Hypothesis Testing for Families of Ergodic Processes

Andrew B. Nobel
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260
Email: nobel@email.unc.edu

May 9, 2004

Abstract

General sufficient conditions for the discernibility of two families of stationary ergodic processes are derived. The conditions involve the weak topology for stationary processes. They are analogous in several respects to existing conditions for the discernibility of families of i.i.d. processes, but require a more refined type of topological separation in the general case. As a first application of the conditions, it is shown how existing discernibility results for i.i.d. processes may be extended to a countable union of uniformly ergodic families. In addition, it is shown how one may use hypothesis testing to study polynomial decay rates for covariance based mixing conditions.

Key words and phrases. Discernibility, ergodic processes, hypothesis testing, mixing conditions, process families.

1 Introduction

Hypothesis testing seeks to distinguish between two competing explanations for the observed behavior of some measured phenomena. Of interest here is a special case of the hypothesis testing problem: how to determine the membership of a sequence of observations in one of two known families of stationary stochastic processes. In general, one or both of the families will contain processes exhibiting long range dependence.

To be more specific, let H_0 and H_1 be disjoint families of stationary ergodic processes, and consider the following game between players A and B. To begin, player B is provided with complete information regarding the joint distributions of every process $\mathbf{X} \in H_0 \cup H_1$. Then player A selects a process $\mathbf{X} = \{X_i : i \geq 1\}$ from $H_0 \cup H_1$ and reveals its elements one by one, in order, to player B. At each time $n \geq 1$ player B is asked to determine the membership of $\mathbf{X}$ in H_0 or H_1, based only on the observed values of $X_1, \ldots, X_n$, and knowledge of the joint distributions of the processes in $H_0 \cup H_1$. Player B is successful if, for every $\mathbf{X} \in H_0 \cup H_1$, she makes only finitely many mistakes with probability one. When a successful strategy for player B exists, we say that H_0 and H_1 are discernible. A more precise definition of discernibility, and several variants, are given in Section 2.

In this paper general sufficient conditions for the discernibility of two families of stationary ergodic processes are derived. The conditions, given in Theorem 1, involve the weak topology on stationary processes. They are analogous in several respects to the conditions of Dembo and Peres (1994) for the discernibility of families of i.i.d. processes, but require a more refined type of topological separability in the general case. As an application of Theorem 1, it is shown that many existing discernibility results for i.i.d. processes can be generalized to weakly uniformly ergodic families. In addition, it is shown how hypothesis testing can be used to study polynomial decay rates for covariance based mixing conditions.

1.1 Outline

The next section contains technical preliminaries, several definitions of discernibility, and an account of previous work. The definitions of discernibility are briefly compared in Section 3. A number of preliminary results on uniformly ergodic families of processes are given in Section 4. Section 5 is devoted to the principal conclusion of the paper, Theorem 1, which gives general sufficient conditions for the almost sure discernibility of two families of ergodic processes. The final two sections of the paper are devoted to applications of Theorem 1. The discernibility of process families determined by finite-dimensional distributions is examined in Section 6. The study of polynomial mixing rates for a general covariance type mixing condition is considered Section 7. The appendix contains the proofs of several results stated in the text.

2 Preliminaries and Discernibility

2.1 Ergodic Processes

In what follows, we restrict our attention to real valued processes; the principal results of the paper may be extended to processes taking values in a complete separable metric space. Let $\mathcal{B}$ denote the Borel subsets of $\mathbb{R}$, and $\mathcal{B}^k$ denote the Borel subsets of $\mathbb{R}^k$. Let $\mathbb{R}^\infty$ be the set of all infinite sequences $\mathbf{x} = x_1, x_2, \ldots$ with $x_i \in \mathbb{R}$, and let $\mathcal{B}^\infty$ denote the usual product sigma field on $\mathbb{R}^\infty$, generated by the finite dimensional cylinder sets $\{A_1 \times \cdots \times A_k \times \mathbb{R} \times \mathbb{R} \cdots : A_l \in \mathcal{B},\, k \geq 1\}$. Each process $\mathbf{X} = X_1, X_2, \ldots \in \mathbb{R}$ will be identified with its induced distribution on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$; two processes with the same finite dimensional distributions will be considered identical. With this identification, define $\mathcal{M}$ to be the family of all stochastic processes $\mathbf{X} = X_1, X_2, \ldots \in \mathbb{R}$, and let $\mathcal{M}_s$ denote the sub-family of (strictly) stationary processes. A process $\mathbf{X} \in \mathcal{M}_s$ is said to be *ergodic* if it satisfies a weak, non-uniform mixing condition, namely for every $k \geq 1$ and every $A, B \in \mathcal{B}^k$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P\{X_1^k \in A, X_{i+1}^{i+k} \in B\} \;\to\; P\{X_1^k \in A\}\, P\{X_1^k \in B\},$$

where $X_1^k = (X_1, \ldots, X_k)$. This definition is equivalent to the standard one (*c.f.* Breiman (1992)) involving triviality of the invariant sigma-field. Let $\mathcal{E}$ be the family of all real-valued stationary ergodic processes.

2.2 Discernibility

Let $H_0, H_1 \subseteq \mathcal{E}$ be disjoint families of ergodic processes. As above, each process is identified with its induced distribution on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$. In particular, different processes may be defined on different underlying probability spaces.

Definition: A sequence of measurable functions $\varphi_n : \mathbb{R}^n \to [0,1]$, $n \geq 1$ will be called a testing scheme. A testing scheme is continuous if each of its constituent functions is continuous. Families H_0 and H_1 are *discernible* with probability one if there exists a testing scheme $\{\varphi_n : n \geq 1\}$ such that

$$\varphi_n(X_1^n) \;\to\; \begin{cases} 0 \text{ wp1} & \text{if } \mathbf{X} \in H_0 \\ 1 \text{ wp1} & \text{if } \mathbf{X} \in H_1. \end{cases} \tag{1}$$

Likewise, H_0 and H_1 are discernible in expectation if $E\varphi_n(X_1^n)$ converges to 0 when $\mathbf{X} \in H_0$, and converges to 1 when $\mathbf{X} \in H_1$. Families H_0 and H_1 are continuously discernible (in either of the above senses) if they are discernible by a continuous testing scheme.

Note that the definitions above place no requirements on the asymptotic behavior of $\varphi_n(X_1^n)$ if $\mathbf{X} \notin H_0 \cup H_1$. We consider test functions with values in the unit interval, rather than binary-valued decision functions, in order to distinguish between continuous and measurable testing schemes. One may readily convert a testing scheme into a decision scheme by thresholding its values, e.g., replacing $\varphi_n(X_1^n)$ by $I\{\varphi_n(X_1^n) > 1/2\}$.

2.3 Previous Work

To date, work on the discernibility of families of stochastic processes has primarily addressed the special case in which each candidate process consists of independent samples from a fixed distribution. In this case, the elements of H_0 and H_1 are fully described by their associated families of one-dimensional distributions, D_0 and D_1, respectively, and we will refer to the discernibility of D_0 and D_1, rather than H_0 and H_1.

Berger (1951) gave necessary and sufficient conditions for the existence of uniformly consistent tests of general families D_0 vs. D_1. Hoeffding and Wolfowitz (1958) gave sufficient, and in some cases necessary, conditions for the existence of tests for D_0 vs. D_1 under stopping criteria stronger than those considered here. Extending the results in Berger (1953), Le Cam and Schwartz (1960) gave necessary and sufficient conditions for the discernibility of D_0 and D_1 in the more general context of consistent point estimation. As noted by Dembo and Peres (1994), their conditions, which involve uniform structures on all the n-fold products of measures in $D_0 \cup D_1$, are not readily verifiable. Fisher and Van Ness (1969) studied the uniform discernibility of a countable family of probability measures in a sequential setting where one must make a "final" decision after a finite number of observations. In recent work, Devroye and Lugosi (2003) studied the problem of distinguishing a family $\mathcal{F}$ of univariate densities from its complement. They showed that some properties, like unimodality, are discernible, while others, like having compact support, are not.

Cover (1972) showed that for any countable family of means $S \subseteq \mathbb{R}$ there exists a (typically unknown) set S_0 of Lebesgue measure zero such that $D_0 = \{\mu : \int x^2\, d\mu < \infty, \int x\, d\mu \in S\}$ and $D_1 = \{\mathbf{X} : \int x^2\, d\mu < \infty, \int x\, d\mu \in S^c \setminus S_0\}$ are discernible. Kulkarni and Zeitouni (1995) looked at the more general problem of discerning D_0 from $D_1 = D_0^c$, for suitable families D_0, when one allows the testing procedure to fail on a set of distributions that is negligible under a given prior distribution. Some related results for discrete time, finite state Markov chains are given in Zeitouni and Kulkarni (1994).

This paper is motivated in part by the work of Dembo and Peres (1994), who provided topological criteria for the discernibility in the i.i.d. and finite state Markov settings. Let $\mathcal{M}(\mathbb{R})$ be the family of probability measures on $(\mathbb{R}, \mathcal{B})$, equipped with the topology of weak convergence. Recall that a family $D \subseteq \mathcal{M}(\mathbb{R})$ is said to be an F_σ if it is a countable union of closed sets. The following result is given in Theorem 2 of Dembo and Peres (1994). An extension of the sufficiency part of the theorem to ergodic processes is given in Theorem 1.

Theorem A *Two families $D_0, D_1 \subseteq \mathcal{M}(\mathbb{R})$ are discernible if they are contained in disjoint F_σ's. The converse is true if every distribution $\mu \in D_0 \cup D_1$ has a density f_μ with respect to Lebesgue measure such that $\int f_\mu^p \, dx < \infty$ for some $p > 1$ that may depend on μ.*

Concerning families of non-independent processes, Kraft (1955) studied consistent tests for families $H_0 = \{P_\theta : \theta \in \Theta\}$ and $H_1 = \{Q_\lambda : \lambda \in \Lambda\}$ of general distributions on $\mathbb{R}^\infty$, a setting that includes dependent, possibly non-stationary, processes. He established that H_0 and H_1 are uniformly discernible in expectation if and only if the n-dimensional projections of their convex hulls are, in a suitable sense, asymptotically orthogonal. In addition, if there exist probability measures α and β on Θ and Λ, respectively, such that the composite distributions $\int P_\theta \, d\alpha(\theta)$ and $\int Q_\lambda \, d\beta(\lambda)$ are mutually singular, then there is a testing scheme consistent in expectation for α-almost every P_θ and β-almost every Q_λ. From this latter result follows an interesting corollary, which also appears, with different proofs, in Barron (1985) and Adams and Nobel (1998).

Lemma A *If H_0 and H_1 are countable and disjoint then they are discernible.*

Ornstein and Weiss (1990) describe an estimation scheme that, given any Bernoulli process $\mathbf{Y} = Y_1, Y_2, \ldots \in \mathbb{R}$, produces a sequence of *processes* $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ such that $\mathbf{Z}_k$ is constructed only from knowledge $Y_1, \ldots, Y_k$, and $\mathbf{Z}_k$ converges in the $\overline{d}$ distance to $\mathbf{Y}$. They also showed that no estimation scheme is $\overline{d}$ consistent for the larger family of K automorphisms. Some extensions of this work can be found in Ornstein and Shields (1994). In recent work, Lim (2002) considers several problems related to discernibility of dependent processes, and detecting non-stationarity in a process of independent coin flips, where the goal is to detect the repeated but infrequent use of a biased coin. He also establishes an analog of Theorem A for m-dependent processes and families with uniformly convergent empirical measures, which is weaker than Theorem 1 below.

3 Forms of Discernibility

Here the different forms of discernibility defined in the previous section are briefly compared. It is clear that continuous discernibility of either type implies the corresponding form of measurable discernibility. Likewise, discernibility with probability one implies discernibility in expectation. As a partial converse, if each process $\mathbf{X} \in H_0 \cup H_1$ is i.i.d. then measurable/continuous discernibility in expectation implies measurable/continuous discernibility with probability one. To see this, let $\{\varphi_n\}$ be a testing scheme discerning H_0 and H_1 in expectation. By an application of Hoeffding's inequality for sums of bounded independent random variables, the testing scheme $\psi_n(X_1^n) = m^{-1} \sum_{j=0}^{m-1} \varphi_k(X_{kj+1}^{kj+k})$ satisfies (1) if $k = n \,/\, 2 \log n$ and $m = \lfloor n/k \rfloor$. Moreover $\{\psi_n\}$ is continuous if $\{\varphi_n\}$ is. A similar argument applies if, for example, the processes in $H_0 \cup H_1$ are α-mixing and share a common mixing rate. On the other hand, the following example shows that continuous discernibility is a stronger notion than discernibility with probability one (or in probability), even for families of i.i.d. processes.

Example: Let H_0 consist of all i.i.d. processes with values in $[0, 1]$ having an absolutely continuous marginal distribution, and let H_1 consist of all i.i.d. processes whose marginal distribution has at least one point mass. The measurable test functions $\psi_n(X_1^n) = 1 - \prod_{i \neq j} I\{X_i \neq X_j\}$ readily discern H_0 from H_1 almost surely. We claim that no continuous testing scheme can distinguish between H_0 and H_1. To see this, let $\{\varphi_n\}$ be any continuous testing scheme, and assume without loss of generality that $E\varphi_n(X_1^n) \to 1$ for every $\mathbf{X} \in H_1$, for otherwise $\{\varphi_n\}$ fails to discern H_0, H_1. Under this assumption, we exhibit a process $\mathbf{X} \in H_0$ such that $E\varphi_n(X_1^n) \not\to 0$. In what follows $\delta_{\{a\}}$ denotes a point mass at a, $U[a, b]$ denotes the uniform distribution on $[a, b]$, and $\mathrm{Bin}(n, p)$ denotes a binomial distribution with parameters $n \geq 1$ and $p \in [0, 1]$.

Let $\mu_0 = \delta_{\{1\}}$, and let $\mathbf{Y} = Y_1, Y_2, \ldots$ be i.i.d. with $Y_i \sim \mu_0$, so that $\mathbf{Y} \in H_1$. By assumption there is an integer n_0 for which $P\{\varphi_{n_0}(Y_1^{n_0}) \geq 3/4\} \geq 3/4$. As φ_{n_0} is continuous, there exists $0 < \epsilon_0 < 1/2$ such that $|\varphi_{n_0}(x_1^{n_0}) - \varphi_{n_0}(y_1^{n_0})| < 1/4$ whenever $|x_i - y_i| < \epsilon_0$ and $x_i, y_i \in [0, 1]$ for $i = 1, \ldots, n_0$. Let $0 < p_0 < 1$ be so large that $P\{\mathrm{Bin}(n_0, 1 - p_0) = 0\} \geq 3/4$.

We now proceed recursively as follows. Suppose that $k \geq 1$, and that we have specified integers $n_0 < n_1 < \cdots < n_{k-1}$, positive constants $\epsilon_0, \ldots, \epsilon_{k-1}$ satisfying $\epsilon_j < 2^{-(j+1)}$, and probabilities $0 < p_0, \ldots, p_{k-1} < 1$. Let $\beta_i = \prod_{r=0}^{i}(1 - p_r)$ for $0 \leq i \leq k - 1$, and set $\beta_{-1} = 1$. Define the distribution $\mu_k = \sum_{i=0}^{k-1} p_i \, \beta_{i-1} \cdot U[2^{-i} - \epsilon_i, 2^{-i}] + \beta_{k-1} \, \delta_{\{2^{-k}\}}$. Let $\mathbf{Y} = Y_1, Y_2, \ldots$ be i.i.d. with $Y_i \sim \mu_k$, so that $\mathbf{Y} \in H_1$. By assumption there exists

$n_k > n_{k-1}$ such that $P\{\varphi_{n_k}(Y_1^{n_k}) \geq 3/4\} \geq 3/4$. Moreover, as φ_{n_k} is continuous there exists $0 < \epsilon_k < 2^{-(k+1)}$ such that $|\varphi_{n_k}(x_1^{n_k}) - \varphi_{n_k}(y_1^{n_k})| < 1/4$ whenever $|x_i - y_i| < \epsilon_k$ and $x_i, y_i \in [0,1]$ for $i = 1, \ldots, n_k$. Choose $0 < p_k < 1$ so that $P\{\text{Bin}(n_k, \beta_k) = 0\} \geq 3/4$, and proceed to stage $k+1$.

The definition of $\{p_k\}$ ensures that $\sum_{k \geq 0} p_k \, \beta_{k-1} = 1$. Define the absolutely continuous distribution $\mu = \sum_{k \geq 0} p_k \, \beta_{k-1} \cdot U[2^{-k} - \epsilon_k, 2^{-k}]$ and let $\mathbf{X} = X_1, X_2, \ldots$ be i.i.d. with $X_i \sim \mu$, so that $\mathbf{X} \in H_0$. For each $i, k \geq 1$ define $Y_{i,k} = X_i$ if $X_i > 2^{-k}$ and $Y_{i,k} = 2^{-k}$ otherwise. Then $\mathbf{Y}_k = Y_{1,k}, Y_{2,k}, \ldots$ is i.i.d. with distribution μ_k, and as $P\{X_i \leq 2^{-(k+1)}\} = \beta_k$,

$$P\{|X_i - Y_{i,k}| < \epsilon_k \text{ for } i = 1, \ldots, n_k\} \;\; \geq \;\; P\{\text{Bin}(n_k, \beta_k) = 0\} \; \geq \; 3/4.$$

For each $k \geq 1$, $P\{\varphi_{n_k}(X_1, \ldots, X_{n_k}) \geq 1/2\}$ is at least

$$\begin{aligned} & P\{\varphi_{n_k}(Y_{1,k}, \ldots, Y_{n_k,k}) \geq 3/4 \;\text{ and }\; |Y_{i,k} - X_i| < \epsilon_k \text{ for } i = 1, \ldots, n_k\} \\ & \quad \geq \;\; P\{\varphi_{n_k}(Y_{1,k}, \ldots, Y_{n_k,k}) \geq 3/4\} \; - \; P\{|Y_{i,k} - X_i| \geq \epsilon_k \text{ some } i = 1, \ldots, n_k\} \end{aligned}$$

which is lower bounded by 1/2. Thus $E\varphi_n(X_1^n)$ does not converge to zero as $n \to \infty$, and as $\{\varphi_n : n \geq 1\}$ was an arbitrary continuous testing scheme, the claim follows.

4 Preliminary Results

4.1 Weak Convergence

We will make use of some basic facts from the theory of weak convergence. For more details and proofs, see Dudley (1989) or Billingsley (1999). Let $\mathcal{C}_k = C_b(\mathbb{R}^k)$ be the family of bounded continuous functions $g : \mathbb{R}^k \to \mathbb{R}$. Recall that a sequence of random vectors $U_1, U_2, \ldots \in \mathbb{R}^k$ (possibly defined on different probability spaces) is said to converge weakly to a random vector $U \in \mathbb{R}^k$, written $U_n \Rightarrow U$, if $Eg(U_n) \to Eg(U)$ as for each $g \in \mathcal{C}_k$. A sequence of processes $\mathbf{X}^1, \mathbf{X}^2, \ldots \in \mathcal{M}_s$ converges weakly to $\mathbf{X} \in \mathcal{M}_s$ if $(X_1^n, \ldots, X_k^n) \Rightarrow (X_1, \ldots, X_k)$ for every $k \geq 1$. Weak convergence defines a metrizable topology on $\mathcal{M}$ with basic open sets $\{\mathbf{X} : |Eh(X_1^k) - a| < \epsilon\}$, with $k \geq 1$, $h \in \mathcal{C}_k, a \in \mathbb{R}$, $\epsilon > 0$. $\mathcal{M}_s$ is a closed subset of $\mathcal{M}$ in the weak topology, but that the family of ergodic processes $\mathcal{E}$ is not.

Recall that a family of processes $\mathcal{S}$ is (uniformly) tight if for every $\epsilon > 0$ there is a compact set $C \subseteq \mathbb{R}^\infty$ such that $\sup_{\mathbf{X} \in \mathcal{S}} P\{\mathbf{X} \in C^c\} \leq \epsilon$. A family $\mathcal{S}$ of stationary processes is tight if and only if its one-dimensional distributions are tight, i.e., $\sup_{\mathbf{X} \in \mathcal{S}} P\{|X_1| > c\} \to 0$ as $c \to \infty$. Prohorov's theorem states that every sequence $\{\mathbf{X}^n : n \geq 1\}$ of processes in a tight family $\mathcal{S}$ has a weakly convergent subsequence. In particular, a family of processes that is closed and tight is compact, and every compact family is tight.

Definition: Let $\mathbf{X} \in \mathcal{M}_s$ be any stationary stochastic process. For each $n, k \geq 1$ and each function $f : \mathbb{R}^k \to \mathbb{R}$ such that $Ef(X_1^k)^2 < \infty$, define $V_n(\mathbf{X}, f) \; = \; \mathrm{Var}\left(n^{-1} \sum_{i=0}^{n-1} f(X_{i+1}^{i+k}) \right)$.

Lemma 1 *If g is a bounded continuous function of k variables, then $V_n(\cdot, g)$ is continuous in the weak topology on $\mathcal{M}_s$,* i.e., *$\mathbf{X}_s \Rightarrow \mathbf{X}$ implies $V_n(\mathbf{X}_s, g) \to V_n(\mathbf{X}, g)$.*

Proof: Let $Y = n^{-1} \sum_{i=0}^{n-1} g(X_{i+1}^{i+k}) - Eg(X_1^k)$, and define Y_s in the same way with elements of $\mathbf{X}_s$ appearing in the average. Let $\theta_s = Eg(X_{1,s}, \ldots, X_{k,s}) - Eg(X_1^k)$. Then $V_n(\mathbf{X}, g) = EY^2$ and $V_n(\mathbf{X}_s, g) = E(Y_s + \theta_s)^2$. If $\mathbf{X}_s \Rightarrow \mathbf{X}$ then the continuous mapping theorem and Slutsky's theorem together imply that $(Y_s + \theta_s)^2 \Rightarrow Y^2$, and the desired convergence follows as Y_s and θ_s are bounded.

The proofs of the following Lemmas are elementary, and are omitted.

Lemma 2 *Let $\mathbf{X} \in \mathcal{M}_s$ be any stationary process, and let f, h be bounded measurable functions of k and l variables, respectively. Then*

$$|V_n^{1/2}(\mathbf{X}, f) - V_n^{1/2}(\mathbf{X}, h)| \;\leq\; E|f(X_1^k) - h(X_1^l)| + (E|f(X_1^k) - h(X_1^l)|^2)^{1/2}.$$

Moreover, the numerical sequence $a_n = \mathrm{SD}(\sum_{i=0}^{n-1} f(X_{i+1}^{i+k}))$, $n \geq 1$, is subadditive, i.e., *$a_{n+m} \leq a_n + a_m$ for every $n, m \geq 1$.*

Lemma 3 *A stationary process $\mathbf{X}$ is ergodic if and only if $V_n(\mathbf{X}, g) \to 0$ for every $g \in \cup_{k \geq 1} \mathcal{C}_k$.*

4.2 Uniform Ergodicity

Existing discernibility results for families of i.i.d. processes are established by appeals to fundamental inequalities describing how rapidly averages of independent random variable converge to their expectations. The ergodic theorem extends the law of large numbers to ergodic processes, but it does not (and can not) guarantee any rate of convergence. One central conclusion of this paper is that, for discernibility, identifiable rates of convergence are unnecessary. It is sufficient that, for each bounded continuous g, the variances $V_n(\mathbf{X}, g)$ decay uniformly over the family of processes $H_0 \cup H_1$.

Definition: A family $\mathcal{S}$ of stationary processes is *uniformly ergodic* if $\sup_{\mathbf{X} \in \mathcal{S}} V_n(\mathbf{X}, g) \to 0$ as $n \to \infty$ for every function $g \in \cup_{k \geq 1} \mathcal{C}_k$. Note that the uniformity is over processes $\mathbf{X} \in \mathcal{S}$, not the functions $g \in \cup_{k \geq 1} \mathcal{C}_k$. The rate of decay of the supremum can depend on g.

Example 1 (I.I.D. Processes): If $\mathbf{X}$ is i.i.d., a routine argument shows that $V_n(\mathbf{X}, g)$ is at most $2(n+k)k/n^2$ times the maximum absolute value of g. Thus every family of i.i.d. processes is uniformly ergodic.

Example 2 (Uniform Cesaro Mixing): A straightforward calculation shows that a family $\mathcal{S}$ of stationary processes is uniformly ergodic if for every $k \geq 1$, and every $A, B \in \mathcal{B}^k$,

$$\sup_{\mathbf{X} \in \mathcal{S}} \left| \frac{1}{n} \sum_{j=1}^{n} P(X_1^k \in A, X_{j+1}^{j+k} \in B) \,-\, P(X_1^k \in A)\, P(X_1^k \in B) \right| \;\to\; 0.$$

Uniform ergodicity is not equivalent to the notion of a uniformly ergodic transformation (*c.f.* Krengel (1985)), in which the supremum above is taken over events $A, B \in \cup_k \mathcal{B}^k$.

Example 3 (Uniform α-mixing): In many statistical applications, stronger mixing conditions than those in Example 2 are more natural. Consider a two-sided stationary process $\mathbf{X} = \{X_i : -\infty < i < \infty\}$. Let $\mathbf{X}_0^- = X_0, X_{-1}, X_{-2}, \ldots$ denote the "past" of the process, starting from time zero, and let $\mathbf{X}_k^+ = X_k, X_{k+1}, X_{k+2}, \ldots$ denote the "future" of the process starting from some time $k \geq 1$. The α-mixing coefficients of $\mathbf{X}$, introduced by Rosenblatt (1956), are defined by

$$\alpha(k : \mathbf{X}) \;=\; \sup_{A,B \in \mathcal{B}^\infty} |P(\mathbf{X}_0^- \in A, \mathbf{X}_k^+ \in B) \,-\, P(\mathbf{X}_0^- \in A)\, P(\mathbf{X}_1^+ \in B)| \quad k \geq 1. \tag{2}$$

The process $\mathbf{X}$ is said to be α-mixing if $\alpha(k : \mathbf{X}) \to 0$ as k tends to infinity. It follows from Example 2 that a family $\mathcal{S}$ of stationary α-mixing processes is uniformly ergodic if there exist non-negative constants $a_1, a_2, \ldots \to 0$ such that $\alpha(k : \mathbf{X}) \leq a_k$ for each $k \geq 1$ and each $\mathbf{X} \in \mathcal{S}$. In particular, a uniformly α-mixing family is also uniformly ergodic. Analogous conclusions hold for stronger mixing conditions (e.g. those defined in terms of β- and ϕ-mixing coefficients).

Example 4 (Moving Average Processes): Let $\mathcal{S} = \{\mathbf{X}_\lambda : \lambda \in \Lambda\}$ be a family of stationary processes such that each $\mathbf{X}_\lambda$ has an infinite moving average representation of the form $X_t^\lambda = \sum_{j \geq 0} a_{j,\lambda} Z_{t-j,\lambda}$, $-\infty < t < \infty$, where $\{Z_{i,\lambda} : -\infty < i < \infty\}$ are i.i.d. random variables with mean zero, and $a_{j,\lambda}$ are real valued constants. One may show that $\mathcal{S}$ is uniformly ergodic if $\lim_{k \to \infty} \sup_{\lambda \in \Lambda} \sum_{j \geq k} |a_{j,\lambda}| = 0$ and $\sup_{\lambda \in \Lambda} \sup_{i \geq 1} E Z_{i,\lambda}^2 < \infty$.

Proposition 1 *If $\mathcal{S} \subseteq \mathcal{E}$ is a compact subset of $\mathcal{M}_s$, then $\mathcal{S}$ is uniformly ergodic.*

Proof: Fix $g \in \mathcal{C}_k$ and define continuous functions $\beta_n : \mathcal{S} \to \mathbb{R}$ by $\beta_n(\mathbf{X}) = V_n^{1/2}(\mathbf{X}, g)$. Lemma 3 ensures that $\beta_n(\mathbf{X}) \to 0$ for each $X \in \mathcal{S}$ since $\mathcal{S} \subseteq \mathcal{E}$, and Lemma 2 implies that

$\beta_{2^{n+1}}(\cdot) \leq \beta_{2^n}(\cdot)$ for $n \geq 1$. Thus $\{\beta_{2^n}(\cdot) : n \geq 1\}$ are continuous functions converging monotonically to zero on a compact set, and it follows that β_{2^n} converges uniformly to zero on $\mathcal{S}$. Define $\alpha_n = \sup_{\mathbf{X}\in\mathcal{S}} \mathrm{SD}(\sum_{i=0}^{n-1} g(X_{i+1}^{i+k}))$. By Lemma 2, the sequence $\{\alpha_n\}$ is subadditive, and therefore

$$0 \leq \lim_{n\to\infty} \sup_{\mathbf{X}\in\mathcal{S}} \beta_n(\mathbf{X}) = \lim_{n\to\infty} \frac{\alpha_n}{n} = \inf_{n\geq 1} \frac{\alpha_n}{n} \leq \inf_{m\geq 1} \sup_{\mathbf{X}\in\mathcal{S}} \beta_{2^m}(\mathbf{X}) = 0.$$

A proof of the following proposition is given in Appendix 8.1.

Proposition 2 *If $\mathcal{S}$ is uniformly ergodic then $\overline{\mathcal{S}}$ is uniformly ergodic, and in particular $\overline{\mathcal{S}} \subseteq \mathcal{E}$. If $\mathcal{S}$ is tight and fails to be uniformly ergodic, then $\overline{\mathcal{S}} \not\subseteq \mathcal{E}$.*

5 Sufficient Conditions for Discernibility

The next theorem is the principal result of the paper. It provides a topological criterion for the discernibility of two ergodic families H_0 and H_1. It may be viewed as an extension of Theorem A, which gives a topological criterion for discernibility in the i.i.d. setting. The conditions of Theorem 1 are evidently satisfied when H_0 and H_1 are countable, and we again recover Lemma A. Several applications of Theorem 1 are considered in Sections 6 and 7 below. The proof is given in Appendix 8.2.

Theorem 1 *Two families, H_0 and H_1, of stationary ergodic processes are continuously discernible if the following two conditions are satisfied.*

(i) $H_0 \cup H_1$ is contained in a countable union of uniformly tight subsets of $\mathcal{M}_s$.

(ii) There exist families $U_1, U_2, \ldots, V_1, V_2, \ldots \subseteq \mathcal{M}_s$ such that

- *a. Each U_i, V_j is contained in $\mathcal{E}$ and closed in $\mathcal{M}_s$;*
- *b. $H_0 \subseteq U = \bigcup_{i\geq 1} U_i$ and $H_1 \subseteq V = \bigcup_{i\geq 1} V_i$;*
- *c. $U \cap V = \emptyset$.*

Condition (i) is equivalent to the condition that $H_0 \cup H_1$ is contained in a sigma-compact subset of $\mathcal{M}_s$. It is satisfied if, for example, there is an increasing function $\gamma(\cdot) \geq 0$, with $\gamma(u) \to \infty$ as $u \to \infty$, such that $E\gamma(|X_1|) < \infty$ for each $\mathbf{X} \in H_0 \cup H_1$. Condition (ii) specifies a form of topological separation that is sufficient to ensure discernibility of two hypotheses. It is important to note that the sets U_i and V_i are assumed to be *simultaneously* contained in $\mathcal{E}$ and closed in $\mathcal{M}_s$. The weaker assumption that U_i and V_i are the intersection of $\mathcal{E}$ with a closed subset of $\mathcal{M}_s$ is not sufficient to guarantee discernibility (see Section 5.1 below). With this distinction in mind, call a set $U \subseteq \mathcal{M}_s$ an E_σ if it is equal to a countable union of sets $U_i \subseteq \mathcal{E}$ that are closed in $\mathcal{M}_s$. Then condition (ii) requires that H_0 and H_1 be contained in disjoint E_σ's.

5.1 A Negative Example

If two families H_0 and H_1 are indiscernible, then any sequential scheme that seeks to distinguish between them must fail. We briefly describe an example of two families of ergodic processes that are not discernible by any measurable testing scheme.

Example: Let H_0^* be the family of stationary ergodic processes $\mathbf{X}$ with values in $[0,1]$ and marginal density $f_0 = 1$. Let H_1^* be the family of stationary ergodic processes $\mathbf{X}$ with values in $[0,1]$ and having marginal densities of the form

$$f_{i,k}(x) \;=\; \sum_{j=0}^{2^k-1} 2 \cdot I\left\{2^{-k}\,(2j+i)\;\leq x <\; 2^{-k}\,(2j+i+1)\right\} \quad i=0,1,\; k \geq 1.$$

Each $f_{i,k}$ is a Rademacher density, taking the value 2 on the odd ($i=1$) or even ($i=0$) dyadic intervals of order k, and zero elsewhere. By a cutting and stacking argument like that in Adams and Nobel (1998), one may show that the families H_0^* and H_1^* are not discernible in expectation (or almost surely) by any measurable testing scheme. The families H_0^* and H_1^* evidently satisfy condition (i) of Theorem 1. For $a \in (0,1)$ define $\mathcal{S}(a)$ to be the family of stationary ergodic processes $\mathbf{X}$ with values in $[0,1]$ such that $EX_1 = a$. Then $H_0^* \subseteq U = \mathcal{S}(1/2)$, while $H_1^* \subseteq V = \bigcup_{j\geq 2} \mathcal{S}(1/2+2^{-j}) \cup \mathcal{S}(1/2+2^{-j})$, and clearly $U \cap V = \emptyset$. The family $\mathcal{S}(a)$ is the intersection of $\mathcal{E}$ with the closed set of stationary processes having mean a, but this intersection is *not* a closed subset of $\mathcal{M}_s$, so the topological condition (ii) of Theorem 1 is violated.

5.2 Infinitely Many Hypotheses

Here we consider the problem of distinguishing between infinitely many families H_0, H_1, $H_2, \ldots$ of ergodic processes. A collection $\{H_i : i \geq 0\}$ of families $H_i \subseteq \mathcal{E}$ is continuously discernible (with probability one) if there exist continuous functions $\varphi_n : \mathbb{R}^n \to [0,\infty)$, $n \geq 1$, such that for each $i \geq 0$ and each $\mathbf{X} \in H_i$, $\varphi_n(X_1^n) \to i$ with probability one.

Theorem 2 *Families H_i, $i \geq 0$ of ergodic processes are continuously discernible if the following two conditions hold: (i) $\cup_{i\geq 0} H_i$ is contained in a countable union of uniformly* *tight sets; (ii) there exist families $\{U_{i,j} : i, j \geq 1\}$ of ergodic processes, each of which is closed in $\mathcal{M}_s$, such that $H_i \subseteq W_i = \bigcup_{j\geq 1} U_{i,j}$, and $W_i \cap W_j = \emptyset$ if $i \neq j$.*

Proof: For each $k \geq 1$ the composite hypotheses $\tilde{H}_0^k = H_0 \cup \cdots \cup H_k$ and $\tilde{H}_1^k = H_{k+1} \cup H_{k+2} \cup \cdots$ satisfy the conditions of Theorem 1. Therefore, there exist continuous maps $\{\varphi_{n,k} : n \geq 1, k \geq 0\}$ such that for every $\mathbf{X} \in \cup_{i\geq 0} H_i$, and every $k \geq 0$,

$$\lim_{n\to\infty} \varphi_{n,k}(X_1^n) \;=\; \begin{cases} 0 \text{ wp1} & \text{if } \mathbf{X} \in \cup_{i=0}^k H_i \\ 1 \text{ wp1} & \text{if } \mathbf{X} \in \cup_{i>k} H_i \end{cases}$$

Let $g(u) = ((2u - 1/4) \wedge 0) \vee 1$. Given observations $X_1, \ldots, X_n$ from some process $\mathbf{X} \in \cup_{i\geq 1} H_i$, define $\varphi_n(X_1^n) = 1 + \sum_{j=1}^n g\left(\min_{1\leq l\leq j} \varphi_{n,l}(X_1^n)\right)$. It is easy to verify that φ_n is continuous, and that $\varphi_n(X_1^n) \to i$ with probability one if $\mathbf{X} \in H_i$.

6 Finite-Dimensional Families

As a first application of Theorem 1, we show how it may be applied to obtain extensions of several existing discernibility results for i.i.d. processes. These extensions illustrate a more general principle: when membership in H_0 and H_1 depends on the k-dimensional distribution of $\mathbf{X}$, we may replace independence by the more general assumption of weak uniform ergodicity. Formally, a family of processes $\mathcal{S}$ is *weakly uniformly ergodic* if it is contained in (or equal to) a countable union of uniformly ergodic families.

6.1 Testing Means

Let $\mathcal{S}$ be a weakly uniformly ergodic family. Then $\mathcal{S} \subseteq \cup_{i\geq 1}\mathcal{S}_i$, where $\mathcal{S}_i$ is uniformly ergodic. By Proposition 2 we can assume, without loss of generality, that each family $\mathcal{S}_i$ is also closed in $\mathcal{M}_s$. Let $\{A_i : i \geq 1\}$ and $\{B_j : j \geq 1\}$ be closed subsets of $\mathbb{R}$. For $l \geq 1$ let $\mathcal{S}_l' = \cup_{i=1}^l \mathcal{S}_i$, $A_l' = \cup_{i=1}^l A_i$, and $B_l' = \cup_{i=1}^l B_i$. Define $U_l = \{\mathbf{X} \in \mathcal{S}_l' : ||X||_{1+l^{-1}} \leq l,\, EX \in A_l'\}$, and define V_l similarly, but with the requirement that $EX \in B_l'$. Here $||X||_p = (E|X|^p)^{1/p}$ denotes the usual L_p-norm of X. Note that U_l and V_l are contained in $\mathcal{E}$ and closed in $\mathcal{M}_s$. Consider the hypotheses

$$\begin{aligned} H_0 &= \{\mathbf{X} \in \mathcal{S} : E|X|^p < \infty \text{ for some } p > 1,\, EX \in \cup_{i\geq 1} A_i\} \subseteq \cup_{l\geq 1} U_l, \\ H_1 &= \{\mathbf{X} \in \mathcal{S} : E|X|^p < \infty \text{ for some } p > 1,\, EX \in \cup_{i\geq 1} B_i\} \subseteq \cup_{l\geq 1} V_l. \end{aligned}$$

The following corollary of Theorem 1 extends the sufficiency part of Theorem 1 of Dembo and Peres (1994) to uniformly ergodic families. An analogous result holds for discerning multivariate means.

Corollary 1 *If* $\cup_i A_i$ *and* $\cup_j B_j$ *are disjoint, then* H_0 *and* H_1 *are discernible.*

6.2 Testing Marginal Distributions

Let $\mathcal{M}(\mathbb{R})$ be the set of probability measures on $(\mathbb{R}, \mathcal{B})$, equipped with the topology of weak convergence, and let $\mathcal{L}(X) \in \mathcal{M}(\mathbb{R})$ denote the distribution of a real-valued random variable X. Let $\{C_i : i \geq 1\}$ and $\{D_j : j \geq 1\}$ be closed subsets of $\mathcal{M}(\mathbb{R})$, and let let $\mathcal{S}$ be a weakly uniformly ergodic family. Then $\mathcal{S} \subseteq \cup_{i \geq 1} \mathcal{S}_i$, where $\mathcal{S}_i$ is uniformly ergodic, and by Proposition 2 we can assume, without loss of generality, that each family $\mathcal{S}_i$ is also closed in $\mathcal{M}_s$. For $l \geq 1$ let $\mathcal{S}'_l = \cup_{i=1}^{l} \mathcal{S}_i$, $C'_l = \cup_{i=1}^{l} C_i$, and $D'_l = \cup_{i=1}^{l} D_i$. Define $U_l = \{\mathbf{X} \in \mathcal{S}'_l : \mathcal{L}(X) \in C'_l\}$ and $V_l = \{\mathbf{X} \in \mathcal{S}'_l : \mathcal{L}(X) \in D'_l\}$. Note that U_l and V_l are contained in $\mathcal{E}$ and closed in $\mathcal{M}_s$. Let $\gamma \geq 0$ be an increasing function such that $\gamma(u) \to \infty$ as $u \to \infty$. Consider the hypotheses

$$\begin{aligned} H_0 &= \{\mathbf{X} \in \mathcal{S} : E\gamma(|X|) < \infty,\ \mathcal{L}(X) \in \cup_{i \geq 1} C_i\} \subseteq \cup_{l \geq 1} U_l. \\ H_1 &= \{\mathbf{X} \in \mathcal{S} : E\gamma(|X|) < \infty,\ \mathcal{L}(X) \in \cup_{i \geq 1} D_i\} \subseteq \cup_{l \geq 1} V_l. \end{aligned}$$

The following corollary of Theorem 1 extends the sufficiency part of Theorem A to uniformly ergodic families, under the additional assumption that $E\gamma(|X|)$ is finite. An analogous result holds for discerning higher-order distributions.

Corollary 2 *If* $\cup_i C_i$ *and* $\cup_j D_j$ *are disjoint, then* H_0 *and* H_1 *are discernible.*

6.3 Testing Unimodality

A probability density f on $\mathbb{R}$ is unimodal if there exists $x_0 \in \mathbb{R}$ such that f is non-decreasing on $(-\infty, x_0]$ and non-increasing on $[x_0, \infty)$. Let F_0 be the family of unimodal densities f such that $\int f^p dx < \infty$ for some $p > 1$, and let F_1 be the family of non-unimodal densities f such that $\int f^p dx < \infty$ for some $p > 1$. Devroye and Lugosi (2003) show that F_0 and F_1 are discernible from i.i.d. samples. It then follows from the necessity part of Theorem A that F_0 and F_1 are contained in disjoint F_σ's in $\mathcal{M}(\mathbb{R})$. Let $\mathcal{S}$ be a weakly uniformly ergodic family. By virtue of the arguments in the previous example, for every increasing function γ such that $\gamma(u) \to \infty$ as $u \to \infty$, the hypotheses

$$H_0 = \{\mathbf{X} \in \mathcal{S} : E\gamma(|X|) < \infty,\, X \sim f \in F_0\}, \quad H_1 = \{\mathbf{X} \in \mathcal{S} : E\gamma(|X|) < \infty,\, X \sim f \in F_1\}$$

are discernible. One may extend the other positive results of Devroye and Lugosi (2003) in a similar manner.

7 Estimating Mixing Rates

In this section it is shown how Theorem 1 can be used to study, and in some cases estimate, polynomial mixing rates for dependent processes. We consider two-sided processes $\mathbf{X} = \{X_i : -\infty < i < \infty\}$. The other results above hold without change in this setting. Let $\mathcal{M}_s$ denote the family of two-sided stationary processes and let $\mathcal{E}$ denote the sub-family of ergodic processes.

All ergodic processes exhibit a weak form of asymptotic independence. In the statistical literature, stronger forms of asymptotic independence (often referred to as mixing conditions) have received a great deal of attention. Mixing conditions quantify how the past and future of a process become independent as the gap between them grows, and mixing rates quantify how fast the limiting independence takes place. Many asymptotic results for i.i.d. processes, *e.g.*, central limit theorems and convergence rates for density and regression estimation, carry over to weakly dependent processes under suitable mixing conditions and rate assumptions. However, it is usually difficult to verify empirically when specific mixing conditions and rate assumptions hold. A good account of several popular mixing conditions, and the relations between them, can be found in Bradley (1986).

Here we take a preliminary look, in the context of hypothesis testing, at the problem of assessing polynomial decay rates for covariance based mixing coefficients. It is shown in Lemma 4 that one may distinguish between suitably separated polynomial rate regimes, and Theorem 3 applies this result to the problem of rate assessment.

Let the sequence space $\mathbb{R}^\infty = \times_{i=1}^\infty \mathbb{R}$ be endowed with the standard product topology, or equivalently, the metric $d(\mathbf{x}, \mathbf{x}') = \sum_{i \geq 1} 2^{-i} |x_i - x_i'|$, and let $C_b(\mathbb{R}^\infty)$ denote the family of bounded, continuous functions $f : \mathbb{R}^\infty \to \mathbb{R}$. Recall that, for a two sided process $\mathbf{X}$, $\mathbf{X}_0^- = X_0, X_{-1}, X_{-2}, \ldots$ denotes the past of the process, starting from time zero, and $\mathbf{X}_k^+ = X_k, X_{k+1}, X_{k+2}, \ldots$ denotes the future starting from some time $k \geq 1$.

Definition: Let $\Theta \subseteq C_b(\mathbb{R}^\infty)$ be a countable family of continuous functions. For each $\mathbf{X} \in \mathcal{M}_s$ and each $k \geq 1$ define the mixing coefficient $\theta(k : \mathbf{X}) = \sup_{g,h \in \Theta} |\operatorname{Cov}(g(\mathbf{X}_0^-), h(\mathbf{X}_k^+))|$. Thus $\theta(k : \mathbf{X})$ measures the dependence between $\mathbf{X}_0^-$ and $\mathbf{X}_k^+$ through the induced covariance of test functions in Θ. Following the usual terminology, a process $\mathbf{X}$ will be called θ-mixing if $\theta(k : \mathbf{X}) \to 0$ as $k \to \infty$. Of primary interest here are families Θ rich enough to ensure that every stationary θ-mixing process is ergodic.

Most asymptotic results for mixing sequences require that the observed process has mixing coefficients that tend to zero at some polynomial or exponential rate. Here we focus attention on polynomial mixing rates and consider the following problem: how can an hypothesis concerning the polynomial decay of $\theta(k : \mathbf{X})$ be tested against a reasonable alternative, based only on observations of $\mathbf{X}$?

Definition: For each $\mathbf{X} \in \mathcal{M}_s$ define the lower polynomial envelope of $\theta(k : \mathbf{X})$ by

$$R_*(\mathbf{X}) \;=\; \sup\left\{\gamma \geq 0 : \sup_{k\geq 1} \theta(k : \mathbf{X})\, k^{\gamma} < \infty\right\},$$

and the upper polynomial envelope of $\theta(k : \mathbf{X})$ by

$$R^*(\mathbf{X}) \;=\; \inf\left\{\gamma \geq 0 : \sup_{g,h\in\Theta} \inf_{k\geq 1} |\operatorname{Cov}(g(\mathbf{X}_0^-), h(\mathbf{X}_k^+))\,|\, k^{\gamma} > 0\right\}.$$

One may readily verify that $R_*(\mathbf{X}) \leq R^*(\mathbf{X})$. A process $\mathbf{X}$ Θ-regular if $R_*(\mathbf{X}) = R^*(\mathbf{X}) > 0$, in which case their common value will be denoted by $R(\mathbf{X})$.

As defined above, $R_*(\mathbf{X})$ is the supremum of those γ for which $\theta(k : \mathbf{X}) = O(k^{-\gamma})$, and is therefore an upper bound on the polynomial decay rate of $\theta(k : \mathbf{X})$. In general the supremum over k of $\theta(k : \mathbf{X})\, k^{R_*(\mathbf{X})}$ may be finite or infinite. By contrast, each rate $\gamma > R^*(\mathbf{X})$ must be "witnessed" by a pair of test functions whose correlations decay more slowly than $k^{-\gamma}$ along some subseqence. If $\theta(k : \mathbf{X})$ tends to zero faster than any polynomial (*e.g.*, at an exponential rate) then $R_*(\mathbf{X}) = \infty$, while if $\theta(k : \mathbf{X})$ tends to zero slower than any polynomial (*e.g.* at a logarithmic rate) then $R_*(\mathbf{X}) = 0$. The proof of the next lemma is given in Appendix 8.3.

Lemma 4 *Suppose that $R_*(\mathbf{X}) > 0$ implies $\mathbf{X}$ is ergodic. If $\mathcal{S}$ is a countable union of uniformly tight subsets of $\mathcal{M}_s$, then for each $c > 0$, $H_{0,c} = \{\mathbf{X} \in \mathcal{S} : 0 < R_*(\mathbf{X}),\, R^*(\mathbf{X}) < c\}$ and $H_{1,c} = \{\mathbf{X} \in \mathcal{S} : R_*(\mathbf{X}) > c\}$ are continuously discernible.*

Theorem 3 *Suppose that $R_*(\mathbf{X}) > 0$ implies $\mathbf{X}$ is ergodic. If $\mathcal{S}$ is a countable union of uniformly tight subsets of $\mathcal{M}_s$, then there exist functions $\psi_n : \mathbb{R}^n \to [0, \infty)$, $n \geq 1$, such that*

$$R_*(\mathbf{X}) \;\leq\; \liminf_n \psi_n(X_1^n) \;\leq\; \limsup_n \psi_n(X_1^n) \;\leq\; R^*(\mathbf{X}) \tag{3}$$

with probability one for every $\mathbf{X} \in \mathcal{S}$ such that $R_*(\mathbf{X}) > 0$. In particular, $\psi_n(X_1^n) \to R(\mathbf{X})$ with probability one if $\mathbf{X} \in \mathcal{S}$ is Θ-regular.

Remark: If every Θ-mixing process is ergodic, then each appearance of the condition $R_*(\mathbf{X}) > 0$ above may be replaced by the (weaker) condition that $\theta(k : \mathbf{X}) = O(a_k)$ for some fixed sequence of numbers $\{a_k\}$ tending to zero with increasing k. The initial tightness condition is satisfied if $E\gamma(|X|) < \infty$ for every $\mathbf{X} \in S$, where $\gamma(\cdot)$ is a positive, unbounded increasing function.

Proof: Let $\{\varphi_{n,c} : n \geq 1\}$ be test functions for the hypotheses $H_{0,c}, H_{1,c}$ of Lemma 4, and let $k = k(n)$ be any sequence of integers tending to infinity with n. Calculation of $\psi_n(X_1^n)$ proceeds iteratively, in k steps. Beginning with $c_1 = 1$, at each stage a new value c_{j+1} is selected based on the value of φ_{n,c_j}, and $\psi_n(X_1^n)$ is set to the value of c_k. Formally, let $c_1 = 1$ and for $j = 1, \ldots, k-1$ do the following.

Case 1: $\varphi_{n,c_j}(X_1^n) \geq 1/2$. If $c_j = \max\{c_1, \ldots, c_j\}$ then let $c_{j+1} = c_j + 1$; otherwise set $c_{j+1} = c_j + \min\{c_i - c_j : c_i > c_j\}/2$.

Case 2: $\varphi_{n,c_j}(X_1^n) < 1/2$. If $c_j = 1$ then let $c_{j+1} = 1/2$; otherwise set $c_{j+1} = c_j - \min\{c_j - c_i : c_j > c_i\}/2$.

Define $\psi_n(X_1^n) = c_k$. Suppose that $R_*(\mathbf{X}) > 0$. Then for each $c > 0$, eventually almost surely $\gamma_{n,c}(X_1^n) < 1/2$ if $R^*(\mathbf{X}) < c$, and $\gamma_{n,c}(X_1^n) > 1/2$ if $R_*(\mathbf{X}) > c$. The recursive procedure above ensures that, for every $\epsilon > 0$, c_k is eventually almost surely less than $R^*(\mathbf{X}) + \epsilon$. This establishes the final inequality in (3), and the first follows in a similar fashion.

If we wish to determine a lower bound on the polynomial decay rate of the mixing coefficients of an observed process $\mathbf{X}$, then the estimates ψ_n of Theorem 3 are informative only if $\mathbf{X}$ is Θ-regular, or if there is a known upper bound on the difference $R^*(\mathbf{X}) - R_*(\mathbf{X})$. The assumption in the definition of $\theta(k : \mathbf{X})$ that Θ is countable is needed to apply Theorem 1 in the proof of Lemma 4. This assumption can be relaxed, though one must then exercise some caution in interpreting $R^*(\mathbf{X})$. Let us write $\Theta' \sim \Theta$ for a family of processes $\mathcal{S}$ if the supremum of $|\operatorname{Cov}(g(\mathbf{X}_0^-), h(\mathbf{X}_k^+))|$ over $g, h \in \Theta'$ is equal to the supremum over $g, h \in \Theta$ for each $k \geq 1$ and $\mathbf{X} \in \mathcal{S}$. If $\Theta' \sim \Theta$, then Θ' and Θ define the same mixing coefficients for $\mathbf{X} \in \mathcal{S}$. We state the following lemma without proof.

Lemma 5 *If $\mathcal{S}$ is a countable union of uniformly tight subsets of $\mathcal{M}_s$, and Θ is a subfamily of $C_b(\mathbb{R}^\infty)$ such that $\sup_{f\in\Theta} ||f||$ is finite, then there exists a countable family $\tilde{\Theta} \subset \Theta$ such that $\tilde{\Theta} \sim \Theta$ on $\mathcal{S}$.*

Concerning the subfamily $\tilde{\Theta}$ above, with the obvious notation, $\tilde{R}_*(\mathbf{X}) = R_*(\mathbf{X})$ for $\mathbf{X} \in \mathcal{S}$. On the other hand, $\tilde{R}^*(\mathbf{X})$ is greater than or equal to $R^*(\mathbf{X})$; the inequality may be strict, and may depend on the choice of approximating family $\tilde{\Theta}$. With this caveat, the coefficients $\theta(k, \mathbf{X})$ defined above are broad enough to encompass several common mixing conditions. For example, if Θ is the family of all $\phi \in C_b(\mathbb{R}^\infty \times \mathbb{R}^\infty)$ with $0 \leq \phi \leq 1$, then $\theta(k : \mathbf{X})$ coincides with the strong mixing coefficients $\alpha(k : \mathbf{X})$ defined in (2) above. Other choices of Θ yield mixing conditions recently introduced by Doukhain and Louhichi [12].

In so far as estimation of the lower polynomial envelope $R_*(\mathbf{X})$ is concerned, the difficulties described above may be unavoidable. We conjecture that no estimation scheme can provide consistent estimates of the lower polynomial α-mixing rate $R_*^\alpha(\mathbf{X})$ for the family of stationary process having finite mean and $R_*^\alpha(\mathbf{X}) > 0$.

8 Appendix

8.1 Proof of Proposition 2

Let $\mathcal{S}$ be uniformly ergodic. Fix $g \in \cup_{k\geq 1}\mathcal{C}_k$ and $\epsilon > 0$. Let $N = N(g, \epsilon)$ be such that $\sup_{\mathbf{X}\in\mathcal{S}} \Gamma_n(\mathbf{X}, g) < \epsilon$ for every $n \geq N$. If $\mathbf{X}^* \in \overline{\mathcal{S}}$ then there exist processes $\mathbf{X}_1, \mathbf{X}_2, \ldots \in \mathcal{S}$ such that $\mathbf{X}_r \Rightarrow \mathbf{X}^*$. As $\Gamma_n(\cdot, g)$ is continuous, for every $n \geq N$,

$$V_n(\mathbf{X}^*, g) \ = \ \lim_{r\to\infty} V_n(\mathbf{X}_r, g) \ \leq \ \sup_{\mathbf{X}\in\mathcal{S}} V_n(\mathbf{X}, g) < \epsilon.$$

Since N depends only on g and ϵ, and not on $\mathbf{X}^*$, it follows that $\overline{\mathcal{S}}$ is uniformly ergodic.

Suppose that $\mathcal{S} \subseteq \mathcal{E}$ is not uniformly ergodic. Then there exists a function $g \in \cup_{k\geq 1}\mathcal{C}_k$ and a constant $\eta > 0$ such that

$$\eta \ < \ \limsup_n \sup_{\mathbf{X}\in\mathcal{S}} V_n(\mathbf{X}, g) \ = \ \inf_{n\geq 1} \sup_{\mathbf{X}\in\mathcal{S}} V_n(\mathbf{X}, g),$$

where the equality follows from the subadditivity of the constants a_n defined in Lemma 2. Recall that $V_{2^m}(\mathbf{X}, g)$ is decreasing in m. The last display implies that for each $m \geq 1$ there is a process $\mathbf{X}_m \in \mathcal{S}$ such that $V_{2^m}(\mathbf{X}_m, g) > \eta$ and $V_{2^l}(\mathbf{X}_m, g) > \eta$ for $1 \leq l \leq m-1$. If $\mathcal{S}$ is tight, then there exists a subsequence $\{\mathbf{X}_{m_k}\}$ of these processes that converges weakly to a process $\mathbf{X}^* \in \mathcal{M}_s$. In particular, $V_{2^l}(\mathbf{X}^*, g) = \lim_k V_{2^l}(\mathbf{X}_{m_k}, g) > \eta$ for every $l \geq 1$. Thus $V_n(\mathbf{X}^*, g)$ does not tend to zero as $n \to \infty$, so $\mathbf{X}^*$ fails to be ergodic.

8.2 Proof of Theorem 1

Lemma 6 *If $U, V \subseteq \mathcal{M}_s$ are disjoint families of processes such that U is compact and V is closed, then there exist integers $k, r \geq 1$, functions $h_1, \ldots, h_r \in \mathcal{C}_k$, constants $a_1, \ldots, a_r \in \mathbb{R}$, and positive $\epsilon_1, \ldots, \epsilon_r > 0$ such that*

1. *For every $\mathbf{X} \in U$, $|Eh_j(X_1^k) - a_j| < \epsilon_j$ for some $1 \leq j \leq r$.*
2. *For every $\mathbf{X} \in V$, $|Eh_j(X_1^k) - a_j| > 2\epsilon_j$ for every $1 \leq j \leq r$.*

Proof: Fix $\mathbf{X} \in U$. By assumption, $\mathbf{X}$ has a neighborhood that is disjoint from V. In particular, there exist $l \geq 1$, $h \in \mathcal{C}_l$, $a \in \mathbb{R}$ and $\epsilon > 0$, such that $Eh(X_1, \ldots, X_l) = a$ and the basic open set $\{\mathbf{X}' : |Eh(X'_1, \ldots, X'_l) - a| < 3\epsilon\}$ is disjoint from V. Define $O(\mathbf{X}) = \{\mathbf{X}' : |Eh(X'_1, \ldots, X'_l) - a| < \epsilon\}$. Then $\mathcal{O} = \{O(\mathbf{X}) : \mathbf{X} \in U\}$ is an open cover of U. As U is compact, $\mathcal{O}$ has a finite subcover $\mathcal{O}' = \{O_1, \ldots, O_r\}$ with $O_j = \{\mathbf{X}' : |Eh_j(X'_1, \ldots, X'_{l_j}) - a_j| < \epsilon_j\}$. Letting $k = \max\{l_j : j = 1, \ldots, r\}$ we may redefine the functions h_j so that each is a function of k arguments. The lemma then follows from the choice of basic open sets $O(\mathbf{X})$.

Proof of Theorem 1: As unions of closed sets are closed, we may assume without loss of generality that $U_i \subseteq U_{i+1}$ and $V_i \subseteq V_{i+1}$. By condition (i), there exist compact sets $W_1 \subseteq W_2 \subseteq \cdots$ such that $H_0, H_1 \subseteq \cup_{l \geq 1} W_l$. Thus $H_0 \subseteq \cup_{i \geq 1} U'_i$ and $H_1 \subseteq \cup_{i \geq 1} V'_i$, where $U'_i = U_i \cap W_i \subseteq \mathcal{E}$ and $V'_i = V_i \cap W_i \subseteq \mathcal{E}$ are compact and disjoint.

We may recursively construct a testing scheme to distinguish between H_0 and H_1 as follows. Suppose that integers $0 = n_0 < n_1 < \cdots < n_{j-1}$ and continuous test functions $\varphi_1, \ldots, \varphi_{n_{j-1}}$ have been selected. Consider the disjoint compact sets U'_j and V'_j. By Lemma 6 there exist integers $r_j, k_j \geq 1$, functions $h_1, \ldots, h_{r_j} \in \mathcal{C}_{k_j}$, constants $a_1, \ldots, a_{r_j} \in \mathbb{R}$, and $\epsilon_1, \ldots, \epsilon_{r_j} > 0$, all depending on U'_j, V'_j, such that (i) for every $\mathbf{X} \in U'_j$, $|Eh_l(X_1^k) - a_l| < \epsilon_l$ for some $1 \leq l \leq r$ and (ii) for every $\mathbf{X} \in V'_j$, $|Eh_l(X_1^k) - a_l| > 2\epsilon_l$ for every $1 \leq l \leq r$. (Here and in what follows we suppress the dependence of r, h_l, a_l, k and ϵ_l on j.) Define $\eta_j = \min\{\epsilon_l\} > 0$ and let $\mathcal{S}_j = U'_j \cup V'_j$. By Lemma 1, $\mathcal{S}_j$ is uniformly ergodic, and therefore there is an integer $n_j > \max\{n_{j-1}, k_j\}$ such that for $l = 1, \ldots, r$,

$$\sup_{\mathbf{X} \in \mathcal{S}_j} P\left\{ \left| \frac{1}{m_j} \sum_{i=0}^{m_j - 1} h_l(X_{i+1}^{i+k}) \; - \; Eh_l(X_1^k) \right| > \eta_j \right\} \;\leq\; \frac{1}{\eta_j^2} \sup_{\mathbf{X} \in \mathcal{S}_j} V_{m_j}(\mathbf{X}, h_l) \;\leq\; \frac{1}{r\, j^2} \qquad (4)$$

where $m_j = n_j - k_j + 1$. Let $\varphi_n(x_1^n) = \varphi_{n_{j-1}}(x_1^{n_{j-1}})$ for $n_{j-1} \leq n < n_j$ and define

$$\varphi_{n_j}(x_1^{n_j}) \;=\; 2 \min_{1 \leq l \leq r} \left\{ \left[\frac{|m_j^{-1} \sum_{i=0}^{m_j - 1} h_l(x_{i+1}^{i+k}) \; - \; a_l|}{2\epsilon_l} \wedge 1 \; - \; \frac{1}{2} \right] \vee 0 \right\}.$$

Repeat this process for $j+1, j+2, \ldots$. By construction, each function $\varphi_n(\cdot)$ takes values in $[0, 1]$ and depends continuously on its n arguments. For a given sequence $x_1, x_2, \ldots$, the value of $\varphi_n(x_1^n)$ changes only at times $n_1, n_2, \ldots$, and is constant for $n_j \leq n < n_{j+1}$. The function φ_{n_j} assesses whether the observed process $\mathbf{X}$ belongs to U'_j or V'_j.

Suppose now that we observe a process $\mathbf{X}$ belonging to H_0. Then $\mathbf{X} \in U'_s$ for some $s \geq 1$. By definition of φ_{n_j}, for every $j \geq s$ and every rational $\delta > 0$, $P\{\varphi_{n_j}(X_1^{n_j}) > \delta\}$ is bounded above by

$$P\left\{\left|m_j^{-1}\sum_{i=0}^{m_j-1} h_l(X_{i+1}^{i+k}) \;-\; a_l\right| \;>\; (1+\delta)\,\epsilon_l \;\text{ for each }\; l=1,\ldots,r\right\}$$

$$\leq\;\; P\left\{\left|m_j^{-1}\sum_{i=0}^{m_j-1} h_l(X_{i+1}^{i+k}) \;-\; Eh_l(\mathbf{X})\right| \;>\; \delta\,\epsilon_l \;\text{ for some }\; l=1,\ldots,r\right\}$$

$$\leq\;\; \sum_{l=1}^{r} P\left\{\left|m_j^{-1}\sum_{i=0}^{m_j-1} h_l(X_{i+1}^{i+k}) \;-\; Eh_l(\mathbf{X})\right| \;>\; \delta\eta_j\right\} \;\leq\; r\cdot\frac{1}{\delta^2\, j^2 r} \;=\; \frac{1}{\delta^2\, j^2}.$$

The first inequality above follows from (i), and the third is a consequence of (4). It follows from the Borel Cantelli lemma that $\phi_{n_j}(X_1^{n_j}) \to 0$ with probability one, and consequently $\phi_n(X_1^n) \to 0$ with probability one as well.

Suppose now that we observe a process $\mathbf{X}$ belonging to H_1. Then $\mathbf{X} \in V_s'$ for some $s \geq 1$, and for every $j \geq s$ and every rational $\delta > 0$,

$$P\{\varphi_{n_j}(X_1^{n_j}) < (1-\delta)\} \;\leq\; P\left\{\min_{1\leq l\leq r}\left|m_j^{-1}\sum_{i=0}^{n-k} h_l(X_{i+1}^{i+k}) \;-\; a_l\right| \;<\; 2\epsilon_l\,(1-\delta/2)\right\}.$$

As $|Eh_l(\mathbf{X}) - a_l| > 2\epsilon_l$, the right hand side above is at most

$$P\left\{\max_{1\leq l\leq r}\left|m_j^{-1}\sum_{i=0}^{n_j-k} h_l(X_{i+1}^{i+k}) \;-\; Eh_l(\mathbf{X})\right| > \delta\,\epsilon_l\right\}$$

$$\leq\;\; \sum_{l=1}^{r} P\left\{\left|m_j^{-1}\sum_{i=0}^{n_j-k} h_l(X_{i+1}^{i+k}) \;-\; Eh_l(\mathbf{X})\right| > \delta\,\epsilon_l\right\} \;\leq\; r\frac{1}{\delta^2\, j^2 r} \;=\; \frac{1}{\delta^2\, j^2}.$$

It follows from the Borel Cantelli lemma that $\phi_{n_j}(X_1^{n_j}) \to 1$ with probability one, and consequently $\phi_n(X_1^n) \to 1$ with probability one as $n \to \infty$.

8.3 Proof of Lemma 4

Proof: For $a > 0$ and $l \geq 1$ define $U(a,l) = \left\{\mathbf{X} \in \mathcal{M}_s : \sup_{k\geq 1}\theta(k:\mathbf{X})\,k^a \leq l\right\}$, or equivalently,

$$U(a,l) \;=\; \bigcap_{k\geq 1}\bigcap_{g,h\in\Theta}\{\mathbf{X}\in\mathcal{M}_s :\; |\,\mathrm{Cov}(g(\mathbf{X}_0^-), h(\mathbf{X}_k^+))|\,k^a \leq l\}$$

As each set in the last intersection is closed in $\mathcal{M}_s$, the same is true of $U(a,l)$. Moreover, $U(a,l) \subseteq \mathcal{E}$ as $\mathbf{X} \in U(a,l)$ implies $R_*(\mathbf{X}) > 0$. For $b \geq 0$ define

$$V(b) \;=\; \{\mathbf{X}\in\mathcal{M}_s : R_*(\mathbf{X}) > b\} \;=\; \bigcup_{\delta>0}\bigcup_{l\geq 1} U(b+\delta, l).$$

Restricting the first union to rational δ, we see that $V(b)$ is an E_σ. For $g, h \in \Theta$ and $a, \epsilon > 0$ define

$$C(g,h,a,\epsilon) \;=\; \bigcap_{k\geq 1}\{\mathbf{X}\in\mathcal{M}_s : |\operatorname{Cov}(g(\mathbf{X}_0^-), h(\mathbf{X}_k^+))|\, k^a \geq \epsilon\},$$

which is closed in $\mathcal{M}_s$. Finally, let

$$D(c) \;=\; \{\mathbf{X}\in\mathcal{M}_s : 0 < R_*(\mathbf{X}),\, R^*(\mathbf{X}) < c\} \;=\; \bigcup_{\delta>0}\bigcup_{\epsilon>0}\bigcup_{g,h\in\Theta} C(g,h,c-\delta,\epsilon)\cap B(0).$$

As $B(0)$ is an E_σ, the same is true of $D(c)$. Moreover, $H_{0,c} \subseteq D(c)$, $H_{1,c} \subseteq B(c)$ and $D(c)\cap B(c) = 0$ since $R_*(\mathbf{X}) \leq R^*(\mathbf{X})$. The discernibility of $H_{0,c}$ and $H_{1,c}$ follows from Theorem 1.

Acknowledgement: The author's work was supported in part by NSF Grant DMS 9971964.

References

[1] Adams, T.M. (1997). Families of ergodic processes without consistent density or regression estimates. Preprint.

[2] Adams, T.M. and Nobel, A.B. (1997). On density estimation from ergodic processes. *Ann. Probab.* **26** 794–804.

[3] Barron, A.R. (1985). Logically Smooth Density Estimation. Ph.D. dissertation, Dept. Electrical Engineering, Stanford University, Stanford CA.

[4] Berger, A. (1951). On uniformly consistent tests. *Ann. Math. Stat.* **18** 289-293.

[5] Berger, A. (1953). On orthogonal probability measures. *Proc. Amer. Math. Soc.* **4** 800-806.

[6] Billingsley, P. (1999). *Convergence of Probability Measures*, Second edition. Wiley, New York.

[7] Bradley, R.C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics*, Eberlein, E., and Taqqu, M.S. (eds.), Progress in Probability and Statistics, **11** 165-192, Birkhauser, Boston.

[8] Breiman, L. (1992). *Probability.* SIAM, Philadelphia, PA.

[9] Cover, T.M. (1973). On determining the irrationality of the mean of a random variable. *Ann. Stat.* **1** 862-871.

[10] Dembo, A. and Peres, Y. (1994). A topological criterion for hypothesis testing. *Ann. Stat.* **22** 106-117.

[11] Devroye, L. and Lugosi, G. (2003). Almost sure classification of densities. *J. Nonparam. Stat* **14** 675-698.

[12] Doukhan, P. and Louhichi, L. (1999). A new weak dependence condition and applications to moment inequalities. Stoch. Proc. Appl. **84** 313-342.

[13] Dudley, R. (1989). *Real Analysis and Probability.* Chapman and Hall, New York.

[14] Fisher, L. and Van Ness, J.W. (1969). Distinguishibility of probability measures. *Ann. Math. Stat.* **40** 381-399.

[15] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58** 13-30.

[16] Hoeffding, W. and Wolfowitz, J. (1958). Distinguishability of sets of distributions. *Ann. Math. Stat.* **29** 700-718.

[17] Kraft, C. (1955). Some conditions for constency and uniform consistency of statistical procedures. *Univ. Cal. Pub. Stat.* **2** 125-142, Berkeley.

[18] Krengel, U. (1985). *Ergodic Theorems.* De Gruyter, Berlin.

[19] Kulkarni, S.J. and Zeitouni, O. (1995). A general classification rule for probability measures. *Ann. Stat.* **23** 1393-1407.

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

[21] Lim, J. (2002). Hypothesis testing in time series. Preprint.

[22] Nobel, A.B. (1999). Limits to classification and regression estimation from ergodic processes. *Annals of Statistics* **27** 262-273.

[23] Ornstein, D.S. and Weiss, B. (1990). How Sampling Reveals a Process. *Annals of Probability,* **18** 905-930.

[24] Ornstein, D.S. and Shields, P.C. (1994). The $\overline{d}$ recognition of processes. *Adv. Math.,* **104** 182-224.

[25] Petersen, K. (1989). *Ergodic Theory.* Cambridge Univ. Press.

[26] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43-47.

[27] Zeitouni, O. and Kulkarni, S.J. (1994). A general classification rule for probability measures. Technical report LIDS-P-2272, Laboratory for Information and Decision Systems, MIT.