TESTING STOCHASTIC PROCESSES:
Stationarity, Independence and Ergodicity

by

Johan Lim

August 2003

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
TESTING STOCHASTIC PROCESSES:
Stationarity, Independence and Ergodicity

by

Johan Lim
Department of Statistics
Stanford University

August 2003

This research was supported in part by
National Science Foundation grant DMS-0072331.

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Testing Stochastic Processes:
Stationarity, Independence and Ergodicity

Johan Lim*

Abstract

Two sets of probability measures $H_0$ and $H_1$ are called discernible, when there exists a sequence of discerning functions $f_n : \mathbb{R}^n \to \{0, 1\}$ such that almost surely only finitely many errors are made. This paper studies the discernibility of various families of stochastic processes. First, three different modes of discernibility are introduced: discernibility with an entire sample path (DES), uniform discernibility with an entire sample path (UDES), and sequential discernibility (SD). These modes are shown to have a strictly nested structure. Second, using phase transition phenomena in random coin tossing, we construct criteria for SD between stationary (independent) and non-stationary (dependent) measures. The proposed criteria are functions of (i) how much non-stationary components deviate from their stationary counterparts and (ii) how frequently non-stationary components occur in a long sequence. Third, we study the SD between the independent fair coin tossing $\mu_0$ and the sparse heterogeneous mixtures $HM(\gamma, \theta) := (1 - \epsilon_n)\text{Bernoulli}(1/2) + \epsilon_n\text{Bernoulli}(1 + \theta)/2$, where $\epsilon_n = n^{-\gamma}$ with $\gamma \in (0, 1)$. It is shown that $HM(\gamma, \theta)$ is sequentially discernible from $\mu_0$ when $\gamma \in (0, 0.5)$, but it is not when $\gamma \in (0.5, 1)$. We also give a negative result for testing ergodicity and a general criterion on SD for a broader class of stochastic processes.

KEY WORDS: Hypothesis Testing; Discernibility; Stochastic Processes; Stationarity; Independence; Ergodicity.

*This work is supported in part by NSF grant #DMS-0072331.
1 Introduction

Suppose one wants to decide between two mutually exclusive sets of probability measures, denoted \( H_0 \) and \( H_1 \), using sequential observations \( X_1, X_2, X_3, \ldots \). At each time \( n \), one decides between \( H_0 \) and \( H_1 \) based on the observations \( X_1, X_2, \ldots, X_n \). When does there exist a sequence of tests for deciding between \( H_0 \) and \( H_1 \) such that almost surely only finite many errors are made?

To date, work on the “discernibility” of probability measures has primarily focused on the special case in which the sequential observations are independent and identically distributed (IID) (e.g. Hoeffding and Wolfowitz, 1957; Fisher and Van Nese, 1969; Cover, 1973; Dembo and Peres, 1994; Kulkarni and Zeitouni, 1996). In this case, the elements of \( H_0 \) and \( H_1 \) are fully described by their associated families of one-dimensional marginal distributions. Unlike IID samples, the issues concerning families of dependent processes have not been much studied. Kraft (1955) studied consistent test for families \( H_0 \) and \( H_1 \) of general distributions on \( \mathbb{R}^n \). 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. Nobel (2003) provided a sufficient condition for testing stationary ergodic processes. In both works these conditions involve infinite-dimensional measures and are thus hard to verify.

In this paper, three problems about testing stochastic processes are discussed. First, when \( n \to \infty \), are the sequential decisions equivalent to those relying on the entire sample path? Second, what is the minimal class of non-stationary measures which are discernible from stationary measures? Third, is there a general criterion for discernibility between two sets of stochastic processes analogous to Theorem 2 of Dembo and Peres (1994) for the IID case?

In Section 2, the three different modes of discernibilities described below are studied:

**Definition 1.1.** Two classes of probability measures \( H_0 \) and \( H_1 \) on \( (\mathcal{B}^\infty, \mathbb{R}^\infty) \) are Discernible with an Entire Sample path (DES), if, for each \( P \in H_0 \) and \( Q \in H_1 \), there exists a measurable function \( f_{P,Q} : \mathbb{R}^\infty \to \{0,1\} \) such that \( f_{P,Q}(X_1^\infty) = 1 \), \( Q \)-a.s and \( f_{P,Q}(X_1^\infty) = 0 \), \( P \)-a.s. When the discerning function \( f_{P,Q} \) does not depend on \( (P,Q) \), \( H_0 \) and \( H_1 \) are called Uniformly Discernible with an Entire Sample path (UES).

**Definition 1.2.** Two classes of probability measures \( H_0 \) and \( H_1 \) are Sequentially Discernible (SD) if there exists a sequence of measurable functions \( f_n : \mathbb{R}^n \to \{0,1\} \) such that \( \lim_{n \to \infty} f_n(x_1, x_2, \ldots, x_n) = 1 \) \( Q \)-a.s. and \( \lim_{n \to \infty} f_n(x_1, x_2, \ldots, x_n) = 0 \) \( P \)-a.s., for every \( P \in H_0 \) and \( Q \in H_1 \).

From the definition, clearly \( (SD) \implies (UES) \implies (DES) \). We show that neither \( (SD) \iff (UES) \) nor
(UDES) \iff (DES) holds (Theorems 2.2-2.3). Hence sequential decisions are different from those relying on an entire sequence.

Section 3.1 discusses the criteria for SD between stationary and non-stationary measures in a special setting of binary sequences. This is done by relating testing stationarity to a seemingly different subject, phase transition in random coin tossing. Suppose \( \{ \Delta_n \} \in \{0, 1\} \) is an unobserved binary sequence. An independent coin with a bias \( \theta > 0 \) \( (P(\text{Heads}) = (1 + \theta)/2) \) is tossed when \( \Delta_n = 1 \); a fair coin is tossed when \( \Delta_n = 0 \). Let \( \mu_{\theta}(l) \) with \( l = (t_1, t_2, \cdots, t_n, \cdots) \) be the probability measure for the record of independent coin tosses, where \( \Delta_n = 1 \), if \( n \in \{t_1, t_2, \cdots \} \), and 0 otherwise. We say that \( l \) has 0-density when \( \lim_{n \to \infty} |l \cap \{1, 2, \cdots, n\}|/n = 0 \), where \( |A| \) is the size of the finite set \( A \). Note that, for every \( l \), \( \mu_0(l) \) is the same measure of fair coin tossing, denoted as \( \mu_0 \).

When \( l \) is from a renewal process (we call it a hidden renewal process since \( l \) is not observed), let \( u_n = P(\Delta_n = 1) \) and consider the set MLPP of all \( \mu_\theta(l) \), where (i) \( \sum u_n^2 = \infty \) or (ii) \( u_n \sim n^\gamma \) for some \( \gamma \in (0.5, 1) \) and \( \theta > 2\gamma / \max \{u_i : i \geq 1\} - 1 \). Here, \( u_n \sim n^{-\gamma} \) implies \( 0 < \lim \inf_{n \to \infty} u_n n^\gamma \leq \lim \sup_{n \to \infty} u_n n^\gamma < \infty \). We relate the set of probability measures which is SD from \( \mu_0 \) (denoted by MSD) to MLPP, MPD (the set of \( \mu_\theta(l) \) where \( l \) has a positive density) and MHKC (the set of \( \mu_\theta(l) \) where \( l \) is from the renewal process satisfying the conditions \( \sum u_n^2 < \infty \) and \( \theta \sum u_n^2 < 1 \)). We show that MPD \( \subset \) MLPP (Lemma 3.1), MHKC \( \not\subset \) MSD (Lemma 2.1), and MLPP \( \subset \) MSD (Theorem 3.2). Theorem 3.2 shows that it is possible to detect non-stationarity in a sequence of independent binary observations, even when the empirical means do not deviate from the alternative stationary value (asymptotic stationary). Figure 1 graphically presents the results in Section 3.1. The above results on random coin tossing are directly extended to general settings (including finite state Markov chain and Gaussian autoregressive processes). The extensions are discussed in Section 3.2.
A similar problem for (independent) Gaussian mixtures has recently been addressed by Jin and Donoho (2003). They obtained the boundary of SD between $H_0 : \text{IID } N(0, 1)$ and $H_1 : \epsilon_n N(0, 1) + (1 - \epsilon_n) N(\mu_n, 1)$ with $\epsilon_n \sim n^{-\gamma}$ and $\mu_n \sim \sqrt{2q \log n}$. In their work, the increasing mean of the heterogeneous (non-stationary) components allows SD when $\gamma > 0.5$ for some $q$. In contrast, in this section, we obtain a positive result for $\gamma \in (0.5, 1)$ with the bounded mean (of non-stationary components) by using the dependence structure induced by a hidden renewal process. Note that the next section shows that, if $\mu_n$ is bounded, $H_1$ is not SD from $H_0$ when $\gamma \in (0.5, 1)$.

In Section 4, we consider independent Bernoulli trials where $\Delta_n$ is independently from Bernoulli($\epsilon_n$) with $\epsilon_n = n^{-\gamma}$ for $\gamma \in (0, 1)$. Denote the corresponding probability measure as $\text{HM}(\theta, \gamma)$. It is shown that $\text{HM}(\theta, \gamma)$ is not DES from $\mu_0$ when $0.5 < \gamma < 1$ from Kakutani's dichotomy (Theorem 3.A); thus, it is neither UDES nor SD from $\mu_0$. On the other hand, it is proven that $\text{HM}(\theta, \gamma)$ is SD from $\mu_0$ when $0 < \gamma < 0.5$ and UDES from $\mu_0$ when $\gamma = 0.5$ (Theorem 4.1). Thus, in IID mixtures with regular components (here, regular implies the two mixture components are fixed for every $n$), UDES is (almost) equivalent to SD with boundary at $\gamma = 0.5$. It is interesting to see that the boundary does not depend on $\theta$. The above results in random coin tossing can be easily extended to general distributions.

In Section 5, the results on testing stationarity are extended to testing independence. We show that testing independence is a special type of testing stationarity. Thus, most of the results on independence can be shown by applying the theorems in Section 3 to a suitable stochastic process. In addition, a negative result is proven for discerning between (strongly mixing) ergodic and non-ergodic measures, using the ergodic decomposition of the stationary non-ergodic measure.

Finally, Section 6 extends Theorem 2 of Dembo and Peres (1994) on IID observations to testing general stochastic processes. Here, a sufficient condition is proposed for testing broader families of stochastic processes (beyond the stationary ergodic measures). When the processes are stationary and ergodic, our criterion is a special case of Theorem 1 in Nobel (2003).

2 Modes of Discernibility

In this section, we discuss three different modes of discernibility in detail and prove that neither $(SD) \iff (UDES)$ nor $(UDES) \iff (DES)$ holds.
It is worth noting that, in IID settings, UDES = DES and they are trivial in the sense that any disjoint families of probability measures are UDES (DES) because an infinite number of samples provides exact knowledge of the true distribution. On the other hand, several different modes of SD have been proposed in previous literature. Hoeffding and Wolfowitz (1957) proposed five different classes of test relying on the sample size function $N$ and the decision function $\phi(X_1, \ldots, X_n)$. Here, $N$ is a stopping time with respect to $\{\sigma(X_1, \ldots, X_n)\}_{n=1}^\infty$, and $\phi \in \sigma(X_1, \ldots, X_n)$ is the testing function in $\{0, 1\}$. They defined $H_0$ and $H_1$ to be "discernible" if there exists a test such that the probability of error is smaller than $\epsilon$ for every $\epsilon > 0$. Among Hoeffding and Wolfowitz’s five classes, the class of tests $P(N < \infty) = 1$ has been widely used in previous literature and known as SD (Fisher and Van Ness, 1969; Cover, 1973; Dembo and Peres, 1994; Kulkarni and Zeitouni, 1996; Nobel, 2003). The same notion of SD is commonly expressed as follows

**Definition 2.A.** Let $H_0$ and $H_1$ be two ensembles of probability measures on $(R^d, R^d)$. We say that $H_0$ and $H_1$ are SD if there exists a sequence of Borel functions $f_n : (R^d)^n \to \{0, 1\}$ such that

$$\lim_{n \to \infty} f_n(X_1, X_2, \ldots, X_n) = j, \mu - a.s.$$ holds, whenever $X_1, X_2, \ldots, X_n, \ldots$ are IID with common probability distribution $\mu \in H_j$.

Unlike in IID samples, the relation between the proposed modes of discernibility is not obvious in time series. In the following, we show the nested structure of the three different modes that have been introduced.

### 2.1 DES does not imply UDES

Let $\{\Delta_n\}_{n=1}^\infty$, $\mu_\theta(I)$, and $\mu_0$ be the binary sequence and their corresponding measures introduced in the Introduction.

**Lemma 2.1.** Let $H_1$ be the collection of probability measures $\mu_\theta(I)$, where $I$ has 0-density with $|I| = \infty$. Then, $H_1$ is not UDES from $H_0 = \{\mu_0\}$.

**Proof.** Suppose $I$ is a random set of times where a renewal process $\{\Delta_n\}_{n=1}^\infty$ visits a distinguished state. Assume the return probabilities $\{u_n\}_{n=1}^\infty$ and the bias $\theta$ satisfy $\sum_{n=1}^\infty u_n = \infty$, $\sum_{n=1}^\infty u_n^2 < \infty$, and $\theta \cdot \sum_{n=1}^\infty u_n^2 < 1$. Then, according to the discrete renewal theorem, $\sum_{n=1}^\infty u_n^2 < \infty$ implies $\sum_{n=1}^\infty \theta u_n^2 = \infty$ (Feller 1970), where $p_n$ is the probability that the first return is at time $n$. Thus, $G$, the probability measure for a renewal process $\{\Delta_n\}_{n=1}^\infty$, has the support on the collection of 0-density sequences $I$ with $|I| = \infty$ (|I| = \infty follows from
\[ \sum_{n=1}^{\infty} u_n = \infty \]. If we assume that \( H_0 \) and \( H_1 \) are UDES with a discerning function \( f : \mathbb{R}^\infty \to \{0,1\} \), then

\[
\mu_\theta \left( \{ f(X_1, X_2, \cdots) = 1 \} \right) = \int \mu_\theta(l) \left( \{ f(X_1, X_2, \cdots) = 1 \} \right) dG(l) = 1,
\]

whereas \( \mu_0 \left( \{ f(X_1, X_2, \cdots) = 1 \} \right) = 0 \). Therefore, \( \mu_0 \) and \( \{ \mu_\theta(l), l \in \text{support of } G \} \) are mutually singular, but this contradicts Theorem 1 in Harris and Keane (1997) (Theorem 3.B in this paper). The theorem states that \( \mu_\theta \) is absolutely continuous with respect to \( \mu_0 \), when \( \sum_{n=1}^{\infty} u_n^2 < \infty \) and \( \theta \cdot \sum_{n=1}^{\infty} u_n^2 < 1 \). 

**Theorem 2.2.** DES does not imply the UDES.

**Proof.** Let \( \theta_n \) be the bias of the coin in the \( n \)-th trial. Then, in Lemma 2.1, \( \sum_{n=1}^{\infty} \theta_n^2 = \infty \) because \( |l| = \infty \).

From Kakutani's theorem, \( \mu_\theta(l) \in H_1 \) and \( \mu_0 \in H_0 \) are mutually singular for \( H_0 \) and \( H_1 \) in Lemma 2.1. Hence, they are DES, but not UDES as proven in the previous lemma. 

**2.2 UDES does not imply SD**

Let \( C = \{ I_j; 1 \leq j \leq m \} \) be an ordered collection of \( m \) disjoint intervals \( I_j = [a_j, a_j + b) \subset [0,1) \), in which the intervals \( I_j \) are stacked on top of one another (i.e. \( I_1 \) at the bottom and \( I_j \) placed directly above \( I_{j-1} \)).

Associated with a column \( C \), define \( T_C : \bigcup_{j=1}^{m-1} I_j \to \bigcup_{j=2}^{m} I_j \) that maps each point in the first \( m-1 \) levels of \( C \) to the point directly above it. \( C' \) is obtained by cutting \( C \) in half vertically and then stacking the intervals to the right of the cut directly on top of those to the left. To be specific, \( C' = \{ I'_1, \cdots, I'_{2m} \} \) is

\[
I'_j = \begin{cases} 
[a_j, a_j + \frac{b}{2}] & \text{if } 1 \leq j \leq m \\
[a_{j-m}, a_{j-m} + \frac{b}{2}] & \text{if } m+1 \leq j \leq 2m.
\end{cases}
\]

Denote \( C' \) as a 2-cut of \( C \). Now, let \( C = \{ I_1, \cdots, I_m \} \) be an initial column with the support \( U = \bigcup_{j=1}^{m} I_j \) and let \( C_1, C_2, \cdots \) be successive 2-cuts of \( C \). Then, the mappings \( T_{C_1}, T_{C_2}, \cdots \) form a chain whose limit \( T_C^\ast(\omega) = \lim_{n \to \infty} T_{C_n}(\omega) \) is defined for each \( \omega \in U \). Each map \( T_{C_n} \) is measurable and \( \lambda(T_{C_n}^{-1}A) = \lambda(A) \) for each Borel subset \( A \) of \([0,1] \). \( \lambda \) is the Lebesgue measure. As pointed out in Adams and Nobel (1988), \( T_C^\ast \) is measurable and preserves the normalized restriction of Lebesgue measure to \( U \). In addition, it follows from Theorem 6.2 of Friedman (1970) that \( T_C^\ast \) is ergodic.

For each \( k \geq 1 \), let \( \pi_{2k} \) be the partition of \([0,1]\) into \( 2k \) left closed, right open sub-intervals of length \( 1/k \). Subsequently, let \( U_1 \) and \( U_2 \) be exclusive subsets of \( \pi_{2k} \), where each has \( k \) subintervals respectively. Furthermore, let \( U_j \) be the support of \( C_j \) and define probability measures \( \mu_j(A) = 2\lambda(A \cap U_j) \) for \( j = 1, 2 \), with densities
\( f_j(x) = 2I_{U_j}(x) \). Define the transformation

\[
T(\omega) = \begin{cases} 
T^*_j(\omega) & \text{if } \omega \in U_1 \\
T^*_j(\omega) & \text{if } \omega \in U_2.
\end{cases}
\]

Then, \( T \) is ergodic and measure preserving on \( ([0, 1), \mathcal{B}, \mu_\omega) \) for \( j = 1, 2 \). Finally, let \( P_1 \) be the collection of the processes of the form \( X_i(\omega) = T^{i-1} \omega \), in which \( T \) is the above ergodic transformation for some \( k, U_1, \) and \( U_2 \). Note that \( \{X_i\}_{i=1}^\infty \) is stationary and ergodic.

On the other hand, let \( C_1^{(1)} = \{[0, 1/2]\} \) and \( C_1^{(2)} = \{[1/2, 1]\} \) be initial columns. Also, suppose we have selected an integer sequence \( m = \{m_1, \ldots, m_n, \ldots\} \). For \( m_1 \geq 1 \), let \( C_2^{(j)} \ldots, C_{m_1}^{(j)} \) be successive 2-cuts of \( C_1^{(j)} \) for each \( j = 1, 2 \). At time \( m_1 \), we also mangle the two columns by stacking \( C_m^{(2)} \) on top of \( C_m^{(1)} \) and then form new \( C_m^{(j)} \) by cutting this in half. For each of these new columns, perform 2-cuts successively by time \( s(2) = m_1 + m_2 \) and obtain \( C_{m_1+m_1+1}^{(j)} \ldots, C_{m_1+m_2}^{(j)} \) for \( j = 1, 2 \). Mingle the current two columns at time \( s(2) \). Continuing in this fashion, we obtain \( \{C_k^{(1)}, C_k^{(2)} : k \geq 1\} \) of pairs of columns and the associated transformation \( T_m \). Using the same arguments in defining \( P_1 \), \( T_m \) is ergodic. Finally, let \( P_2 \) be the family of processes defined by \( X_i(\omega) = T_n^*_m(\omega) \) for \( \omega \in [0, 1] \), where \( m \) ranges over all sequences of non-negative integers. Then, each element of \( P_2 \) is stationary and ergodic with uniform marginal distribution on \( [0, 1] \).

**Theorem 2.3.** SD does not imply UDES.

**Proof.** As pointed out in Adams and Nobel (1998), there exist consistent density estimation procedures \( \Phi_1 \) and \( \Phi_2 \) for \( P_1 \) and \( P_2 \), respectively. Suppose \( P_1 \) and \( P_2 \) are SD with discerning functions \( \{f_n(X_1, \ldots, X_n)\}_{n=1}^\infty \), i.e. \( \lim_{n \to \infty} f_n(X_1, \ldots, X_n) = j \), \( P_j - a.s. \), for \( j = 0, 1 \). Then, defining a density estimation procedure \( Q_n \) as \( Q_n = f_n \cdot \Phi_1 + (1 - f_n) \cdot \Phi_2 \), we obtain a consistent density estimation procedure for \( P_1 \cup P_2 \). This contradicts Corollary 1 in Adams & Nobel (1998).

Every measure \( \mu_1 \in P_1 \) (or \( \mu_2 \in P_2 \)) has a common support denoted as \( S_1 \) (or \( S_2 \)), where \( \lambda(S_1) = 1/2 \) and \( \lambda(S_2) = 1 \). Then, according to the ergodicity of the sequence \( \{X_n\}_{n=1}^\infty \), the set \( A = \{x_1, x_2, \ldots\} \) is equivalent to the support of \( X_1 \). Hence, the function \( f(X_1, X_2, \ldots) = I(\lambda(\{X_1, X_2, \ldots\}) > 3/4) \) uniformly discerns \( P_1 \) and \( P_2 \).

\( \square \)

**3 Testing Stationarity**

This section discusses the criterion for SD between stationary and non-stationary measures. First, the issue in random coin tossing is studied, where the criterion relies on two factors: (i) how much non-stationary
components deviate from others, and (ii) how frequently non-stationary components appear in a long sequence. Subsequently, the results in random coin tossing are extended to general settings.

3.1 Random Coin Tossing

Phase transition in random coin tossing has been discussed extensively in the literature. The well-known Kakutani's dichotomy for independent coin tosses reduces to the following:

Theorem 3.A. [Kakutani (1948)]

Let \( \mu_0 \) be the distribution of IID fair coin tosses on \( \{-1,1\}^N \), and let \( \mu_\theta \) be the distribution of independent coin tosses with biases \( \{\theta_n\}_{n=1}^\infty \). Then, \( \mu_\theta \) and \( \mu_0 \) are mutually singular, when \( \sum_{n=1}^\infty \theta_n^2 = \infty \); otherwise, \( \mu_\theta \) is absolutely continuous with respect to \( \mu_0 \).

Harris and Keane (1997) extended the results of Kakutani (1948) to more general cases by introducing hidden renewal processes (which were introduced in Introduction and Lemma 2.1). Here, \( \{\Delta_n\}_{n=1}^\infty \) is a hidden renewal process with return probabilities \( \{u_n\}_{n=1}^\infty \). Subsequently, whenever \( \Delta_n = 1 \), an independent coin with bias \( \theta > 0 \) is tossed; otherwise, a fair coin is used. Under this specific dependency, it was shown that

Theorem 3.B. [Harris and Keane (1997)]

(i) If \( \sum_{n=1}^\infty u_n^2 = \infty \), then \( \mu_\theta \) and \( \mu_0 \) are mutually singular \( (\mu_\theta \perp \mu_0) \).

(ii) If \( \sum_{n=1}^\infty u_n^2 < \infty \) and \( \theta^2 \cdot \sum_{n=1}^\infty u_n^2 < 1 \), \( \mu_\theta \) is absolutely continuous with respect to \( \mu_0 \) \( (\mu_\theta \ll \mu_0) \).

Recently, Levin et al. (2001) further extended the results of Harris and Keane (1997) and pointed out that the singularity between \( \mu_\theta \) and \( \mu_0 \) depends not only on the renewal probability \( \{u_n\}_{n=1}^\infty \), but also on \( \theta \).

Theorem 3.C. [Levin et al. (2001)]

Let \( 1/2 < \gamma < 1 \). Suppose that the return probability \( u_n \sim n^{-\gamma} \) and \( \max\{u_i : i \geq 1\} > 2^{\gamma-1} \). Then, \( \mu_\theta \perp \mu_0 \), if \( \theta > \frac{2^{\gamma}}{\max\{u_i : i \geq 1\}} - 1 \).

In Theorem 3.C, only \( 1/2 < \gamma < 1 \) is considered because other cases are covered by Theorem 3.A and Theorem 3.B. Hereafter, in this paper, KD-condition denotes \( \sum_{n=1}^\infty \theta_n^2 = \infty \), HK-condition denotes \( \sum_{n=1}^\infty u_n^2 = \infty \), and LPP-condition denotes the condition on \( u_n \) and \( \theta \) in Theorem 3.C. The conditions on \( \theta \), \( \max\{u_r \}^{-2r} < 1 + \theta \), is used for showing \( E(D_n) \to \infty \) (see Proposition 4.1 in Levin et al., 2001). A weaker condition on \( \theta \) is also presented in Theorem 5.1 in Levin et al. (2001), but it is not used here. As pointed out in Remark for
Proposition 4.1 in Levin et al. (2001), the above conditions on \( \theta \) and \( \{u_n\}^\infty_{n=1} \) are not vacuous.

Phase transition phenomena in random coin tossing are closely related to discerning two sets of probability measures in time series. In the Kakutani's theorem, \( \mu_\theta \) and \( \mu_0 \) correspond to the non-stationary and the stationary measures, respectively. Accordingly, when an entire sequence \( \{X_i\}^\infty_{i=1} \) is observed, the mutual singularity between \( \mu_\theta \) and \( \mu_0 \) is equivalent to the DES between non-stationary and stationary measures. Also, Harris and Keane (1997) posed a conjecture that "in Theorem 3.B, the singularity of the two laws \( \mu_\theta \) and \( \mu_0 \) depends only on the renewal probabilities \( \{u_n\}^\infty_{n=1} \), not on \( \theta \)," which is equivalent to the conjecture on UDES between \( \mu_0 \) and the set of non-stationary measures satisfying the HK-condition. In addition, Theorem 3.C by Levin et al. (2001) disproved this conjecture and showed UDES between \( \mu_\theta \) and the set satisfying the LPP-condition. It is worth noting that the results presented in this section pertain to testing a simple null hypothesis against a composite alternative hypothesis. Extensions to a composite null hypothesis are discussed in the next section.

It can be easily shown that MHKC are not UDES from Lemma 2.1. In the following, Lemma 3.1 shows that MPD \( \subsetneq \) MHK, and Theorem 3.2 shows that MLPP is SD from \( \mu_0 \).

**Lemma 3.1.** For each \( \delta \in (0, 1) \), let \( S_\delta \) be the collection of sequences \( l = (t_1, t_2, \ldots) \) such that \( \lim_{n \to \infty} \frac{|l \cap \{1, \ldots, n\}|}{n} = \delta \). Then, (i) there exists a renewal process having support on \( S_\delta \), with \( u_n = P(\Delta_n = 1) \) satisfying \( \sum u_n^p = \infty \); and (ii) the converse is not true.

**Proof.** (i) is obvious from considering the IID binary sequences \( \{\Delta_n\}^\infty_{n=1} \) with \( P(\Delta_n = 1) = \delta \). Note that \( P(S_\delta) = 1 \) a.s. For (ii), it suffices to show that there exists a renewal process having a support on 0-density but satisfying the condition \( \sum u_n^2 = \infty \). Let \( \{S_n\}^\infty_{n=1} \) be the simple random walk on the plane and \( \{\Delta_n\}^\infty_{n=1} \) be the recordings of events returning to the origin. That is, \( \{\Delta_n = 1\} = \{S_n = 0\} \). Then, \( u_n = P(\Delta_n = 1) = O(1/\sqrt{n}) \), and \( p_n \) (the probability of the event that "the first return occurs at \( n \)) is \( O(1/n^{1/\delta}) \). Thus, the renewal process has a support on 0-density, but \( \sum u_n^2 = \infty \).

**Theorem 3.2.** MLPP is SD from \( \mu_0 \).

According to Lemma 2.1, MPD is also SD from \( \mu_0 \). However, for the consistency with Theorem 2 in Dembo and Peres (1994) (Theorem 5.A in this paper), a different proof is presented in Section 5, where the empirical measure plays a key role in building discerning functions \( \{f_n(X_1, \ldots, X_n)\}^\infty_{n=1} \).

**Corollary 3.3.** Let \( H_1 \) be the collection of probability measures \( \mu_\theta(l) \), where \( l = (t_1, t_2, \ldots, t_n, \ldots) \) has a
positive density. Then, \( H_1 \) is SD from \( H_0 = \{\mu_0\} \).

**Proof on SD of MLPP**

**Proof.** SD for the case \( \sum_{n=1}^{\infty} u_n^2 = \infty \) directly follows from Theorem 6.1 in Levin et al. (2001). On the other hand, to prove SD for \( \{u_n\}_{n=1}^{\infty} \) satisfying LPP-condition, it suffices to show that

(1) \( R_n \) is smaller than \( (1 + \delta) \cdot \log_2 n \) eventually \( \mu_0 - a.s \) and

(2) for every \( \mu \in \text{MLPP} - \text{MHK} \), \( R_n \) is larger than \( (1 + \delta) \cdot \log_2 n \) eventually \( \mu - a.s \),

where for every \( k \), \( L_k = \max \{ j : X_k = X_{k+1} = \cdots , X_{k+j-1} = 1 \} \) and \( R_1 = \max_{1 \leq k \leq n} L_k \). (1) is obvious from the Erdős and Rényi (1970) law, which tells \( \lim \sup_{n \to \infty} I_n = \log_2 n \ a.s \) under \( \mu_0 \). Thus, it suffices to show that, for every \( \mu \in \text{MLPP} - \text{MHK} \),

\[
\mu(R_n < (1 + \delta) \cdot \log_2 n \text{ infinitely often}) = 0. \tag{3.1}
\]

As in Levin et al. (2001), for every \( a \geq 1 \), let

\[
\Lambda^*(a) \equiv \lim_{m \to \infty} -\log_2 P \left[ T_1 + T_2 + \cdots + T_m \leq ma \right] / m, \tag{3.2}
\]

where \( T_i \) is the inter-arrival time between the \( i \)-th and the \((i - 1)\)-th renewal of \( \{\Delta_n\}_{n=1}^{\infty} \). Subsequently, define

\[
\psi_0 \equiv \inf \left\{ 0 < \xi \leq 1 : \tilde{\psi}(\varphi, \xi) = \psi(\xi) \right\},
\]

where

\[
\tilde{\psi}(\varphi, \xi) = \xi (\varphi - \Lambda^*(\xi^{-1})), \psi(\varphi) = \sup_{0 < \xi \leq 1} \tilde{\psi}(\varphi, \xi), \text{ and } \varphi = \log_2 (1 + \theta).
\]

It was shown at Section 5 in Levin et al. (2001) that the maximum of \( \tilde{\psi}(\varphi, \cdot) \) over \((0, 1]\) is attained; accordingly, \( \psi_0 \) is well-defined. Let \( k_n = (1 + \delta) \cdot \log_2 n \), \( I^n_i = [ik_n, (i + 1)k_n - 1] \cap \mathbb{Z}^+ \), and \( \tau_i = \tau^n_i \) be the time of the \([\psi_0 k_n]\)-th renewal after time \( ik_n - 1 \). Then, the event \( G^n_i \) of a good run in the block \( I^n_i \) occurs when

- there is a renewal at time \( ik_n : \Delta_{ik_n} = 1 \),
- there are at least \( \psi_0 k_n \) renewals in \( I_i : \tau_i \leq (i + 1)k_n - 1 \),
- until time \( \tau_i \), all observations are "heads" : \( X_j = 1 \) for \( ik_n \leq j \leq \tau_i \),
- the remainder observation in the block \( I_i \) are also "heads" : \( X_j = 1 \) for \( \tau_i < j \leq (i + 1)k_n - 1 \).
Also, let $D_n = \sum_j I(G^n_j)$, $\lambda_n = E(D_n)$, $R_{j,k_n}^{(j+1)k_n-1} = \max_{i \in [j,k_n, (j+1)k_n-1]} L_i$, $\alpha_i^n = E(G^n_i)$, and $N_i$ be the number of renewals at the $i$-th interval. The expectation is with respect to the measure $\mu_\theta$. Then,

$$P(R_n < k_n) = P\left(\max_{1 \leq k \leq n} R_k < k_n\right) \leq P\left(\forall j, \ 1 \leq j \leq \left\lfloor \frac{n}{k_n}\right\rfloor \ R_{j,k_n}^{(j+1)k_n-1} < k_n\right) = P\left(D_n = 0\right) \leq \exp(-\lambda_n) + e_1 + e_2,$$

where

$$e_1 = 2 \cdot \sum_{i=1}^{n/k_n} \alpha_i^n \cdot \alpha_i^n$$

$$e_2 = 2 \cdot \sum_{i=1}^{n/k_n} E\left\{ (I(G^n_i) - \alpha_i^n) \cdot (1/\lambda_n)^{V(i)+1} \right\}$$

Here, $V^{(i)} = \sum_{j \neq i} I(G^n_j)$ and Inequality (3.3) results from Equation (2.6) in Chen (1975) with $m = 0$.

$$e_1/2 = \sum_{i=1}^{n/k_n} \alpha_i^n \cdot \alpha_i^n = \sum_{i=1}^{n/k_n} 2^{-2k_n} (1 + \theta)^{2\psi_0 k_n} P\left(N_1 \geq \psi_0 k_n \mid \Delta_{k_n} = 1\right)^2 u_{ik_n}^2$$

$$\leq 2^{-2k_n} (1 + \theta)^{2\psi_0 k_n} 2^{-2\psi_0 k_n} \left(\Lambda^{*}(\psi^{-1}) - \epsilon\right) \sum_{i=1}^{n/k_n} u_{ik_n}^2$$

$$= \exp\left\{ -\log_2 n \left[ 2 + 2\delta + 2(1 + \delta)\psi_0 \left( \log_2 (1 + \theta) - \Lambda^* (\psi^{-1}) \right) \right] \right\} \cdot n^{1-2\gamma},$$

11
which is summable with respect to $n$. Inequality (3.4) is from Equation (3.2). On the other hand, it can be shown that

\[
e_2/2 = \sum_{i=1}^{n/k_n} E \left\{ (I(G^n_i) - \alpha^n_i) \cdot (1/\lambda_n)^{V^{(i+1)}} \right\}
\]

\[
= P(D_n = 0) \sum_{i=1}^{n/k_n} \left\{ -\alpha^n_i \frac{1}{\lambda_n} \right\} + \sum_{i=1}^{n/k_n} \left\{ (1 - \alpha^n_i) \frac{1}{\lambda_n} - \sum_{j \neq i} \alpha^n_j \frac{1}{\lambda_n^2} \right\} P(G^n_i, D_n = 1)
\]

\[
+ \sum_{i,j} \left\{ (1 - \alpha^n_i) \frac{1}{\lambda_n^2} + (1 - \alpha^n_j) \frac{1}{\lambda_n^2} - \sum_{k \neq i,j} \alpha^n_k \frac{1}{\lambda_n^3} \right\} P(G^n_i \cap G^n_j, D_n = 2)
\]

\[
+ \sum_{i,j,k} \left\{ (1 - \alpha^n_i) \frac{1}{\lambda_n^2} + (1 - \alpha^n_j) \frac{1}{\lambda_n^2} + (1 - \alpha^n_k) \frac{1}{\lambda_n^2} - \sum_{l \neq i,j,k} \alpha^n_l \frac{1}{\lambda_n^3} \right\} P(G^n_i \cap G^n_j \cap G^n_k, D_n = 2)
\]

\[+ \cdots
\]

\[
= -P(D_n = 0) + \left( \frac{1}{\lambda_n^2} - \frac{1}{\lambda_n^2} \right) \sum_{i,j} \alpha^n_i \cdot P(G^n_i, D_n = 1) + \frac{1}{\lambda_n^2} \sum_{i,j} P(G^n_i \cap G^n_j, D_n = 2)
\]

\[
+ \left( \frac{1}{\lambda_n^3} - \frac{1}{\lambda_n^3} \right) \sum_{i,j} (\alpha^n_i + \alpha^n_j) P(G^n_i \cap G^n_j, D_n = 2) + \frac{1}{\lambda_n^3} \sum_{i,j,k} P(G^n_i \cap G^n_j \cap G^n_k, D_n = 2)
\]

\[
+ \left( \frac{1}{\lambda_n^4} - \frac{2}{\lambda_n^4} \right) \sum_{i,j,k} (\alpha^n_i + \alpha^n_j + \alpha^n_k) P(G^n_i \cap G^n_j \cap G^n_k, D_n = 2) + \cdots
\]

\[
\leq -P(D_n = 0) + \sum_{k=1}^{n/k_n-1} \left\{ \left( \frac{1}{\lambda_n^{k+1}} - \frac{1}{\lambda_n^k} \right) \alpha^n_k P(D_n = k) + \frac{1}{\lambda_n^{k+1}} P(D_n = k+1) \right\},
\]

which is smaller than 0 from $\lambda_n \sim n^q$ for some positive $q$ (Levin et al. 2001). Therefore,

\[
\sum_n P(R_n < k_n) \leq \sum_n P(D_n = 0) \leq \sum_n \left\{ \exp(-\lambda_n) + e_1 \right\} < \infty.
\]

Hence, according to the Borel-Cantelli lemma, $R_n$ is larger than $(1 + \delta) \log_2 n$ eventually $\mu$-a.s. \Box

### 3.2 Extension of Random Coin Tossing to General Settings and Examples

The results on random coin tossing can be extended in two ways: an extension to a general distribution (beyond a binary distribution) and an extension to a composite null hypothesis.

First, suppose $F_1(\cdot)$ is the conditional distribution function of $X_n$ given $\Delta_n = 1$ and $F_0(\cdot)$ is that given $\Delta_n = 0$. Then, there exists a set $A \in R$ and $\theta > 0$ such that

\[
\int_A dF_1(x) = \frac{1 + \theta}{2} \quad \text{and} \quad \int_A dF_0(x) = \frac{1}{2}.
\]

Defining $Y_n = I(X_n \in A)$, we can directly apply Theorem 3.2 to prove that the MLPP with $F_1$ is SD from IID measure with $F_0$. 

12
In contrast, the extension to a composite null hypothesis is not obvious. Let $\mathcal{F}$ be the set of measures and $\{\Delta_n\}_{n=1}^{\infty}$ be a binary sequence. Suppose a non-stationary process $\{X_n\}_{n=1}^{\infty}$ is generated from $\{F_1, F_2, \ldots, F_k\} \subset \mathcal{F}$ if $\Delta_n = 1$, and from $F_0 \in \mathcal{F}$ otherwise. The collection of such $\{X_n\}_{n=1}^{\infty}$ is called non-stationary measures generated by $\mathcal{F}$ and denoted as $\text{NS}(\mathcal{F}, \{\Delta_n\}_{n=1}^{\infty})$ for notational simplicity. Then, $\text{NS}(\mathcal{F}, \{\Delta_n\}_{n=1}^{\infty})$ is SD from the stationary measure $F_0$, when $\{\Delta_n\}_{n=1}^{\infty}$ satisfies the PD-condition and $\mathcal{F}$ is a subset of the family satisfying the IDEN-condition that any finite number of measures $\{F_1, \ldots, F_m\} \in \mathcal{F}$ are linearly independent, in the sense that, for every $(a_1, \ldots, a_m) \in \mathbb{R}^m$, $\sum_{i=1}^{m} a_i F_i \neq 0$. Then, defining the sequence of disjoint closed sets $\{A_n, B_n\}_{n=1}^{\infty}$ in $M(\mathbb{R})$ as

$$A_n = \left\{ \nu \in M(\mathbb{R}); \nu = \delta_n F_1 + (1 - \delta_n) F_2, \exists F_1, F_2 \in \mathcal{F} \text{ and } 0 \leq \delta_n \leq \frac{1}{2n} \right\}$$

$$B_n = \left\{ \nu \in M(\mathbb{R}); \nu = \delta_n F_1 + (1 - \delta_n) F_2, \exists F_1, F_2 \in \mathcal{F} \text{ and } \delta - \frac{1}{2n} \leq \delta_n \leq \delta + \frac{1}{2n} \right\},$$

$H_0$ and $H_1$ are contained in disjoint sequential $F_\varepsilon-$sets. Hence, Theorem 2 in Dembo and Peres (1994) (Theorem 5.A. in this paper) can be applied. Note that the IDEN-condition holds for Poisson, Gaussian with a fixed variance, exponential, and binomial (with a sufficient number of trials) distributions.

**Example A** Let $M \in \mathbb{R}^{d \times d}$ be the set of all non-negative, irreducible $d \times d$ Markov transition matrices and $S(P_0, P_1, \{\Delta_n\}_{n=1}^{\infty})$ be the probability measure associated with the process $\{X_n\}_{n=1}^{\infty}$. Here, the transition probability $X_n$ to $X_{n+1}$ is $P_0$ (or $P_1$), when $\Delta_n = 0$ (or $\Delta_n = 1$).

Suppose $\{\Delta_n\}_{n=1}^{\infty}$ satisfies the LPP condition and $||P_0 - P_1|| = \max_i |P_0(i, i) - P_1(i, i)| > \theta$, where $P(i, j)$ denotes the $(i, j)$-th component of $P$. Without loss of the generality, let the first diagonal element of $P_0$, say $p$, be larger than that of $P_1$, say $q$. Subsequently, let $L_k = \max\{ j : X_k = X_{k+1} = \ldots, X_{k+j-1} = 1 \}$ and $R_n \geq \max_{1 \leq j \leq n} L_k$. Then, according to Corollary 2 in Foulser and Karlin (1987), $\limsup_{n \to \infty} L_n = \log_{1/p} n \text{ a.s.}$, when $\{X_n\}_{n=1}^{\infty}$ is from the stationary MC with $P_0$. Accordingly, $R_n$ is smaller than $(1 + \delta) \cdot \log_{1/p} n$ eventually $\mu_0-$a.s. where $\mu_0$ is the measure for the MC with $P_0$.

On the other hand, define a new renewal process $\{\Delta'_n\}_{n=1}^{\infty}$ as $\Delta'_n = I(X_n = 1, \Delta_n = 1)$, whose renewal probability $u'_n = P(\Delta'_n = 1)$ has the same order as that of $\{\Delta_n\}_{n=1}^{\infty}$; i.e. $u'_n \sim u_n$. Then, using the same procedures in the proof of Theorem 3.2, it can be shown that $\mu(R_n < (1 + \delta) \cdot \log_{1/p} n \text{ i.o.}) = 0$, where $\mu \in \text{NS}(P_0, P_1, \{\Delta_n\}_{n=1}^{\infty})$. Thus, the collection of non-stationary measures $\text{NS}(P_0, P_1, \{\Delta_n\}_{n=1}^{\infty})$ with the LPP-condition is SD from the stationary MC with $P_0$.

**Example B** Suppose $Q(\rho_0, \rho_1, \{\Delta_n\}_{n=1}^{\infty})$ ($\rho_0$ and $\rho_1$ unknown) is the probability measure associated with
the process \( \{X_n\}_{n=1}^\infty \), where, given \( X_n \) and \( \Delta_n = 0 \) (or \( \Delta_n = 1 \)), \( X_{n+1} = \rho_0 \cdot X_n + \varepsilon_n \) (or \( X_{n+1} = \rho_1 \cdot X_n + \varepsilon_n \)). \( \varepsilon_n \) is from a distribution satisfying the IDEN-condition. Then, by applying Corollary 3.3 to \( \{(Y_n, \Delta_n)\} \) with \( Y_n = (X_n, X_{n+1}) \), the stationary AR process is SD from the collection of measures \( Q(\rho_0, \rho_1, \{\Delta_n\}_{n=1}^\infty) \) with \( \{\Delta_n\}_{n=1}^\infty \) satisfying the PD-condition. When \( (\rho_0, \rho_1) \) is known, the class of non-stationary measures can be extended to that associated with \( \{\Delta_n\}_{n=1}^\infty \) satisfying the LPP-condition.

### 4 Detecting Sparse Heterogeneous Mixtures

Recall that \( \mu_0 \) denotes fair coin tossing and HM(\( \gamma, \theta \)) is the collection of the binary processes for which \( \Delta_n \) is independently from Bernoulli(\( n^{-\gamma} \)). Again, an independent coin with a bias \( \theta > 0 \) (\( P(\text{Head}) = (1 + \theta)/2 \)) is tossed when \( \Delta_n = 1 \), a fair coin is tossed when \( \Delta_n = 0 \). In HM(\( \gamma, \theta \)), the average portion of biased coins in the first \( n \) draws is \( n^{-\gamma} \) and, thus, a finite dimensional empirical measure does not deviate from that of the fair coin tossing when \( n \to \infty \).

**Theorem 4.A.** HM(\( \theta, \gamma \)) is neither DES, UDES, nor SD from \( \mu_0 \) when \( \gamma \in (0.5, 1) \)

**Proof.** Since SD \( \Rightarrow \) UDES \( \Rightarrow \) DES from their definitions, it suffices to show that HM(\( \theta, \gamma \)) is not DES from \( \mu_0 \) when \( \gamma \in (0.5, 1) \). Suppose \( X_n \) is the observation from HM(\( \theta, \gamma \)). It has the same law as independent random coin tossing \( Y_n \) with \( Y_n \sim \text{Bernoulli}((1 + \varepsilon_n \theta)/2) \). The bias of the \( n \)-th coin (denoted by \( \theta_n \)) is \( n^{-\gamma} \theta \) and \( \sum_{n=1}^\infty \theta_n^2 < \infty \). Thus, HM(\( \theta, \gamma \)) is not DES from \( \mu_0 \) from Theorem 3.A.

**Theorem 4.1.** HM(\( \theta, \gamma \)) and \( \mu_0 \) is SD when \( 0 < \gamma < 0.5 \) and UDES when \( \gamma = 0.5 \)

**Proof.** Let \( S_n = \sum_{i=1}^n X_i \). In fair coin tossing, it can be shown from the law of the iterated logarithm that \( S_n \) is eventually smaller than \( (1 + \varepsilon)\sqrt{2n \log \log n} \), \( \mu_0 \)-a.s. Thus, it suffices to show that, for every \( \mu \in \text{HM}(\gamma, \theta) \),

\[
\mu \left\{ S_n > (1 + \varepsilon)\sqrt{2n \log \log n} \text{ eventually} \right\} = 1, \tag{4.5}
\]

To show (4.5), under \( \mu_1 \),

\[
(4.5) = P(S_n < (1 + \varepsilon)\sqrt{2n \log \log n})
= P\left(\{a_n - S_n\} < \{a_n - (1 + \varepsilon)\sqrt{2n \log \log n}\}\right) \tag{4.6}
\leq \text{Const.} \cdot \exp \left\{ -2(a_n - (1 + \varepsilon)\sqrt{\log \log n})^2/n \right\} \tag{4.7}
= \text{Const.} \cdot \exp \left\{ -2 \cdot n^{1-2\gamma} \right\}, \tag{4.8}
\]

14
where $a_n = E(S_n) = \sum_{k=1}^{n} k^{-\gamma} \sim n^{1-\gamma}$, and (4.7) is from the Hoeffding inequality for the sum of independent bounded random variables (Hoeffding 1963). Since (4.7) has a finite sum for $\gamma \in (0, 0.5)$, $S_n$ is larger than $(1 + \epsilon)\sqrt{2n\log \log n}$ eventually from the Borel-Cantelli lemma. Finally, HM($\gamma, \theta$) is SD from $\mu_0$ with discerning functions $f_n(X_1, \cdots, X_n) = I(S_n > (1 + \epsilon)\sqrt{2n\log \log n})$. On the other hand, when $\gamma = 0.5$, HM($\theta, \gamma$) can be shown to be UDES from $\mu_0$ using the martingale arguments in Theorem 1 of Harris and Keane (1997). As pointed out in Remark 1 of Harris and Keane (1997), the mutually singularity between HM($\theta, \gamma$) and $\mu_0$ is from the fact that

$$P \otimes P(\Delta_k \Delta_k' = 1 \text{ for infinitely many } k) = 1,$$

where $\Delta_k'$ is an independent copy of $\Delta_k$. Equation (4.9) is true from

$$P \otimes P(\Delta_k \Delta_k' = 1) = k^{-2\gamma}$$

and the Borel Cantelli lemma. □

The results on random coin tossing can directly be extended to a general distribution (beyond a binary distribution). First, suppose $F_1(\cdot)$ is the conditional distribution function of (the observation) $X_n$ given $\Delta_n = 1$ and $F_0(\cdot)$ is the conditional distribution function given $\Delta_n = 0$. Then, without loss of generality, we can find a set $A \in R$ and $\theta > 0$ such that

$$\int_A \frac{1 + \theta}{2} \text{ and } \int_A dF_0(x) = \frac{1}{2}.$$

Defining $Y_n = I(X_n \in A)$, we can directly apply Theorem 2.A to prove that HM($\theta, \gamma$) is SD from $\mu_0$ when $\gamma \in (0, 0.5)$ and $(1 + \theta^2) \cdot \alpha(\gamma) < 1$. Theorem 4.1 can be extended similarly.

5 Test on Independence and Ergodicity

This section discusses testing independence and ergodicity. It can be shown that violating independence is a type of non-stationarity. Accordingly, when testing independence, the results in the previous section can be used directly. In addition, a negative result for testing ergodicity is constructed.

Suppose an entire process $\{X_n\}_{n=1}^{\infty}$ is partitioned into independent sub-blocks as

$$\prod_{k=1}^{\infty} \left(X_{t_k+1}, \cdots, X_{t_k+d_k}\right) \prod_{m \in \text{others}} (X_m).$$

(5.10)
Then, \( \{X_n\} \) is called stationary dependent, if every k-dimensional independent component has the same marginal distribution \( P_n \), which is an n-dimensional marginal distribution of a stationary measure on \( R^{2^k} \), say \( Q \); otherwise, it is non-stationary dependent. Accordingly, \( l = (t_1, t_2, \cdots, t_n, \cdots) \), \( d = (d_1, d_2, \cdots, d_n, \cdots) \), and \( Q \) defines the product measure \( \mu(l, d, Q) \). In the followings, \( \forall k, d_k = 2 \) and \( P(X_{t_k}, X_{t_k+1}) \neq P(X_{t_k}) \cdot P(X_{t_k+1}) \) are assumed for notational simplicity.

**Theorem 5.1.** Suppose \( \alpha \) is the support of a renewal process \( \{\Delta_n\}_{n=1}^{\infty} \) satisfying the LPP-condition. Then, the collection of measures \( \mu(l, d, Q) \) with \( l \in \alpha \) (\( Q \) is known) is SD from the independent product measure \( P_1 \otimes P_1 \otimes \cdots \), where \( P_1 \) is the 1-dimensional marginal distribution of \( Q \).

**Proof.** Let \( A \in B^2 \) be the set satisfying \( (P_1 \otimes P_1)(A) = 1/2 \) and \( P_2(A) > (1 + \theta)/2 \) for some \( \theta > 0 \). \( P_2 \) is the 2-dimensional marginal distribution of \( Q \). Subsequently, let \( Y_n = 1((X_n, X_{n+1}) \in A) \). Then, the SD can be shown from Theorem 3.2 with \( \{(Y_n, \Delta_n)\}_{n=1}^{\infty} \).

When \( Q \) is unknown, the LPP-condition in Theorem 5.1 should be strengthened to PD-condition. Also, as in Example A in the previous section, the results can be easily extended to finite MC.

Unlike stationary dependence, detecting non-stationary dependence may be challenging due to the generality of non-stationary dependence. However, having additional information on the dependent structure reduces the set of probability measures to be considered and allows some positive results. For example, let \( Z_k \) be IID random variables from Bernoulli(1/2) and \( P_n \) be the measure for the binary sequence \((Y_1, \cdots, Y_n)\) with

\[
Y_k = \prod_{i=1}^{n_k} Z_i / Z_k.
\]

Subsequently, define a non-stationary dependence measure NSD(l) as the measure of the sequence

\[
\{(Y_1, \cdots, Y_{k_1}), (Y_{k_1+1}, \cdots, Y_{k_1+k_2}), \cdots, (Y_{\sum_{i=1}^{n_l-1} k_i+1}, \cdots, Y_{\sum_{i=1}^{n_l} k_i}), \cdots\},
\]

where \( l = (k_1, k_2, \cdots) \). Then, \( U_n = \prod_{i=1}^{n_l} Y_i \) follows independent Bernoulli(1/2) if \( n_l \neq \sum_{i=1}^{l} k_i \), for some \( l \). Otherwise, \( U_n = 1 \). Thus, using Theorem 3.2, it can be shown that the collection of NSD(l) with \( l \) satisfying LPP condition is SD from \( \mu_0 \) (the fair coin tossing in Section 3).

Finally, let \( Q \) be a stationary non-ergodic measure. Then, according to the ergodic decomposition, \( Q \) can be represented as a convex combination of (stationary) ergodic measures. Without loss of generality, \( Q = \lambda \cdot P_1 + (1 - \lambda) \cdot P_2 \) for some \( \lambda \in (0, 1) \). Suppose \( Q \) and \( \{P_1, P_2\} \) are SD with the sequence of discerning
functions \( \{f_n(x_1, x_2, \cdots, x_n)\}_{n=1}^{\infty} \): \( f_n \) converges to 1, \( Q \)– a.s and converges to 0, \( P_1 \) and \( P_2 \)– a.s. Then,

\[
1 = \lim_{n \to \infty} Q \left( \{f_n(X_1, X_2, \cdots, X_n) = 0\} \right) \\
= \lim_{n \to \infty} \lambda \cdot P_1 \left( \{f_n(X_1, X_2, \cdots, X_n) = 0\} \right) + (1 - \lambda) \cdot P_2 \left( \{f_n(X_1, X_2, \cdots, X_n) = 0\} \right) \\
= 0,
\]

which is not true. Thus, we can conclude that the class of all stationary ergodic measures is not SD even from a single fixed stationary non-ergodic measure \( Q \).

6 General Criterion for Hypothesis Testing in Time Series

This section extends the following theorem of Dembo and Peres (1994) to time series settings:

**Theorem 6.A.** Let \( H_0 \) and \( H_1 \) be two disjoint families of probability measures on \( \mathbb{R}^d \). Then, if \( H_0 \) and \( H_1 \) are contained in disjoint \( F_\alpha \)–sets of measures, then they are distinguishable. The converse holds under the assumption that all measures in \( H_0 \cup H_1 \) are absolutely continuous with respect to the Lebesgue measure and, for each \( \mu \in H_0 \cup H_1 \), there is some \( p > 1 \) such that the density of \( \mu \) is in \( L^p(\mathbb{R}^d) \).

When the sequential observations are IID from a fixed distribution, testing \( \mu \in H_0 \) and \( \mu \in H_1 \) has been addressed by many authors. Hoeffding and Wolfowitz (1957) explored the various notions of discernibility in terms of the sample size function \( N \) and the decision function \( \phi(X_1, \cdots, X_n) \). \( N \) is a stopping time with respect to \( \sigma(X_1, \cdots, X_n) \). Subsequently, they showed that a sufficient condition for discernibility between \( H_0 \) and \( H_1 \) is that they 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 a variational metric on probability measures. Here, a KS-open set is the topology induced by the Kolmogorov-Smirnov distance between distribution functions. Extending Hoeffding and Wolfowitz (1958), Le Cam and Schwartz (1959) proposed a necessary and sufficient condition for the consistent estimation of real valued parameters. Subsequently, Fisher and Van Ness (1969) provided a necessary and (or) sufficient condition in terms of a variety of measure metrics. However, all these conditions involve all \( n \)-fold product measures obtained from measures in \( H_0 \cup H_1 \) and are thus difficult to verify. Cover (1973) considered distributions on \([0,1]\) with \( A \in R_{[0,1]} \) (more generally, the case of countable \( A \)), the set of rational numbers in \([0,1]\), and provided a test which would commit only a finite number of mistakes for every \( A \) and \( A^c \cup N \), where \( N \) is a set of Lebesgue measure 0. Recently, several authors extended his results to general metric spaces beyond \( \mathbb{R} \) (with the Lebesgue measure), and proposed simple topological (structural)
conditions for discerning \( H_0 \) and \( H_1 \). Dembo and Peres (1994) constructed a simple topological criterion using the metric discernibility of two disjoint increasing sequences of closed sets and the uniform convergence of empirical measures (with an exponential rate). Similarly, Kulkarni and Zeitouni (1996) proposed a procedure which leads a.s. to a correct decision for any \( H_0 \) satisfying certain topological (structural) assumptions; the input sequences are parsed into subsequences whose empirical measures fall within certain error bounds. The decision rules of Cover (1973), Dembo and Peres (1994), and Kulkarni and Zeitouni (1996) are fundamentally as follows. Suppose \( \{F_m\}_{m=1}^\infty \) and \( \{G_m\}_{m=1}^\infty \) are a sequence of hypotheses increasing to \( H_0 \) and \( H_1 \), respectively. Then, for each \( m \), we will have enough data in forming the empirical measure to make the probability of an incorrect decision between \( F_m \) and \( G_m \) less then \( 1/m^2 \), so that only finitely many errors are committed. Accordingly, the topological (structural) conditions on \( H_0 \) and \( H_1 \) are sufficient for discernibility from the uniform convergence of empirical measures with IID samples. However, constructing necessity conditions is much more challenging that constructing sufficient conditions.

Let \( M(R^{2+}) \) be a collection of all probability measures on \( R^{2+} \) which is equipped with the weak topology with basic open sets

\[
U_{\delta \phi \nu} = \left\{ \nu \in M(R^{2+}) : \left| \int_{R^{2+}} \phi d\nu - x \right| < \delta \right\},
\]

where \( \phi \in C_b(R^{2+}) \) (=continuous bounded function on \( R^{2+} \)), and \( \delta > 0 \). Note that, since \( M(R^{2+}) \) is Polish, the weak topology is well defined. Furthermore, the Lipschitz metric is defined as

\[
\rho(\mu, \nu) = \sup_{f \in BL_1} \left| \int_{\Sigma} f d\mu - \int_{\Sigma} f d\nu \right|,
\]

where \( BL_1 \) is the class of bounded Lipschitz continuous functions \( f : \Sigma \rightarrow R \) with a Lipschitz constant of at most one. \( \Sigma \) is \( R^{2+} \) for \( \mu, \nu \in M(R^{2+}) \) and is \( R^m \) for \( \mu, \nu \in M(R^m) \). Let \( \mathcal{E} \) be the set of measure \( \mu \in M(R^{2+}) \) satisfying that, for every \( k \), the empirical measures of \( k \)-tuples converge with exponential rate to \( \mu^* \in M(R^k) \) (asymptotically stationary measures); the rate can depend on \( m \). Let \( H_0 \) and \( H_1 \) be disjoint sets of measures in \( \mathcal{E} \).

**Theorem 6.1.** Suppose there exists sequences \( \{A_m\}_{m=1}^\infty \) and \( \{B_m\}_{m=1}^\infty \) in \( M(R^m) \) such that (i) \( A_m \) and \( B_m \) are disjoint closed sets, and (ii) \( H_0 = \bigcup_{m=1}^\infty \{ \mu \in \mathcal{E} : \mu^*_m \in A_m \} \) and \( H_1 = \bigcup_{m=1}^\infty \{ \mu \in \mathcal{E} : \mu^*_m \in B_m \} \). Then, \( H_0 \) and \( H_1 \) are SD.

**Proof.** Let the space \( M(R^{2+}) \) be equipped with the Lipschitz metric \( \rho \) and let \( \{X_i\}_{i=1}^\infty \) be from \( \mu \in H_0 \cup H_1 \in \mathcal{E} \).
Then, according to the definition of $\mathcal{E}$, for any $\epsilon > 0$ and $m$, there exist $K(\epsilon)$ and $\delta(\epsilon)$ such that

$$P\left(\rho(\nu_{n,m}, \mu^*_m) \geq \epsilon\right) \leq K(\epsilon)e^{-nd^2(\epsilon)/2},$$

where $K(\epsilon)$ and $\delta(\epsilon)$ are independent of $\mu$, but can be dependent on $m$. Here, $\nu_{n,m}$ is the $m$-dimensional empirical measure for $\mu^*_m$ from $\{X_i\}_{i=1}^n$. Then, as shown in Dembo and Peres (1994), there exists a sequence $\{n_m(k)\}_{k=1}^\infty$ such that, when $n > n_m(k)$,

$$P\left(\rho(\nu_{n,m}, \mu^*_m) \geq 1/k\right) < \eta_n \quad \text{uniformly in } \mu \in A_m \cup B_m,$$

(6.11)

where $\eta_n$ is summable. Furthermore, the testing function $g_m^n(x_1, \ldots, x_n)$ is $I(\nu_{n,m} \in \mathcal{N}(B_m, \epsilon_n))$, where $\mathcal{N}(B_m, \epsilon_n) = \{\nu_m : \nu_m \in \rho(\mu_m, \mu_m) < \epsilon_n"}, for some $\mu_m \in B_m$. Here, $\epsilon_n = 1/k$ when $n_m(k) \leq n < n_m(k+1)$.

Now, we may construct testing functions for $H_0$ and $H_1$. Let $n^*_k = \max\{n_m(k) : m \leq k\}$ and the index sets $I_n = \{k : n^*_k < n\}$ which increases slowly. It will be shown that the functions

$$f_n(x_1, \ldots, x_n) = I(\nu_{n,m} \in \bigcup_{m \in I_n} \mathcal{N}(B_m, \epsilon_n))$$

(6.12)

sequentially discern $H_0$ and $H_1$. First, suppose $\mu \in B_m \subset H_1$ and $\epsilon_n^* = 1/k$ when $n^*_k \leq n < n^*_k+1$. Then, for $n > n^*_m$

$$P\left(f_n(x_1, \ldots, x_n) = 0\right) \leq P\left(\rho(\nu_{n,m}, \mu^*_m) \geq \epsilon_n\right) \leq P\left(\rho(\nu_{n,m}, \mu^*_m) \geq \epsilon_n^*\right) \leq \eta_n.$$

(6.13)

Here, $\eta_n$ is that corresponding to $\epsilon_n$ in the above, and is summable with respect to $n$. Thus, $f_n(x_1, \ldots, x_n)$ equals 1 eventually from the Borel-Cantelli Lemma. When $\mu \in A_m \subset H_0$, Equation (6.13) is true for all $n > n^*_m$ with the same $\{\epsilon_n\}$. Hence, $f_n(x_1, \ldots, x_n)$ converges to 0 a.s.

When $\mathcal{E}$ is the set of stationary and uniformly ergodic measures introduced in Nobel (2003), $\mu^*_m$ is the m-dimensional marginal probability measures of $\mu$. Accordingly, $U_m = \{\mu \in \mathcal{E} : \mu^*_m \in A_m\}$ and $V_m = \{\mu \in \mathcal{E} : \mu^*_m \in B_m\}$ are closed sets in $\mathcal{E}$. Therefore, Theorem 5.1 becomes a special case of (may be equivalent to) Theorem 1 in Nobel (2003). However, $\mathcal{E}$ can be beyond the class of stationary and ergodic measures, as can be seen from Corollary 3.3. In the remainder of the Section, the corollary is proven in an alternative way, where Theorem 5.1 is applied to the class of non-stationary measures with PD condition.
Alternative Proof of Corollary 3.3

Proof. Let the space of probability measure $M_1(R)$ be equipped with the Lipschitz metric $\rho$, compatible with convergence in law. Let $X_1, X_2, \cdots, X_n, \cdots$ be independent observations from $\mu \in M(R)$. Letting $L_n = (1/n) \sum_{i=1}^{n} \delta_{X_i}$ denote the n-th empirical measure, Dembo and Peres (1994) showed that there exists a sequence $\{\epsilon_n\}_{n=1}^{\infty}$ satisfying $\rho(L_n, \mu) < \epsilon_n$ for all sufficiently large $n$'s.

First, assume $X_1, X_2, \cdots, X_n, \cdots$ are from $\mu_{\theta}(l) \in H_1$, where $l$ has a positive density $\delta$. Let $\nu_{\theta}(\delta)$ denote the marginal probability measure $\delta \mu_{\theta} + (1 - \delta) \mu_0$ and $\eta_n$ be $\epsilon_n + \left(\frac{\ln(1, \cdots, n)}{n} - \delta\right)$. Then,

$$P\left(\rho(L_n, \nu_{\theta}(\delta)) \geq 2\eta_n\right) = P\left(\sup_{f \in BL_1} \left| \left(\frac{1}{n}\right) \sum_{i=1}^{n} f(X_i) - \delta \mu_{\theta}(f) - (1 - \delta) \mu_0(f) \right| \geq 2\eta_n\right)$$

$$\leq P\left(\sup_{f \in BL_1} \left| \frac{1}{n} \sum_{i \in A_1} f(X_i) - \mu_{\theta}(f) \right| \geq \epsilon_n\right)$$

$$+ P\left(\sup_{f \in BL_1} \left| \frac{1}{n} \sum_{i \in A_2} f(X_i) - \mu_0(f) \right| \geq \epsilon_n\right), \quad (6.14)$$

where $A_1 = l \cap \{1, 2, \cdots, n\}$, $A_2^n = \{1, 2, \cdots, n\} - A_1$, and $BL_1$ is the class of Lipschitz functions with a Lipschitz constant one and a uniform bound of one. According to the proof of Theorem 2 in Dembo and Peres (1994), (6.14) is summable and

$$\rho(L_n, \nu_{\theta}(\delta)) < 2\eta_n \quad \text{for all sufficiently large } n. \quad (6.15)$$

Letting two sequences of closed sets $A_n$ and $B_n$ in $M(R^n)$ be $A_n = \{\nu \in \Sigma(R^{Z}) : \rho(\nu, \mu_0) < \eta_n\}$ and $B_n = \{\nu \in \Sigma(R^{Z}) : \rho(\nu, \mu_{\theta}) < \eta_n\}$, (a version of) Theorem 6.1 concludes the proof. Alternatively, define a discerning function $g(X_1, X_2, \cdots, X_n)$ as $I(\rho(L_n, \mu_0) > 3\eta_n)$. Then, when $\{X_n\}_{n=1}^{\infty}$ is from $\mu_0$, $\lim_{n \to \infty} g(X_1, X_2, \cdots, X_n) = 0$ by Theorem 2 in Dembo and Peres (1994). Also, when $\{X_n\}_{n=1}^{\infty}$ is from $\mu_{\theta}(l)$, $\lim_{n \to \infty} g(X_1, X_2, \cdots, X_n) = 1$ by Inequality (6.15). \hfill \Box

7 Discussion

We conclude the paper with two interesting unresolved questions. First, whether or not $HM(\gamma, \theta)$ is SD from fair coin tossing on the boundary $\gamma = 0.5$ is still unclear. Second, Theorem 3.2 and 4.1 only address a simple null hypothesis ($H_0$ has one element) and its extendibility to a composite null hypothesis has not been answered yet.
Acknowledgment

I am grateful to my advisor Amir Dembo for many helpful comments, and Andrew Nobel for providing me the preprint “Hypothesis testing for ergodic processes”. I also thank Eugenio Regazzini for his encouragement and Dylan Small for comments on style.

References


