Prob}\Big(\rho(L_n,\mu) \geq \varepsilon_n\Big) < \infty,$$

and by the Borel–Cantelli lemma almost surely, $\rho(L_n, \mu) < \varepsilon_n$ for all n large enough. Finally, given hypotheses H_0, $H_1 \subset M_1(\mathbb{R}^d)$ which are separated by F_σ-sets, we invoke Lemma 3 to obtain a discerning sequence of functions g_n: $M_1(\mathbb{R}^d) \to \{0, 1\}$ corresponding to $\{\varepsilon_n\}$ above. Setting

$$f_n(x_1, \dots, x_n) = g_n\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)$$

shows that H_0 and H_1 are discernible.

(ii) We are given hypotheses $H_0, H_1 \subset M_1(\mathbb{R}^d)$ and functions f_n: $\mathbb{R}^{dn} \to \{0, 1\}$ which exhibit their discernibility.

Step 1. Choose continuous functions ϕ_n: $\mathbb{R}^{dn} \to [0, 1]$ such that

$$\mathcal{L}_{dn}\{v \in \mathbb{R}^{dn}\colon \phi_n(v) \neq f_n(v)\} < n^{-n},$$

where $\mathcal{L}_{dn}$ is Lebesgue measure. We claim that, for any $\mu \in H_0 \cup H_1$,

$$\lim_{n\to\infty} \int_{\mathbb{R}^{dn}} |f_n(x_1, \dots, x_n) - \phi_n(x_1, \dots, x_n)|\; d\mu^n = 0.$$

Indeed, letting ψ denote the density $d\mu/d\mathcal{L}_d$, there is, by assumption, some $p > 1$ for which $\psi \in L^p(\mathbb{R}^d)$. Defining q by $1/p+1/q = 1$, we have by Hölder's inequality,

$$\begin{aligned}\int |f_n - \phi_n|\,\psi(x_1)\psi(x_2)\cdots\psi(x_n)\; d\mathcal{L}_{dn} \\ \leq \left[\int |f_n - \phi_n|^q \; d\mathcal{L}_{dn}\right]^{1/q} \left[\int \psi(x_i)^p \cdots \psi(x_n)^p \; d\mathcal{L}_{dn}\right]^{1/p} \\ \leq n^{-n/q}||\psi||_p^n \to 0 \quad \text{as} \quad n \to \infty,\end{aligned}$$

establishing the claim.

Step 2. By the definition of convergence in law, it is easily shown that, for any finite k, the mapping taking $\mu \in M_1(\mathbb{R}^d)$ to $\mu^k \in M_1(\mathbb{R}^{dk})$ is continuous. Therefore the sets $\{\mu \in M_1(\mathbb{R}^d)\text{: } \int_{\mathbb{R}^{dk}} \phi_k d\mu^k \leq \frac{1}{3}\}$ are closed for all k (recall that ϕ_k are continuous and bounded functions). Thus,

$$\mathcal{B} = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\mu \in M_1(\mathbb{R}^d)\text{: } \int_{\mathbb{R}^{dk}} \phi_k \; d\mu^k \leq \tfrac{1}{3}\right\}$$

and

$$\mathcal{C} = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\mu \in M_1(\mathbb{R}^d)\text{: } \int_{\mathbb{R}^{dk}} \phi_k \; d\mu^k \geq \tfrac{2}{3}\right\}$$

are disjoint F_σ-sets in $M_1(\mathbb{R}^d)$.

Step 3. It only remains to verify that $H_0 \subset \mathcal{B}$ and $H_1 \subset \mathcal{C}$. Let $\mu \in H_0$ and let $X_1,X_2,X_3, \ldots$ be i.i.d. observations from μ. By discernibility,

$$f_n(X_1,X_2, \ldots,X_n) \to 0 \quad \text{a.s. as } n \to \infty.$$

Taking expectations using the bounded convergence theorem and utilizing Step 1, we conclude that

$$\int_{\mathbb{R}^{dn}} \phi_n(x_1, \ldots,x_n)\; d\mu^n \to 0 \quad \text{as } n \to \infty.$$

This shows that $\mu \in \mathcal{B}$; as $H_1 \subset \mathcal{C}$ is established similarly, the proof is complete. □

PROOF OF COROLLARY 2. (i) Assume that $\{H_i\}_{i\in I}$ are contained in disjoint F_σ-sets. Then so are $H_1^{(m)} = \bigcup_{i\leq m} H_i$ and $H_0^{(m)} = \bigcup_{i>m} H_i$, for any $m \in I$. By Theorem 2 the sets $H_0^{(m)}$ and $H_1^{(m)}$ are discernible (for any $m \in I$). Let $h_n^{(m)}\text{: } \mathbb{R}^{dn} \to \{0,1\}$ be the "discerning" functions corresponding to $H_0^{(m)}$ and $H_1^{(m)}$ and construct $f_n\text{: } \mathbb{R}^{dn} \to I$ as follows:

$$f_n(\mathbf{x}) = m \text{ iff } h_n^{(m)}(\mathbf{x}) = 1 \text{ and } h_n^{(i)}(\mathbf{x}) = 0, \text{ for all } i < m,$$

where $f_n(\mathbf{x})$ is arbitrary if $h_n^{(i)}(\mathbf{x}) = 0$ for all $i \in I$.

If $\mu \in H_m$, then $\mu \in \{\bigcap_{i<m} H_0^{(i)}\} \cap H_1^{(m)}$. Thus, if $X_1, X_2, X_3, \ldots$ are i.i.d. with law μ, then almost surely for all n large enough, $h_n^{(m)}(X_1, \ldots, X_n) = 1$ while $h_n^{(i)}(X_1, \ldots, X_n) = 0$, $i = 1, 2, \ldots, m-1$. Since this holds for all $m \in I$ the collection $\{H_i\}_{i \in I}$ is discernible.

(ii) Assume that $\{H_i\}_{i \in I}$ are discernible. Then definitely so are $H_0^{(m)}$ and $H_1^{(m)}$, for any $m \in I$ (simply construct $h_n^{(m)} = 1_{f_n \le m}$). By Theorem 2 the sets $H_0^{(m)}$ and $H_1^{(m)}$ are separated by disjoint F_σ-sets, say, $H_0^{(m)} \subset B_m$ and $H_1^{(m)} \subset C_m$. Now, $\{\bigcap_{i=1}^{m-1} B_i\} \cap C_m$, for $m \in I$, are disjoint F_σ-sets which separate $\{H_m\}_{m \in I}$. □

5. Extensions. So far we have considered only i.i.d. observations. However, the definition of discernibility extends naturally to two families of stochastic processes and the same general principle applies, as long as a universal rate of convergence ε_n can be established. For example, the following proposition is the analogue of Theorem 2 for finite-state Markov chains.

PROPOSITION 6. *Let $\mathcal{M} \subset \mathbb{R}^{d^2}$ be the set of all nonnegative, irreducible stochastic $d \times d$ matrices. With each $P \in \mathcal{M}$ associate the Markov chains $\{X_n\}$ over $\{1, \ldots, d\}$ whose transition matrix is P, where the initial state $X_0 \in \{1, \ldots, d\}$ is determined according to some arbitrary law. Then the hypotheses $H_0 = \{P \in A_0 \subset \mathcal{M}\}$ and $H_1 = \{P \in A_1 \subset \mathcal{M}\}$ are discernible iff A_0 and A_1 are contained in disjoint F_σ-subsets of $\mathcal{M}$.*

PROOF. (i) Suppose $A_0 \subset B$ and $A_1 \subset C$, where B and C are disjoint F_σ subsets of $\mathcal{M}$. For any $P \in \mathcal{M}$, there is a unique stationary distribution π which is strictly positive. The mapping $\{P_{ij}\} \to \{\pi_i P_{ij}\}$ is then an invertible closed mapping (i.e., the image of a closed subset of $\mathcal{M}$ is closed). By the law of the iterated logarithm for Markov chains (see [1], Sections 14–16), almost surely for any $0 < \alpha < \frac{1}{2}$ and sufficiently large n,

$$\sup_{i,j,x_0} \left| \frac{1}{n} \sum_{t=0}^{n-1} 1_{x_t = i,\ x_{t+1} = j} - \pi_i P_{ij} \right| < n^{-\alpha}.$$

Thus, the discernibility of H_j follows as in the proof of Theorem 1.

(ii) If H_0 and H_1 are discernible, then there exists $f_n \colon \mathbb{R}^n \to \{0, 1\}$ such that $A_0 \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} B_k$ and $A_1 \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} C_k$, where

$$B_k = \bigcap_{x_0} \{P \in \mathcal{M} \colon E_P[f_k(X_1, \ldots, X_k) | X_0 = x_0] \le \tfrac{1}{3}\},$$

$$C_k = \bigcap_{x_0} \{P \in \mathcal{M} \colon E_P[f_k(X_1, \ldots, X_k) | X_0 = x_0] \ge \tfrac{2}{3}\}.$$

Since $E_P[f_k(X_1, \ldots, X_k) | X_0 = x_0] = \sum_{\{\mathbf{x}:\, f_k(x_1, \ldots, x_k) = 1\}} \prod_{t=0}^{k-1} P_{x_t x_{t+1}}$ are continuous functions of $\{P_{ij}\}$, indeed B_k and C_k are (disjoint) closed sets, that is, A_0 and A_1 are contained in disjoint F_σ-subsets of $\mathcal{M}$. □

Similar ideas apply for Gaussian processes when the empirical covariances converge almost surely to the underlying covariance with a uniform rate.

Returning to the i.i.d. setup, note that in Theorem 2(ii) Lebesgue measure may be replaced by any reference measure. In this formulation the theorem holds with $\mathbb{R}^d$ replaced by any Polish space $\mathcal{E}$, since $M_1(\mathcal{E})$ is still a metric space with respect to convergence in law and Lusin's theorem is applicable (see [12]).

QUESTION. In Theorem 2(ii), can one remove the assumption that the densities of measures $\mu \in H_0 \cup H_1$ are in L^p, for some $p > 1$?

REFERENCES

[1] CHUNG, K. L. (1960). *Markov Chains with Stationary Transition Probabilities*. Springer, New York.

[2] COVER, T. (1973). On determining the irrationality of the mean of a random variable. *Ann. Statist.* **1** 862–871.

[3] DURRETT, R., KESTEN, H. and LAWLER, G. (1991). Making money from fair games. In *Random Walks, Brownian Motion and Interacting Particle Systems. A Festschrift in Honor of Frank Spitzer* (R. Durrett and H. Kesten, eds.) 255–267. Birkhäuser, Boston.

[4] FELLER, W. (1946). A limit theorem for random variables with infinite moments. *Amer. J. Math.* **68** 257–262.

[5] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[6] HOEFFDING, W. and WOLFOWITZ, J. (1958). Distinguishability of sets of distributions (the case of independent and identically distributed chance variables). *Ann. Math. Statist.* **29** 700–718.

[7] KULKARNI, S. R. and ZEITOUNI, O. (1991). Can one decide the type of the mean from the empirical measure? *Statist. Probab. Lett.* **12** 323–327.

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

[9] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[10] MAULDIN, R. D., PREISS, D. and WEIZSACKER, H. V. (1983). Orthogonal transition kernels. *Ann. Probab.* **11** 970–988.

[11] MAULDIN, R. D. and WEIZSACKER, H. V. (1991). Some orthogonality preserving kernels which are not completely orthogonal. *Ann. Probab.* **19** 396–400.

[12] NEVEU, J. (1964). *Bases Mathematiques du Calcul des Probabilités*. Masson, Paris.

[13] RUDIN, W. (1974). *Real And Complex Analysis*, 2nd ed. McGraw-Hill, New York.

[14] ZEITOUNI, O. and KULKARNI, S. R. (1990). A general classification rule for probability measures. Unpublished manuscript.

DEPARTMENT OF STATISTICS
AND DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720