

## STATIONARITY AS A PATH PROPERTY

BY

YI SHEN\* (WATERLOO) AND TONY S. WIRJANTO (WATERLOO)

*Abstract.* Traditionally, stationarity refers to shift invariance of the distribution of a stochastic process. In this paper, we rediscover stationarity as a path property instead of a distributional property. More precisely, we characterize a set of paths, denoted by  $A$ , which corresponds to the notion of stationarity. On one hand, the set  $A$  is shown to be large enough, so that for any stationary process, almost all of its paths are in  $A$ . On the other hand, we prove that any path in  $A$  will behave in the optimal way under any stationarity test satisfying some mild conditions.

**2010 AMS Mathematics Subject Classification:** Primary: 60G10; Secondary: 60G17, 62M10.

**Key words and phrases:** Stationarity, path property.

### 1. INTRODUCTION

Stationarity is an important notion in probability and statistics. Let us assume that  $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}_0}$  is a discrete-time stochastic process, where  $\mathbb{N}_0$  stands for the set of all non-negative integers.  $\mathbf{X}$  is called (*strictly*) *stationary* if all its finite-dimensional distributions are invariant under translation, in other words, if  $(X_{n_1}, \dots, X_{n_k}) \stackrel{d}{=} (X_{n_1+m}, \dots, X_{n_k+m})$  for all  $k, m \in \mathbb{N}$  and  $n_1, \dots, n_k \in \mathbb{N}_0$ . This definition shows that stationarity is, by its very nature, a distributional property of a stochastic process.

In this paper, we make an attempt to understand stationarity from a more path-oriented perspective. The motivation for and background of this effort come from time series analysis, where several stationarity tests for time series exist. These tests are different from their counterparts for stochastic processes in general, where a number of independent or correlated paths are often available. For time series, typically only one path is available, and all of the conclusions about the time series must be drawn based on the information extracted from this single path. Thus,

---

\* This work is partially supported by NSERC grant No. 469065.

in some sense, stationarity tests for time series naturally transform stationarity from a distributional property into a path property, with each particular stationary test dividing the path space into a “stationary/acceptance region” and a “non-stationary/rejection region”.

Logically, a stationarity test for time series should capture some “essential” properties satisfied by “typical” (e.g., almost all) paths of stationary processes, and it should be used to verify whether the given path has this property. Equivalently, the test can also be used to verify the existence of some traits which should be essentially absent in a stationary process, and utilize this result as a basis to reject the null hypothesis of stationarity. The critical question is, what properties are deemed to be “essential” in distinguishing between stationarity and non-stationarity, and whether we will obtain the same result for a given path when different properties are used for evaluation?

In principle, any property which is satisfied by all of the stationary processes with a higher probability than the non-stationary processes, or the opposite, should work. There are so many of them, so that it seems to be hopeless to come up with a clear idea about how such a property should look like. On the other hand, interestingly, it seems that we have a relatively clear notion about which paths are “stationary”, or more precisely, which are not. Let us consider the following examples:

Let  $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}_0}$  be a time series, and  $H$  be the path space  $\mathbb{R}^{\mathbb{N}_0}$  equipped with the cylindrical  $\sigma$ -field.

**EXAMPLE 1.1.** If  $\mathbf{x} = \{x_n\}_{n \in \mathbb{N}_0}$  is strictly increasing, then the corresponding time series should not be stationary, since  $P(\mathbf{X} \text{ is strictly increasing}) = 0$  for any stationary time series  $\mathbf{X}$ .

**EXAMPLE 1.2.** If there exists  $k$  such that  $x_k > \sup_{i \in \mathbb{N}_0, i \neq k} x_i$ , then the time series should not be stationary. Intuitively, with probability one, a stationary time series does not have a peak which is never attainable again.

Given the above examples, it might be tempting to argue that since each path is special in a certain sense, it will be rejected for stationarity by some tests. In other words, the abundance of the criteria which can be used for stationarity will result in an empty intersection for their acceptance regions in the path space. If this is the case, then stationarity should not be considered as a path property, because it means that the result of a stationarity test for a given path solely depends on the properties upon which the test is constructed. This, however, turns out not to be the case. In fact, there exist paths which should not be excluded from stationarity in any case, as shown by the following examples.

**EXAMPLE 1.3.** Let  $\mathbf{x} = (c, c, \dots)$  be a sequence of constants  $c \in \mathbb{R}$ . Then one should not conclude that  $\mathbf{x}$  is not stationary. Actually, if a stationarity test rejects such a path, then for this constant stationary process, its type I error will be one.

EXAMPLE 1.4. Let  $\mathbf{x} = (x_0, x_1, \dots)$ , where  $x_n = \sin(n\theta + \varphi_0)$ ,  $n \in \mathbb{N}_0$ . This is a wave with period  $2\pi/\theta$  and phase  $\varphi_0$ , observed at integer times. Notice that if we make  $\varphi_0$  to be random and uniformly distributed on  $[0, 2\pi)$ , then  $\mathbf{x}$  becomes a stationary process. Therefore, if we consider that all of the phases are equal in determining whether the path  $\mathbf{x}$  is stationary, which seems to be an irrefutable argument, then such  $\mathbf{x}$  should not be rejected for stationarity when tested. This example extends to all of the periodic functions observed at integers.

The examples above show how a strong, intuitive distinction between stationary and non-stationary paths exists in our mind, which enables us to tell the non-stationary paths from the stationary ones even before we venture into finding an appropriate set of criteria to discriminate them. Thus, such an intuitive distinction should be built upon some principles more fundamentally than the numerous specific path properties such as monotonicity, the number of peaks, etc.

The goal of our paper is to flash out these principles. In particular, we point out that there are three conditions which determine if a given path should be classified as “stationary”. Roughly speaking, the first condition requires that for any event of a certain type, if it happens once, it must happen infinitely many times along the path, with a positive limiting frequency; the second condition is a tightness condition which prevents any non-negligible part of the path from escaping to infinity; and the third condition is more of a technical nature, and is related to ergodicity.

The three conditions mentioned above identify a set of paths, denoted by  $A$ . We show that this is exactly the set of all of the paths which should be classified as “stationary”. We firstly prove that the set  $A$  is large enough, such that it contains almost all of the paths of any stationary process; then we show that the set  $A$  is also small enough, such that it only includes those paths which yield the best possible results under any given stationarity test. Thus, this justifies the idea that the notion of stationarity can be transformed profitably into a path property, and that the path space can be divided into an “essentially stationary” part and its complement. These results also show how the three proposed conditions can usefully serve as a basis for our intuition about the distinction between stationarity and non-stationarity, and provide a unified framework to understand and assess the existing stationarity tests.

We would like to emphasize that the results obtained in this paper should be understood in a statistical context, with time series tests serving as the background. They are not intended to replace the current definition of stationarity as a distributional property. Rather, based on this definition and existing probabilistic and ergodic theoretical results, the results in this paper are intended to offer a new perspective, which helps to clarify what is indeed being tested in the time series tests for stationarity, and to show that there is a limited number of ways in which these tests can behave.

The rest of the paper is organized as follows. In Section 2 we introduce the basic set-up and construct the set  $A$  of all the “stationary” paths. Section 3 shows

that the set  $A$  is large enough to contain almost all of the paths for any stationary process. A practical criterion to check one of the conditions that defines  $A$  is also established. Finally, in Section 4 we prove that  $A$  is also small enough, so that any path in  $A$  will be statistically indistinguishable with a typical path of a certain stationary process, in the sense that it will behave optimally under any stationarity test satisfying some mild conditions.

## 2. BASIC SET-UP

Let  $\mathbf{x} = \{x_n\}_{n \in \mathbb{N}_0}$  be a numerical sequence in  $\mathbb{R}$ . For  $k \in \mathbb{N}$ , consider  $\mathbf{I} = I_0 \times \dots \times I_{k-1} \in \mathcal{I}^k$ , where  $\mathcal{I}$  is the collection of open intervals on the real line. Define a set  $S_k^{\mathbf{I}} = S_k^{\mathbf{I}}(\mathbf{x})$  of non-negative integers by

$$S_k^{\mathbf{I}}(\mathbf{x}) := \{n \geq 0 : x_n \in I_0, \dots, x_{n+k-1} \in I_{k-1}\}.$$

Denote by  $N_k^{\mathbf{I}} = \{N_k^{\mathbf{I}}(n)\}_{n \in \mathbb{N}}$  the counting function of  $S_k^{\mathbf{I}}$ . That is,

$$N_k^{\mathbf{I}}(n) = |S_k^{\mathbf{I}} \cap [0, n-1]|,$$

where  $|\cdot|$  for a set gives the number of elements in the set. Note that for both  $S_k^{\mathbf{I}}$  and  $N_k^{\mathbf{I}}$ ,  $k$  is determined by  $\mathbf{I}$ , hence the only independent parameter would be  $\mathbf{I}$ . Here we also include  $k$  in the notation to indicate the dimension.

We say that **Property E** holds for  $\mathbf{x}$ , with parameters  $k$  and  $\mathbf{I}$ , if the corresponding  $N_k^{\mathbf{I}}$  is such that either  $N_k^{\mathbf{I}}(\infty) = 0$  or  $\lim_{n \rightarrow \infty} \frac{N_k^{\mathbf{I}}(n)}{n} > 0$ .

Define the density in  $\mathbb{N}_0$  of a set  $S \subseteq \mathbb{N}_0$  as  $\lim_{n \rightarrow \infty} \frac{|S \cap [0, n-1]|}{n}$  if the limit exists. Then Property E says that  $S_k^{\mathbf{I}}$  is either empty or of a positive density in  $\mathbb{N}_0$ .

Let  $A_0$  be the set of all of the numerical sequences such that Property E holds for all  $k \in \mathbb{N}$  and  $\mathbf{I} \in \mathcal{I}^k$ .

We further add a tightness condition, called **Property T**:

$$\lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{[0, K)}(|x_i|) = \lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{N_1^{(-K, K)}(n)}{n} = 1.$$

Intuitively, Property T prevents the “main part” of the sequence from escaping to infinity. Let  $A_1$  be the subset of  $A_0$  consisting of all of the sequences in  $A_0$  which have Property T.

Denote by  $F_n^1$ ,  $n \in \mathbb{N}$ , the marginal empirical measures of a sequence  $\mathbf{x} \in A_1$ , determined by

$$F_n^1(I) = \frac{N_1^I(n)}{n}, \quad I \in \mathcal{I}.$$

The fact that  $\mathbf{x} \in A_0$  implies that  $\lim_{n \rightarrow \infty} F_n^1(I)$  always exists, Property T then guarantees that the sequence of measures  $\{F_n^1\}_{n \in \mathbb{N}}$  is tight, and so  $\lim_{n \rightarrow \infty} F_n^1(I)$

generates a probability measure. More generally, for any  $k \in \mathbb{N}$ , the  $k$ -dimensional empirical measure  $F_n^k$  is defined by

$$F_n^k(\mathbf{I}) = \frac{N_k^{\mathbf{I}}(n)}{n}, \quad \mathbf{I} \in \mathcal{I}^k.$$

It is easy to see that Property T also assures the tightness of any finite-dimensional empirical measures, and thus  $\lim_{n \rightarrow \infty} F_n^k(\mathbf{I})$  generates a probability measure on  $\mathbb{R}^k$ .

Together, the family of limiting probability measures  $\{\lim_{n \rightarrow \infty} F_n^k\}_{k \in \mathbb{N}}$  satisfies the consistency condition, and thus by Kolmogorov's existence theorem, there exists a stationary process  $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{N}_0}$  such that any finite-dimensional distribution of  $\mathbf{Y}$ ,  $F_{Y_0, \dots, Y_{k-1}}$ , is given by

$$F_{Y_0, \dots, Y_{k-1}} = \lim_{n \rightarrow \infty} F_n^k.$$

The process  $\mathbf{Y} = \mathbf{Y}^{\mathbf{x}}$  is unique in distribution since all of its finite-dimensional distributions are completely determined by the empirical measures of the sequence  $\mathbf{x}$ . We call  $\mathbf{Y}^{\mathbf{x}}$  the stationary process induced by the numerical sequence  $\mathbf{x} \in A_1$ .

Define the set

$$A := \{\mathbf{x} \in A_1 : \mathbf{Y}^{\mathbf{x}} \text{ is ergodic}\}.$$

Also, notice that to make  $\mathbf{Y}^{\mathbf{x}}$  well-defined, we only need a weaker version of Property E, where  $\lim_{n \rightarrow \infty} \frac{N_k^{\mathbf{I}}(n)}{n}$  exists for any  $k \in \mathbb{N}$  and  $\mathbf{I} \in \mathcal{I}^k$ , but  $N_k^{\mathbf{I}}(\infty) > 0$  does not necessarily imply  $\lim_{n \rightarrow \infty} \frac{N_k^{\mathbf{I}}(n)}{n} > 0$ .

### 3. COVERAGE BY $A$ OF PATHS FROM STATIONARY PROCESSES

The following theorem shows that the set  $A$  is large enough, so that every discrete-time stationary process puts mass one on  $A$ .

**THEOREM 3.1.** *Let  $\mathbf{X} = \{X_n\}_{n=0,1,\dots}$  be a stationary process. Then we have  $P(\mathbf{X} \in A) = 1$ .*

**Proof.** Firstly, by ergodic decomposition, it suffices to prove the result for the case where  $\mathbf{X}$  is ergodic. Moreover, for the ergodic process  $\mathbf{X}$ , once we prove that  $P(\mathbf{X} \in A_0) = 1$ , it follows immediately that  $P(\mathbf{X} \in A) = 1$  as well, since Property T and the ergodicity of the path are guaranteed by the pointwise ergodic theorem. Thus, it suffices to prove that  $P(\mathbf{X} \in A_0) = 1$ .

The fact that Property E holds for any fixed  $k$  and any single  $\mathbf{I}$  almost surely is a trivial consequence of the pointwise ergodic theorem. As a result, Property E also holds for any countable set of  $(k, \mathbf{I})$  almost surely. What is in question is whether this implies that Property E is satisfied for all  $(k, \mathbf{I})$ . In the rest of the proof, for

ease of notation, we will focus on the case where  $k = 1$ , and prove that Property E holds for all  $I \in \mathcal{I}$  almost surely. The cases for  $k > 1$  follow in a similar way.

Let  $F_1$  be the marginal distribution of  $X_k$  for any  $k = 0, 1, \dots$ . Denote by  $D_1$  the set of atoms of  $F_1$ ,

$$D_1 = \{a \in \mathbb{R} : F_1(\{a\}) > 0\},$$

and put  $D = D_1 \cup \mathbb{Q} \cup \{-\infty, \infty\}$ ; then both  $D_1$  and  $D$  are at most countable sets. Hence the set

$$A_2 := \{\mathbf{x} \in \mathbb{R}^{\mathbb{N}_0} : \text{Property E holds for } k = 1 \text{ and any } I = (a, b), a, b \in D\}$$

satisfies  $P(\mathbf{X} \in A_2) = 1$ . Thus, from now on we can assume that the paths are in  $A_2$ .

For any open interval  $(a, b)$ , there exists an increasing sequence of open intervals  $\{(a_i, b_i)\}_{i=1,2,\dots}$  such that  $a_i, b_i \in D$  for  $i = 1, 2, \dots$ , and  $(a, b) = \bigcup_i (a_i, b_i) = \lim_{i \rightarrow \infty} (a_i, b_i)$ . Let the corresponding sets be  $S$  and  $S_i$ , and the corresponding counting functions be  $N(n)$  and  $N_i(n)$ . By construction,  $S = \lim_{i \rightarrow \infty} S_i$ , and  $N(n) = \lim_{i \rightarrow \infty} N_i(n)$  for  $n \in \mathbb{N}$ . Suppose  $N(\infty) > 0$  but  $\lim_{n \rightarrow \infty} \frac{N(n)}{n} = 0$  for some path in  $A_2$ ; then  $\lim_{n \rightarrow \infty} \frac{N_i(n)}{n} = 0$  for any  $i$ , and  $N_i(\infty) > 0$  for  $i$  large enough, which contradicts the construction of  $A_2$ . Therefore, the only possibility that a path  $\mathbf{x}$  is in  $A_2 \setminus A$  is that the corresponding ratio  $\frac{N(n)}{n}$  does not admit a limit as  $n \rightarrow \infty$ .

By the pointwise ergodic theorem, for any fixed open interval  $I$ , we have

$$\frac{\sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in I\}}}{n} \rightarrow E(\mathbf{1}_{\{X_0 \in I\}}) = P(X_0 \in I)$$

almost surely. Thus, if we define the set

$$B := \left\{ \mathbf{x} : \sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in I\}} / n \rightarrow P(X_0 \in I) \text{ for all } I = (a, b), a, b \in D \right\},$$

then  $P(B) = P(A_2 \cap B) = 1$ . As a result, we can almost surely assume that  $\mathbf{x} \in A_2 \cap B$ .

Suppose that for such an  $\mathbf{x}$  and for an open interval  $I = (a, b)$ ,  $a, b \in \overline{\mathbb{R}}$ , the corresponding ratio  $\frac{N(n)}{n}$  does not admit a limit as  $n \rightarrow \infty$ . Without loss of generality, assume that  $a \in D$  and  $b \notin D$ . The cases where  $a \notin D$ ,  $b \in D$  and  $a \notin D$ ,  $b \notin D$  are similar. The non-existence of the limit implies that

$$u := \limsup_{n \rightarrow \infty} \frac{N(n)}{n} \neq \liminf_{n \rightarrow \infty} \frac{N(n)}{n} =: d.$$

By definition, for any  $b' \in D \cap (a, b)$ ,

$$(3.1) \quad \lim_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in (a, b')\}}}{n} \leq \liminf_{n \rightarrow \infty} \frac{N(n)}{n} = d.$$

On the other hand, for  $b'' \in D \cap (b, \infty)$ ,

$$(3.2) \quad \lim_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in (a, b'')\}}}{n} \geq \limsup_{n \rightarrow \infty} \frac{N(n)}{n} = u.$$

The limit above exists because

$$\sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in (a, b'')\}} = \sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in (a, \infty)\}} - \sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in (b'', \infty)\}}.$$

Subtracting (3.1) from (3.2), we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \mathbf{1}_{\{x_i \in [b', b'']\}}}{n} \geq u - d > 0$$

for any  $b', b'' \in D$  and  $b' < b < b''$ . Recall that since we work with  $A_2 \cap B$ , this also implies that

$$P(X_0 \in [b', b'']) \geq u - d.$$

Because  $D$  is dense in  $\mathbb{R}$ , we can take  $b' \uparrow b$  and  $b'' \downarrow b$ , leading to the result

$$P(X_0 = b) \geq u - d > 0.$$

However, since  $b \notin D$ ,  $b$  is not an atom of  $F_1$ . Thus  $P(X_0 = b) = 0$ , which is a contradiction. Hence the assumption is almost surely false and the limit exists with probability one. ■

In reality, checking the ergodicity of  $\mathbf{Y}^x$  for a given  $x$  by definition firstly requires us to fully recover the distribution of  $\mathbf{Y}^x$  from  $x$ , then determine whether the process  $\mathbf{Y}^x$  is ergodic given its distribution. Unfortunately, none of these two steps is practical enough to implement. However, for a given sequence  $x$ , we can derive an equivalent characterization of the ergodicity, which is directly built upon the behavior of the sequence rather than the property of the measure it induces.

**DEFINITION 3.1.** An *asymptotically proportional contraction* of the index set  $\mathbb{N}_0$  is a subset  $G$  of  $\mathbb{N}_0$  consisting of disjoint intervals  $G_i$  of consecutive integers,

$$G = \bigcup_{i=1}^{\infty} G_i,$$

satisfying:

1.  $G_i, i \in \mathbb{N}$ , are increasingly ordered, that is,  $\min\{n : n \in G_{i+1}\} > \max\{n : n \in G_i\}, i \in \mathbb{N}$ ;
2.  $|G_i| \rightarrow \infty$  as  $i \rightarrow \infty$ , where  $|\cdot|$  is the number of elements (integers) in a set;
3.  $\frac{|[0, n-1] \cap G|}{n} \rightarrow c > 0$  as  $n \rightarrow \infty$ .

DEFINITION 3.2. An *asymptotically proportional contraction* of a numerical sequence  $\mathbf{x} = \{x_n\}_{n=0,1,\dots}$  is a subsequence  $\{x_{n_i}\}_{n_i \in G}$  of  $\{x_n\}_{n \in \mathbb{N}_0}$ , where  $G$  is an asymptotically proportional contraction of the index set  $\mathbb{N}_0$ .

Intuitively, an asymptotically proportional contraction of a numerical sequence consists of pieces of the original sequence with length of the pieces going to infinity and the fraction of coverage converging to a fixed positive level.

THEOREM 3.2. Let  $\mathbf{x}$  be a numerical sequence in  $A_1$ . Then  $\mathbf{x} \in A$  if and only if all of its asymptotically proportional contractions induce the same process as the original sequence. That is, for any asymptotically proportional contraction  $\mathbf{x}'$ ,  $k \in \mathbb{N}$  and  $\mathbf{I} \in \mathcal{I}^k$ ,

$$\lim_{n \rightarrow \infty} \frac{N_k^{\mathbf{I}}(n)}{n} = \lim_{n \rightarrow \infty} \frac{N_k^{\mathbf{I}}(n)}{n},$$

where  $N'$  is the counting function defined in the same way as previously but for the subsequence  $\mathbf{x}'$ .

To prove Theorem 3.2, let us first introduce the following lemma. A similar result was presented in [5]. However, the proof to be presented below is much simpler, due to the difference in the framework used in this paper and that used in [5], and the fact that we only need a one-directional result.

LEMMA 3.1. Let  $\mathbf{x}$  be a path in  $A$ , therefore  $\mathbf{Y}^{\mathbf{x}}$  be ergodic. Let  $k \in \mathbb{N}$ ,  $\mathbf{I} = I_0 \times \dots \times I_{k-1} \in \mathcal{I}^k$  and  $S_k^{\mathbf{I}} = S_k^{\mathbf{I}}(\mathbf{x})$  be defined as previously. Then for every  $\epsilon > 0$ , there is an  $N$  such that the set

$$R_{k,N}^{\mathbf{I}} := \left\{ n \in \mathbb{N} : \left| \frac{1}{N} \sum_{i=n}^{n+N-1} \mathbf{1}_{S_k^{\mathbf{I}}}(i) - p_k^{\mathbf{I}} \right| > \epsilon \right\}$$

has a density in  $\mathbb{N}_0$  smaller than  $\epsilon$ , where the constant  $p_k^{\mathbf{I}}$  equals  $P(Y_0^{\mathbf{x}} \in I_0, \dots, Y_{k-1}^{\mathbf{x}} \in I_{k-1})$ .

Proof. Note that the existence of the density in  $\mathbb{N}_0$  for the sets  $R_{k,N}^{\mathbf{I}}$ , that is,

$$\lim_{n \rightarrow \infty} \frac{\sum_{m=0}^{n-1} \mathbf{1}_{R_{k,N}^{\mathbf{I}}}(m)}{n},$$

is guaranteed by Property E. Moreover, by the ergodicity of the path, the density of a set  $R_{k,N}^{\mathbf{I}}$  in  $\mathbb{N}_0$  is exactly the probability of the corresponding event. In other

words,

$$\lim_{n \rightarrow \infty} \frac{\sum_{m=0}^{n-1} \mathbf{1}_{R_{k,N}^{\mathbf{I}}}(m)}{n} = P\left(\left|\frac{1}{N} \sum_{i=0}^{N-1} \prod_{j=0}^{k-1} \mathbf{1}_{I_j}(Y_{i+j}^{\mathbf{x}}) - p_k^{\mathbf{I}}\right| > \epsilon\right) = P\left(\left|\frac{1}{N} \sum_{i=0}^{N-1} \mathbf{1}_{\{\theta^i \circ \mathbf{Y} \in A_k^{\mathbf{I}}\}} - p_k^{\mathbf{I}}\right| > \epsilon\right),$$

where  $\theta$  is the shift operator, and  $A_k^{\mathbf{I}}$  is a subset of the path space  $H$  defined as

$$A_k^{\mathbf{I}} = \{\mathbf{x} \in H : x_i \in I_i, i = 0, \dots, k-1\}.$$

Assume that the result in Lemma 3.1 is not true. Then it follows that there exists  $\epsilon > 0$  such that, for any  $N \in \mathbb{N}$ , either  $\{n \in \mathbb{N} : \frac{1}{N} \sum_{i=n}^{n+N-1} \mathbf{1}_{S_k^{\mathbf{I}}}(i) - p_k^{\mathbf{I}} > \epsilon\}$  or  $\{n \in \mathbb{N} : \frac{1}{N} \sum_{i=n}^{n+N-1} \mathbf{1}_{S_k^{\mathbf{I}}}(i) - p_k^{\mathbf{I}} < -\epsilon\}$  has a density in  $\mathbb{N}_0$  which is greater than or equal to  $\frac{\epsilon}{2}$ . Without loss of generality, assume that the set

$$\left\{n \in \mathbb{N} : \frac{1}{N} \sum_{i=n}^{n+N-1} \mathbf{1}_{S_k^{\mathbf{I}}}(i) - p_k^{\mathbf{I}} > \epsilon\right\}$$

has a density in  $\mathbb{N}_0$  greater than or equal to  $\frac{\epsilon}{2}$  for infinitely many  $N \in \mathbb{N}$ , which will be denoted by  $\{N_i\}_{i \in \mathbb{N}}$ . By ergodicity of the path  $\mathbf{x}$ , this implies that

$$P\left(\frac{1}{N_i} \sum_{j=0}^{N_i-1} \mathbf{1}_{\{\theta^j \circ \mathbf{Y} \in A_k^{\mathbf{I}}\}} > p_k^{\mathbf{I}} + \epsilon\right) \geq \frac{\epsilon}{2}$$

for  $i \in \mathbb{N}$ . As a result, the event

$$\left\{\frac{1}{N} \sum_{j=0}^{N-1} \mathbf{1}_{\{\theta^j \circ \mathbf{Y} \in A_k^{\mathbf{I}}\}} > p_k^{\mathbf{I}} + \epsilon \text{ for infinitely many } N\right\}$$

has a probability greater than or equal to  $\frac{\epsilon}{2}$ . This implies that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{\theta^j \circ \mathbf{Y} \in A_k^{\mathbf{I}}\}} \geq p_k^{\mathbf{I}} + \epsilon$$

happens with a probability greater than or equal to  $\frac{\epsilon}{2}$ .

However, since  $\mathbf{Y}$  is ergodic,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{\theta^j \circ \mathbf{Y} \in A_k^{\mathbf{I}}\}} = p_k^{\mathbf{I}}$$

almost surely, which is a contradiction. Therefore, we conclude that the assumption is invalid and the result in Lemma 3.1 holds. ■

**Proof of Theorem 3.2.** Let  $\mathbf{x} \in A$ . For  $k \in \mathbb{N}$ ,  $\mathbf{I} = I_0 \times \dots \times I_{k-1} \in \mathcal{I}^k$ , define  $S_k^{\mathbf{I}}(\mathbf{x})$  as previously. Let  $\mathbf{x}' = \{x_{n_i}\}_{n_i \in G}$  be an asymptotically proportional contraction of  $\mathbf{x}$ , where  $G = \bigcup_i G_i$  is the corresponding asymptotically proportional contraction of  $\mathbb{N}_0$ . To prove the “only if” direction, our goal is to show that the set

$$\begin{aligned} S_k^{\mathbf{I}}(\mathbf{x}') &= \{n \geq 0 : x'_n \in I_0, \dots, x'_{n+k-1} \in I_{k-1}\} \\ &= \{i \geq 0 : x_{n_i} \in I_0, \dots, x_{n_i+k-1} \in I_{k-1}\} \end{aligned}$$

has the same density in  $\mathbb{N}_0$  as  $S_k^{\mathbf{I}}(\mathbf{x})$ . Let  $c = \lim_{n \rightarrow \infty} \frac{|[0, n-1] \cap G|}{n}$ . By Lemma 3.1, for any  $\epsilon > 0$ , there exists  $N$  such that the set

$$R_{k,N}^{\mathbf{I}} = \left\{ n \in \mathbb{N}_0 : \left| \frac{1}{N} \sum_{j=n}^{n+N-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{x})}(j) - p_k^{\mathbf{I}} \right| > \epsilon \right\}$$

has a density in  $\mathbb{N}_0$  smaller than  $\epsilon$ . Hence, the upper density of  $R_{k,N}^{\mathbf{I}}$  in the set  $G$ , defined as

$$\limsup_{n \rightarrow \infty} \frac{|R_{k,N}^{\mathbf{I}} \cap [0, n-1] \cap G|}{|[0, n-1] \cap G|},$$

is smaller than  $\frac{\epsilon}{c}$ . Similar to  $R_{k,N}^{\mathbf{I}}$ , one can define

$$R'_{k,N}^{\mathbf{I}} := \left\{ n_i \in G : \left| \frac{1}{N} \sum_{j=i}^{i+N-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{x}')} (j) - p_k^{\mathbf{I}} \right| > \epsilon \right\}.$$

Since the operation of contraction will join different segments of the original path together,  $R'_{k,N}^{\mathbf{I}}$  and  $R_{k,N}^{\mathbf{I}}$  will not completely agree in  $G$ . However, since  $\lim_{n \rightarrow \infty} |G_n| = \infty$ , the two sets will have the same upper density in  $G$ . Therefore, the upper density of  $R'_{k,N}^{\mathbf{I}}$  in  $G$  is also smaller than  $\frac{\epsilon}{c}$ . It is easy to see that

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{x}')} (i) \\ &\leq \limsup_{n \rightarrow \infty} \frac{|R'_{k,N}^{\mathbf{I}} \cap [0, n-1] \cap G|}{|[0, n-1] \cap G|} \cdot 1 + 1 \cdot (p_k^{\mathbf{I}} + \epsilon) \leq p_k^{\mathbf{I}} + \epsilon \left( 1 + \frac{1}{c} \right). \end{aligned}$$

Since  $\epsilon$  can be arbitrarily small, we must have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{x}')} (i) \leq p_k^{\mathbf{I}}.$$

Symmetrically,  $\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{x}')} (i) \geq p_k^{\mathbf{I}}$ . Thus

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{x}')} (i) = p_k^{\mathbf{I}},$$

which shows that  $S_k^{\mathbf{I}}(\mathbf{x}')$  always has the same density in  $\mathbb{N}_0$ , being also the density of  $S_k^{\mathbf{I}}(\mathbf{x})$  in  $\mathbb{N}_0$ .

Conversely, assume that  $\mathbf{x} \in A_1$  but  $\mathbf{x} \notin A$ . Thus  $\mathbf{x}$  induces a stationary process  $\mathbf{Y} = \mathbf{Y}^{\mathbf{x}}$ , but it is not ergodic. Therefore, there exists  $p \in (0, 1)$  and stationary processes  $\mathbf{Z}$  and  $\mathbf{W}$  with distinct distributions such that  $F_{\mathbf{Y}} = pF_{\mathbf{Z}} + (1-p)F_{\mathbf{W}}$ . In particular, there exists  $k \in \mathbb{N}$  and  $\mathbf{I} = I_0 \times \dots \times I_{k-1} \in \mathcal{I}^k$  such that  $z := P(Z_i \in I_i, i = 0, \dots, k-1) \neq P(W_i \in I_i, i = 0, \dots, k-1) =: w$ . Without loss of generality, assume that  $z > w$ . Note that since  $\mathbf{x}$  induces  $\mathbf{Y}$ , we have

$$\lim_{n \rightarrow \infty} \frac{|S_k^{\mathbf{I}}(\mathbf{x}) \cap [0, n-1]|}{n} = P(Y_i \in I_i, i = 0, \dots, k-1) = pz + (1-p)w.$$

For  $m \in \mathbb{N}$ , define

$$V_0 := \left\{ j \in \mathbb{N}_0 : \frac{|S_k^{\mathbf{I}}(\mathbf{x}) \cap [j, j+m-1]|}{m} \geq \frac{(1+p)z + (1-p)w}{2} \right\}.$$

Intuitively,  $V_0$  is the set of the starting points of the segments of length  $m$  in  $\mathbf{x}$  for which the local density of the points in  $S_k^{\mathbf{I}}(\mathbf{x})$  is greater than or equal to  $\frac{(1+p)z + (1-p)w}{2}$ , being a level between  $z$  and  $pz + (1-p)w$ . It is clear by the construction of  $A_0$  that  $V_0$  has a density in  $\mathbb{N}_0$ .

Consider the process  $\mathbf{Z}$ . Similar to  $\mathbf{x}$ , we now have a random set

$$S_k^{\mathbf{I}}(\mathbf{Z}) = \{n \geq 0 : Z_{n+i} \in I_i, i = 0, \dots, k-1\}.$$

Then

$$\begin{aligned} z &= P(Z_i \in I_i, i = 0, \dots, k-1) = E(\mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{Z})}(0)) = E\left(\frac{1}{m} \sum_{j=0}^{m-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{Z})}(j)\right) \\ &\leq P\left(\frac{1}{m} \sum_{j=0}^{m-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{Z})}(j) \geq \frac{(1+p)z + (1-p)w}{2}\right) \cdot 1 \\ &\quad + 1 \cdot \frac{(1+p)z + (1-p)w}{2}. \end{aligned}$$

Hence, we have

$$\begin{aligned} &P\left(\frac{|S_k^{\mathbf{I}}(\mathbf{Z}) \cap [0, m-1]|}{m} \geq \frac{(1+p)z + (1-p)w}{2}\right) \\ &= P\left(\frac{1}{m} \sum_{j=0}^{m-1} \mathbf{1}_{S_k^{\mathbf{I}}(\mathbf{Z})}(j) \geq \frac{(1+p)z + (1-p)w}{2}\right) \geq \frac{(1-p)(z-w)}{2}. \end{aligned}$$

Since  $\mathbf{Y}$  is a mixture of  $\mathbf{Z}$  and  $\mathbf{W}$ , we obtain

$$P\left(\frac{|S_k^{\mathbf{I}}(\mathbf{Y}) \cap [0, m-1]|}{m} \geq \frac{(1+p)z + (1-p)w}{2}\right) \geq \frac{p(1-p)(z-w)}{2}.$$

This implies that the density of  $V_0$  in  $\mathbb{N}_0$  is greater than or equal to  $\frac{p(1-p)(z-w)}{2}$  since  $\mathbf{Y}$  is generated by  $\mathbf{x}$ . Denote the elements of  $V_0$  in an increasing order by  $V_0 = \{v_0, v_1, \dots\}$ , and define a subset  $V_1$  of  $V_0$  as

$$V_1 = \{v_{im}, i \in \mathbb{N}\}.$$

That is, we only take each  $m$ -th element in  $V_0$  to form  $V_1$ . Then  $V_1$  has a density in  $\mathbb{N}_0$  which is greater than or equal to  $\frac{p(1-p)(z-w)}{2m}$ . Moreover, the construction of  $V_1$  guarantees that the intervals  $[j, j+m-1], j \in V_1$ , are disjoint. We further take a subset of  $V_1$ , denoted by  $V_2$ , which has a density in  $\mathbb{N}_0$  exactly equal to  $\frac{p(1-p)(z-w)}{2m}$ . Finally, define

$$H = \bigcup_{j \in V_2} [j, j+m-1];$$

then  $H$  consists of disjoint sections of integers, each with length (number of integers)  $m$ , and the set  $H$  has density  $\frac{p(1-p)(z-w)}{2}$  in  $\mathbb{N}_0$ .

Recall that  $V_0, V_1, V_2$  and  $H$  all depend on  $m$ , so we can also denote them respectively by  $V_0(m), V_1(m), V_2(m)$  and  $H(m)$ . Notice, however, that the density of  $H(m)$  in  $\mathbb{N}_0$  does not depend on  $m$ . Now, we construct an asymptotically proportional contraction  $G$  of the index set  $\mathbb{N}_0$  in the following inductive way:

1. Define the set  $G(1) = H(1)$ . Since  $G(1)$  has a density in  $\mathbb{N}_0$  given by  $d := \frac{p(1-p)(z-w)}{2}$ , for any  $\epsilon_1 > 0$ , there exists  $N(1) \in \mathbb{N}$  such that  $N(1) \in G(1)$  and

$$\left| \frac{|G(1) \cap [0, n]|}{n+1} - d \right| \leq \frac{\epsilon_1}{3}$$

for any  $n \geq N(1)$ . Moreover, since  $H(2)$  also has a density in  $\mathbb{N}_0$  given by  $d$ , we can take  $N(1)$  large enough so that

$$\left| \frac{|H(2) \cap [0, n]|}{n+1} - d \right| \leq \frac{\epsilon_1}{3}$$

for any  $n \geq N(1)$ .

2. Let  $\{\epsilon_i\}$  be a sequence of positive numbers decreasing to zero. Assume that we already have a set  $G(m)$  and a positive integer  $N(m)$ , where  $G(m)$  consists of separate sections of consecutive integers with lengths of the sections increasing

to  $m$ , and has a density in  $\mathbb{N}_0$  given by  $d$ ;  $N(m)$  is the endpoint of a section with length  $m$  in  $G(m)$ :  $N(m) - i \in G(m)$ ,  $i = 0, \dots, m - 1$ , and satisfies

$$\left| \frac{|G(m) \cap [0, n]|}{n+1} - d \right| \leq \frac{\epsilon_m}{3}$$

and

$$\left| \frac{|H(m+1) \cap [0, n]|}{n+1} - d \right| \leq \frac{\epsilon_m}{3}$$

for  $n \geq N(m)$ . Then define

$$G(m+1) = (G(m) \cap [0, N(m)]) \cup \bigcup_{\substack{i \in V_2(m+1), \\ i \geq N(m)+1}} [i, i+m].$$

That is,  $G(m+1)$  is obtained by joining the part of  $G(m)$  before  $N(m)$  and the part of  $H(m+1)$  after  $N(m)$ , but the area around the joint point is modified so that only the whole intervals in  $H(m+1)$  are kept. Note that such defined  $G(m+1)$  consists of intervals of integers with lengths increasing to  $m+1$ . Since both  $H(m+1)$  and  $H(m+2)$  have a density in the set  $\mathbb{N}_0$  given by  $d$ , there exists  $N(m+1) > N(m)$  such that  $N(m+1) - i \in G(m+1)$ ,  $i = 0, \dots, m$ ,

$$\left| \frac{|G(m+1) \cap [0, n]|}{n+1} - d \right| \leq \frac{\epsilon_{m+1}}{3}$$

and

$$\left| \frac{|H(m+2) \cap [0, n]|}{n+1} - d \right| \leq \frac{\epsilon_{m+1}}{3}$$

for  $n \geq N(m+1)$ .

3. Define  $G$  as

$$G = \lim_{m \rightarrow \infty} G(m) = \bigcup_{m=1}^{\infty} (G(m) \cap [N(m-1)+1, N(m)]),$$

where  $N(0) = -1$ .

The set  $G$  that we constructed consists of sections of consecutive integers with lengths going to infinity. Moreover, it is not difficult to see that we can make  $G$  to have a density  $d$  in  $\mathbb{N}_0$ . Indeed, for  $m \in \mathbb{N}$  and any  $n \in [N(m-1)+1, N(m)]$ ,

$$\begin{aligned} \left| \frac{|G \cap [0, n]|}{n+1} - d \right| &= \left| \frac{|G(m) \cap [0, n]|}{n+1} - d \right| \\ &\leq \left| \frac{|H(m) \cap [0, n]|}{n+1} - d \right| \\ &\quad + \left| \frac{|G(m-1) \cap [0, N(m-1)]|}{N(m-1)+1} - \frac{|H(m) \cap [0, N(m-1)]|}{N(m-1)+1} \right| + O(m/n) \end{aligned}$$

$$\begin{aligned}
&\leq \left| \frac{|H(m) \cap [0, n]|}{n+1} - d \right| + \left| \frac{|G(m-1) \cap [0, N(m-1)]|}{N(m-1)+1} - d \right| \\
&\quad + \left| \frac{|H(m) \cap [0, N(m-1)]|}{N(m-1)+1} - d \right| + O(m/n) \\
&\leq \frac{\epsilon_{m-1}}{3} + \frac{\epsilon_{m-1}}{3} + \frac{\epsilon_{m-1}}{3} + O(m/n) = \epsilon_{m-1} + O(m/n).
\end{aligned}$$

The error term  $O(m/n)$  comes from the possible difference between  $H(m)$  and  $G(m)$  over  $[N(m-1)+1, N(m)]$  due to the modification made around the joint point, and converges to zero if we choose  $\{N(m)\}_{m \in \mathbb{N}}$  such that  $\frac{m}{N(m-1)} \rightarrow 0$  as  $m \rightarrow \infty$ .

As a result,  $G$  is an asymptotically proportional contraction of the index set  $\mathbb{N}_0$ . Moreover, by construction, it is clear that the lower density of  $S_k^{\mathbf{I}}(\mathbf{x})$  in  $G$ , defined as

$$\liminf_{n \rightarrow \infty} \frac{|S_k^{\mathbf{I}}(x) \cap G \cap [0, n-1]|}{|G \cap [0, n-1]|},$$

is greater than or equal to  $\frac{(1+p)z+(1-p)w}{2}$ . As before, let  $\mathbf{x}'$  be the asymptotically proportional contraction of  $\mathbf{x}$  determined by  $G$ . Then  $S_k^{\mathbf{I}}(\mathbf{x}')$  will have the same limiting behavior as  $S_k^{\mathbf{I}}(\mathbf{x})$  restricted in  $G$ . Hence, either  $S_k^{\mathbf{I}}(\mathbf{x}')$  has a density in  $\mathbb{N}_0$  greater than or equal to  $\frac{(1+p)z+(1-p)w}{2}$  or it does not have a density in  $\mathbb{N}_0$ , while  $S_k^{\mathbf{I}}(\mathbf{x})$  has a density in  $\mathbb{N}_0$  given by  $pz + (1-p)w$ . Thus, we have found an asymptotically proportional contraction of  $\mathbf{x}$  which does not induce the same process as the original sequence  $\mathbf{x}$ . ■

#### 4. TESTING STATIONARITY FOR PATHS IN $A$

The previous section shows that the set of functions  $A$  is large enough, such that any stationary process must put mass one on  $A$ . In this section, our goal is to show that the set  $A$  is also small enough, in the sense that it only contains the “essentially stationary” paths. As the results in this paper should be understood in the context of time series and the set  $A$  is characterized mostly asymptotically, the arguments used in this section, which serve as the last piece needed for the general picture, take a time series and asymptotic perspective. To this end, we consider the stationarity tests applied to the paths in  $A$  and prove that the results cannot be distinguished from those for the typical paths from stationary processes.

Let  $T$  be a hypothesis test for sample size  $n$  and consider the null hypothesis  $H_0$ :  $\mathbf{X} = \{X_0, \dots, X_{n-1}\}$  is stationary, or more precisely,  $H_0$ :  $\mathbf{X}$  is from a stationary time series defined on  $\mathbb{R}^{\mathbb{N}_0}$  or  $\mathbb{R}^{\mathbb{Z}}$ . From a probabilistic theoretic point of view,  $T$  is a mapping from  $\mathbb{R}^n$  to  $\{0, 1\}$ , where 0 and 1 correspond to “acceptance” and “rejection” of the null hypothesis, respectively. Alternatively,  $T$  can be represented as  $\mathbf{1}_{C_T}(x_0, \dots, x_{n-1})$ , where  $C_T \in \mathcal{C}_{\mathbb{R}^n}$  is the critical region (or, equivalently, the

rejection region) of the test,  $\mathcal{C}_{\mathbb{R}^n}$  being the cylindrical  $\sigma$ -field in  $\mathbb{R}^n$ . Define

$$\alpha_T(P) = P(T(\mathbf{X}) = 1) = P(C_T)$$

for  $P \in \mathcal{P}_0$ , the collection of stationary probability measures restricted to  $\mathbb{R}^n$ ; then the size of the test  $T$  is

$$\alpha = \sup_{P \in \mathcal{P}_0} \alpha_T(P).$$

We further define  $g_n = g_{n,0}$  to be the projection:  $g_n(\mathbf{x}) = (x_0, \dots, x_{n-1})$ ,  $\mathbf{x} \in \mathbb{R}^{\mathbb{N}_0}$ , and  $g_{n,i} := g_n \circ \theta^i$ . Thus,  $g_{n,i}$  is the operation of taking the moving window of size  $n$  starting from  $x_i$ .

**THEOREM 4.1.** *Let  $\mathbf{x} \in A$ . Let  $T$  be a given test for stationarity of size  $\alpha$  and with a given sample size  $n$ . Assume that  $T$  satisfies one of the following two conditions:*

1. *the critical region  $C_T$  is closed; or*
2. *the boundary of the critical region,  $\text{bd}(C_T)$ , is a null set under any  $P \in \mathcal{P}_0$ .*

*Then, the upper density of the index set*

$$\{i \in \mathbb{N}_0 : g_{n,i}(\mathbf{x}) \in C_T\}$$

*in  $\mathbb{N}_0$  is smaller than or equal to  $\alpha$ .*

Theorem 4.1 shows that if we apply a “well-behaved” stationarity test, in the sense that it satisfies one of the two conditions listed in the theorem, to a moving window with length  $n$  of any path  $\mathbf{x}$  in the set  $A$ , then the limiting frequency that the null hypothesis of stationarity is rejected should not exceed the size of the test. Intuitively, this ensures that when we apply a stationarity test to a path in  $A$ , we get the best possible result that we come to expect. More precisely, notice that the size  $\alpha$  can be approached by the rejection rate of the null hypothesis even if it is true. Then, by the ergodic decomposition, for arbitrarily small  $\epsilon > 0$ , there exists an ergodic process for which the rejection rate is larger than  $\alpha - \epsilon$ . Interpreting ergodicity as the equivalence between the mean across time and the mean across space, for a typical path of this ergodic process, the null hypothesis should be rejected with a limiting frequency greater than  $\alpha - \epsilon$  when the window of length  $n$  moves from the origin to  $+\infty$ . Therefore, having a limiting frequency of rejection smaller than or equal to  $\alpha$  is the best that we should expect to get. Any further requirement will exclude typical paths from certain stationary processes. Here, a “typical path” from a stationary process refers to a path such that the paths with the same statistical property happen with positive probability in the stationary process. Mathematically, a path  $\mathbf{x}$  is a *typical* path from a stationary process  $\mathbf{X}$  if the ergodic process that  $\mathbf{x}$  induces is a component with positive probability in the ergodic decomposition of  $\mathbf{X}$ . The simplest way would be directly taking  $\mathbf{X}$  as the ergodic process induced by  $\mathbf{x}$ , then almost all of its paths will have the same limiting statistical behavior as  $\mathbf{x}$ .

The goal of Theorem 4.1 is to compare the behavior of a path in  $A$  and that of a path sampled from a certain stationary process. As our paths are of an infinite length while any test for stationarity only uses a time series with a finite length, a natural idea for a performance evaluation would be to look at the limiting behavior of the test applied to a moving window with a fixed length. From this perspective, the significance of Theorem 4.1 resides in the conclusion that if a path  $\mathbf{x}$  is known to belong to the set  $A$ , then it is “statistically indistinguishable” with a typical path from a stationary process, in the sense that its performance under any stationarity test satisfying the conditions of Theorem 4.1 will be at least as good as the path from the stationary process, measured by the limiting rejection rate of the null hypothesis of stationarity. In other words, we should not expect to find any statistical method to be able to discriminate between  $\mathbf{x}$  and a typical path from some stationary process.

**Proof of Theorem 4.1.** Let  $\mathbf{x} \in A$  and  $\mathbf{Y}^{\mathbf{x}}$  be the ergodic process that  $\mathbf{x}$  induces. Define

$$\mathcal{J}_n = \left\{ J \in \mathcal{C}_{\mathbb{R}^n} : \lim_{m \rightarrow \infty} \frac{\sum_{i=0}^{m-1} \mathbf{1}_J(g_{n,i}(\mathbf{x}))}{m} = P(g_n(\mathbf{Y}^{\mathbf{x}}) \in J) \right\},$$

where  $P$  is the stationary measure induced by  $\mathbf{x}$ .

By the definition of the set  $A$ ,  $\mathcal{J}_n$  includes all of the  $n$ -dimensional open cylinder sets (i.e., open orthotopes for which each face and any given axis of  $\mathbb{R}^n$  are either perpendicular or parallel). In other words,  $\mathcal{I}^n \subset \mathcal{J}_n$ . Moreover,  $\mathcal{J}_n$  clearly has the following properties:

1.  $\emptyset \in \mathcal{J}_n, \mathbb{R}^n \in \mathcal{J}_n$ ;
2.  $J_1, J_2 \in \mathcal{J}_n, J_1 \supseteq J_2$  implies  $J_1 \setminus J_2 \in \mathcal{J}_n$ ;
3.  $J_1, J_2 \in \mathcal{J}_n, J_1 \cap J_2 = \emptyset$  implies  $J_1 \cup J_2 \in \mathcal{J}_n$ .

This is to say that  $\mathcal{J}_n$  is closed under a true difference and a finite disjoint union. The following proposition is a simple consequence of the fact that the Euclidean space  $\mathbb{R}^n$  with its usual topology is completely separable.

**PROPOSITION 4.1.** *Let  $C$  be a  $\mathcal{C}_{\mathbb{R}^n}$ -measurable set and  $P$  be a probability measure on  $(\mathbb{R}^n, \mathcal{C}_{\mathbb{R}^n})$ . Then for any  $\epsilon > 0$  there exists  $J \in \mathcal{J}_n$  such that  $J \subseteq C$  and  $P(J) \geq P(\overset{\circ}{C}) - \epsilon$ .*

**Proof.** The proof of this proposition is fundamental. Here we only provide a sketch of the proof. Consider the collection of all open cylinder sets whose vertices have rational coordinates. This is a countable topological basis of  $\mathbb{R}^n$  with its usual topology. Thus, for any  $C$ , its interior  $\overset{\circ}{C}$ , as an open set, can be expressed as the (countable) union of some members of this topological basis, denoted by  $B_1, B_2, \dots$ . For any  $\epsilon > 0$ , there exists a finite number  $k(\epsilon)$  such that  $P(\bigcup_{i=1}^{k(\epsilon)} B_i) > P(\overset{\circ}{C}) - \epsilon$ . Repartitioning  $\bigcup_{i=1}^{k(\epsilon)} B_i$  into finite disjoint hypercubes completes the proof. ■

The proof of Theorem 4.1 now becomes simple. Let  $T$  be a given test of size  $\alpha$  and with a sample size  $n$ , and let  $P$  be the stationary measure induced by  $\mathbf{x}$ . Thus,  $P(C_T) \leq \alpha$ . If  $T$  satisfies one of the two conditions listed in the theorem, then  $P((C_T^c)^\circ) = P(C_T^c) \geq 1 - \alpha$ , where  $(C_T^c)^\circ$  is the interior of  $C_T^c$ . For  $\epsilon > 0$ , by Proposition 4.1, there exists  $J \in \mathcal{J}_n$ ,  $J \subseteq C_T^c$ , such that

$$P(g_n(\mathbf{Y}^{\mathbf{x}}) \in J) \geq P(C_T^c) - \epsilon \geq 1 - \alpha - \epsilon.$$

Since  $J \in \mathcal{J}_n$ , the set  $\{i \in \mathbb{N}_0 : g_{n,i}(\mathbf{x}) \in J\}$  has a density in  $\mathbb{N}_0$  which is greater than or equal to  $1 - \alpha - \epsilon$ . This implies that  $\{i \in \mathbb{N}_0 : g_{n,i}(\mathbf{x}) \in C_T^c\}$  has a lower density in  $\mathbb{N}_0$  which is greater than or equal to  $1 - \alpha - \epsilon$ . Since  $\epsilon$  can be taken arbitrarily small, the lower density of  $\{i \in \mathbb{N}_0 : g_{n,i}(\mathbf{x}) \in C_T^c\}$  in  $\mathbb{N}_0$  is at least  $1 - \alpha$ . In other words, the upper density of  $\{i \in \mathbb{N}_0 : g_{n,i}(\mathbf{x}) \in C_T\}$  in  $\mathbb{N}_0$  is smaller than or equal to  $\alpha$ . ■

In practice, many stationarity tests introduce additional assumptions on the stochastic processes (or time series) in their null hypotheses or alternative hypotheses in constructing the tests or in analyzing their powers. For example, the classical unit root tests such as those proposed by [3] and [7] focus on linear autoregressive models; the papers [4] and [6] rely on a Markov structure, and some results in [2] require a certain mixing condition. A close examination of the proof of Theorem 4.1 reveals that such additional assumptions should not affect the result of the theorem. That is, if we can check that the process  $\mathbf{Y}^{\mathbf{x}}$  satisfies the additional assumptions of a test, then applying the test to a moving window of the path  $\mathbf{x} \in A$  will still lead to a limiting frequency of rejection no larger than the size of the test. Intuitively, the fact that the path  $\mathbf{x}$  is in  $A$  still guarantees the stationarity; if the test results in a higher frequency of rejection, this is due to the violation of the additional assumptions rather than an evidence of non-stationarity.

On the other hand, the two conditions in Theorem 4.1 are very general. As a matter of fact, a good test should have  $\text{bd}(C_T)$  to be a null set under the null hypothesis after all, and this is almost always the case in practice. Thus, the result of Theorem 4.1 applies to every existing time series test for stationarity. In some sense, what we have shown is that all of these tests reduce to checking whether or not the given path is in the set  $A$ .

In the above results, the stationarity of a path is evaluated using a moving window with a fixed length, i.e., a fixed sample size. Next we discuss two types of asymptotic behaviors of paths in  $A$ .

The first kind of asymptotic behavior does not require any additional assumption or technical result. Many stationarity tests used in practice do not have a known exact size, but only an asymptotic size. In other words, there are sequences of tests with sample sizes  $n$  increasing to infinity, such that although the size for any test with a fixed sample size is unknown, there exists a limiting size as  $n \rightarrow \infty$ . In this case, Theorem 4.1 immediately allows us to claim the following result.

**COROLLARY 4.1.** *Let  $\mathbf{x} \in A$ . Assume that  $\{T_n\}_{n \in \mathbb{N}}$  is a sequence of tests for stationarity, where  $T_n$  is for sample size  $n$  and has size  $\alpha_n$ . If  $\lim_{n \rightarrow \infty} \alpha_n = \alpha$ , and for each  $n \in \mathbb{N}$ , one of the two conditions in Theorem 4.1 is satisfied by the critical region  $C_{T_n}$  of  $T_n$ , then for any  $\epsilon > 0$  there exists  $N_\epsilon \in \mathbb{N}$  such that the upper density in  $\mathbb{N}_0$  of the index set*

$$\{i \in \mathbb{N}_0 : g_{n,i}(\mathbf{x}) \in C_{T_n}\}$$

*is smaller than  $\alpha + \epsilon$  for any  $n \geq N_\epsilon$ .*

The second kind of asymptotic result is more challenging. For a fixed path  $\mathbf{x}$ , we apply stationarity tests to a longer and longer fraction of the path, always starting from the first term  $x_0$ , and look at the limiting behavior of the results of these tests. Such limiting results are typically strong and require more assumptions about the tests, as well as some more powerful technical advances. To obtain the results, it is helpful to consider the cylindrical  $\sigma$ -field  $\mathcal{C}$  over the whole path space  $\mathbb{R}^{\mathbb{N}_0}$ , and define

$$\mathcal{J} = \left\{ J \in \mathcal{C} : \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbf{1}_J(\theta^i(\mathbf{x})) = P(\mathbf{Y}^{\mathbf{x}} \in J) \right\},$$

where  $\theta$  is the shift operator, so that  $\mathcal{C}$  and  $\mathcal{J}$  do not correspond to any fixed  $n$ . We can improve Proposition 4.1 with the following result.

**PROPOSITION 4.2.** *Let  $C$  be a  $\mathcal{C}$ -measurable set and let  $P$  be a probability measure on  $(\mathbb{R}^{\mathbb{N}_0}, \mathcal{C})$ . Then for any  $\epsilon > 0$  there exists  $J \in \mathcal{J}$  such that  $J \subseteq C$  and  $P(J) \geq P(C) - \epsilon$ .*

**Proof.** For each  $n \in \mathbb{N}$ , let  $\mathcal{J}_n$  be defined as in the proof of Theorem 4.1. Denote by  $\mathcal{G}_n$  the collection of the sets  $C \in \mathcal{C}_{\mathbb{R}^n}$  such that for any  $\epsilon > 0$  there exists  $J \in \mathcal{J}_n$ ,  $J \subseteq C$ , for which  $P(J) \geq P(C) - \epsilon$ . Clearly,  $\mathcal{J}_n \subseteq \mathcal{G}_n$ . In particular, all of the  $n$ -dimensional open cylinder sets are in  $\mathcal{G}_n$ . Indeed, it is not difficult to verify that all of the  $n$ -dimensional cylinder sets, regardless of the openness/closedness of the boundaries, are all in  $\mathcal{J}_n \subseteq \mathcal{G}_n$ . Moreover,  $\mathcal{G}_n$  is closed under finite disjoint unions. To see this, let  $C_1, \dots, C_m$  be disjoint sets in  $\mathcal{G}_n$ . Let  $J_1, \dots, J_m$  be the sets satisfying Proposition 4.1 for  $C_1, \dots, C_m$  and  $\epsilon_i = 2^{-i}\epsilon$ ,  $i = 1, \dots, m$ . Then  $J = \bigcup_{i=1}^m J_i$  is in  $\mathcal{J}_n$ ,  $J \subseteq C$  and satisfies  $P(J) \geq P(C) - \epsilon$ . Denote by  $\mathcal{F}_n$  the field generated by the  $n$ -dimensional cylinder sets. Then a result in [1] shows that each member in  $\mathcal{F}_n$  can be expressed as a finite union of disjoint cylinder sets. As a result,  $\mathcal{F}_n \subseteq \mathcal{G}_n$ .

Next we prove that  $\mathcal{F}_n \subseteq \mathcal{J}_n$ . Note that  $\mathbb{R}^n \in \mathcal{F}_n \cap \mathcal{J}_n$ , and  $C \in \mathcal{F}_n \cap \mathcal{J}_n$  implies  $C^c \in \mathcal{F}_n \cap \mathcal{J}_n$ . Furthermore,  $\mathcal{F}_n \cap \mathcal{J}_n$  is closed under union. To see this, let  $C_1, C_2 \in \mathcal{F}_n \cap \mathcal{J}_n$ ; then  $C_1 \cup C_2$  and  $(C_1 \cup C_2)^c$  are both in  $\mathcal{F}_n \subseteq \mathcal{G}_n$ . Consequently, for each  $\epsilon > 0$ , there exist  $J_{1,\epsilon}, J_{2,\epsilon} \in \mathcal{J}_n$ ,  $J_{1,\epsilon} \subseteq C_1 \cup C_2$ ,  $J_{2,\epsilon} \subseteq$

$(C_1 \cup C_2)^c$ , such that  $P(J_{1,\epsilon}) \geq P(C_1 \cup C_2) - \epsilon$ ,  $P(J_{2,\epsilon}) \geq P((C_1 \cup C_2)^c) - \epsilon$ . Therefore, we have

$$\begin{aligned} P(C_1 \cup C_2) - \epsilon &\leq P(J_{1,\epsilon}) = \lim_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbf{1}_{J_{1,\epsilon}}(g_{n,i}(\mathbf{x}))/m \\ &\leq \liminf_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbf{1}_{C_1 \cup C_2}(g_{n,i}(\mathbf{x}))/m. \end{aligned}$$

Sending  $\epsilon$  to zero leads to the following result:

$$\liminf_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbf{1}_{C_1 \cup C_2}(g_{n,i}(\mathbf{x}))/m \geq P(C_1 \cup C_2).$$

Symmetrically, using  $J_{2,\epsilon}$ , we have

$$\liminf_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbf{1}_{(C_1 \cup C_2)^c}(g_{n,i}(\mathbf{x}))/m \geq P(C_1 \cup C_2)^c.$$

Thus,  $\lim_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbf{1}_{C_1 \cup C_2}(g_{n,i}(\mathbf{x}))/m$  exists and is equal to  $P(C_1 \cup C_2)$ . Hence  $C_1 \cup C_2 \in \mathcal{F}_n \cap \mathcal{J}_n$ .  $\mathcal{F}_n \cap \mathcal{J}_n$  is a field. Since  $\mathcal{F}_n$  is the field generated by the  $n$ -dimensional cylinder sets, and all of the cylinder sets are both in  $\mathcal{F}_n$  and  $\mathcal{J}_n$ , we must have  $\mathcal{F}_n \subseteq \mathcal{J}_n$ .

Finally, let  $\mathcal{F} = \bigcup_{n \in \mathbb{N}} \mathcal{F}_n$  be the field on  $\mathbb{R}^{\mathbb{N}_0}$  generated by all the open cylinder sets. Notice that since any members in  $\mathcal{F}$  only have a finite number of finite-dimensional constraints,  $\mathcal{F} \subseteq \bigcup_{n \in \mathbb{N}} \mathcal{J}_n \subseteq \mathcal{J}$ . Denote by  $\mathcal{C}'$  the collection of sets  $C$  in  $\mathcal{C}$  such that for each  $\epsilon > 0$  there exists  $J \in \mathcal{F}$ ,  $J \subseteq C$ , for which  $P(J) \geq P(C) - \epsilon$ . By definition, it is easy to see that  $\mathcal{C}'$  contains  $\emptyset$  and  $\mathbb{R}^{\mathbb{N}_0}$ . Moreover, let  $C_1, C_2, \dots \in \mathcal{C}'$ ; then for any  $\epsilon > 0$  there exists  $N \in \mathbb{N}$  and  $J_1, \dots, J_N \in \mathcal{F}$  such that  $P(\bigcup_{i=N+1}^{\infty} C_i \setminus \bigcup_{i=1}^N C_i) \leq \frac{\epsilon}{2}$  and  $P(J_i) \geq P(C_i) - 2^{-i}\epsilon$  for  $i = 1, \dots, N$ . The set  $J = \bigcup_{i=1}^N J_i$  is in  $\mathcal{F}$  and satisfies  $P(J) \geq P(\bigcup_{i \in \mathbb{N}} C_i) - \epsilon$ . Hence  $\bigcup_{i \in \mathbb{N}} C_i \in \mathcal{C}'$ . Similarly, it is easy to see that  $\mathcal{C}'$  is closed under finite intersections. As a result,  $\mathcal{C}'$  is a topology. Since clearly  $\mathcal{F} \subseteq \mathcal{C}'$ ,  $\mathcal{C}'$  contains the topology generated by  $\mathcal{F}$ , which is the natural topology on  $\mathbb{R}^{\mathbb{N}_0}$ . Thus, we can conclude that for any  $\mathcal{C}$ -measurable set  $C$ , for the open set  $\overset{\circ}{C}$ , there exists a set  $J \in \mathcal{F} \subseteq \mathcal{J}$  such that  $P(J) \geq P(\overset{\circ}{C}) - \epsilon$ . ■

Proposition 4.1 and its consequence, Theorem 4.1, show that for any time series stationarity test with a fixed sample size satisfying some mild conditions, a path in the set  $A$  will behave as well as a typical path from a stationary process. Proposition 4.2 allows us to generalize this statement to any asymptotic property. For instance, let  $\{T_n\}_{n \in \mathbb{N}}$  be a sequence of stationarity tests with sample sizes  $n$  and satisfying the conditions in Theorem 4.1. At the risk of abusing the notation,

we also use  $T_n$  for the corresponding test statistics. Then, for  $\mathbf{x} \in A$ , the limiting behavior of  $T_n(\mathbf{x})$  as  $n \rightarrow \infty$  will be comparable to that of  $\mathbf{Y}^{\mathbf{x}}$ , which is a stationary process.

EXAMPLE 4.1. If for any stationary time series  $\mathbf{X}$ , the limiting rejection rate of  $T_n$ ,

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n T_i(g_i(\mathbf{X}))/n,$$

almost surely exists and is bounded from above by a constant  $\alpha$ , then Proposition 4.2 implies that for any  $\mathbf{x} \in A$  and generic  $m \in \mathbb{N}$ ,

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n T_i(g_{i,m}(\mathbf{x}))/n$$

exists and is bounded from above by  $\alpha$ . “Generic” means that the set of  $m$  for which the result does not hold has a limiting density zero in  $\mathbb{N}$ . If the assumption is relaxed to the existence of the upper/lower limit of the rejection rate and their bounds, the corresponding results hold as well for the paths in  $A$ .

**Acknowledgments.** The authors would like to thank Brett Coburn for being a research assistant for this project.

## REFERENCES

- [1] P. Billingsley, *Probability and Measure*, Wiley, 1995.
- [2] F. Busetti and A. C. Harvey, *Tests of strict stationarity based on quantile indicators*, J. Time Series Analysis 31 (2010), pp. 435–450.
- [3] D. A. Dickey and W. A. Fuller, *Distribution of the estimators for autoregressive time series with a unit root*, J. Amer. Statist. Assoc. 74 (1979), pp. 427–431.
- [4] I. Domowitz and M. A. El-Gamal, *A consistent nonparametric test of ergodicity for time series with applications*, J. Econometrics 102 (2001), pp. 365–398.
- [5] H. Furstenberg, *Stationary Processes and Prediction Theory*, Princeton University Press, Princeton 1960.
- [6] S. Kanaya, *A nonparametric test for stationarity in continuous-time Markov processes*, Job Market Paper, University of Oxford, 2011.
- [7] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, *Testing the null hypothesis of stationarity against the alternative of a unit root*, J. Econometrics 54 (1992), pp. 159–178.

Yi Shen  
University of Waterloo  
200 University Avenue West  
Waterloo, Ontario N2L 3G1, Canada  
E-mail: yi.shen@uwaterloo.ca

Tony S. Wirjanto  
University of Waterloo  
200 University Avenue West  
Waterloo, Ontario N2L 3G1, Canada  
E-mail: twirjanto@uwaterloo.ca

Received on 27.2.2018;  
revised version on 28.6.2018