Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2011, Vol. 47, No. 4, 1219–1225
DOI: 10.1214/11-AIHP426

ANNALES DE L'INSTITUT HENRI POINCARÉ PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Testing stationary processes for independence

Gusztáv Morvai[a,1] and Benjamin Weiss[b]

[a] *MTA-BME Stochastics Research Group, Egry József utca 1, Building H, Budapest, 1111 Hungary. E-mail: morvai@math.bme.hu*
[b] *Institute of Mathematics, Hebrew University of Jerusalem, Jerusalem 91904, Israel. E-mail: weiss@math.huji.ac.il*

Received 27 October 2009; revised 21 December 2010; accepted 15 March 2011

Abstract. Let H_0 denote the class of all real valued i.i.d. processes and H_1 all other ergodic real valued stationary processes. In spite of the fact that these classes are not countably tight we give a strongly consistent sequential test for distinguishing between them.

Résumé. Soit H_0 la classe de tous les processus indépendants et équidistribués à valeurs réelles, et H_1 la classe complémentaire dans l'ensemble des processus ergodiques. Nous donnons un test séquentiel fortement consistant pour les distinguer.

MSC: 62M07

Keywords: Independent processes; Hypothesis testing

1. Introduction

In sequential testing for distinguishing between two competing hypotheses, the usual framework consists of two classes of stochastic processes H_0 and H_1, and a sequence of n observations $X_1, X_2, \ldots, X_n$ which represent a sampling from a single process. A test is usually a sequence of functions g_n of n arguments that take the values zero and one and it is said to be *weakly consistent* if the sequence converges to 0/1 in probability according to whether the process belongs to H_0/H_1. It is said to be *consistent* if the convergence is pointwise with probability one. Much classical work (see [2,3,5]) was done in the case where the classes H_i consisted of i.i.d. processes with the classes being distinguished by the type of the underlying distribution. More recently (see [11], [6], [12] and [10]) the more general problem was considered where the classes H_i are no longer restricted to being i.i.d. but are assumed to contain certain ergodic stationary processes. In particular in the latter two papers some general sufficient conditions were given on the classes which ensure the existence of consistent tests.

Perhaps the simplest question of this type that one can put is to distinguish between H_0, the class of all i.i.d. processes and its complement in the class of ergodic stationary processses, namely H_1 is taken to be all such processes that are not independent. If we restrict attention to finite valued processes with a fixed number of states then D. Bailey in his thesis [1] gave just such a consistent test that was based on his universal scheme for estimating the entropy of an unknown process. If however we take H_0 to be the class of all real valued independent processes then none of these earlier results apply. For example, the main result in [10] assumes that the union of the two classes is contained in a countable union of uniformly tight processes whereas the class of all distribution functions on the real line is not countably tight. Our purpose in this note is to answer just this question by providing a consistent sequential test for deciding whether an arbitrary stationary ergodic process is independent or not. In fact our result applies to vector-

[1] Supported by OTKA grant No. K75143 and Bolyai János Research Scholarship.

valued processes as well, indeed all that one needs to assume is that the values taken by the process lie in a countably generated measurable space which is known a priori.

In the first section we shall show how an appropriate quantization enables one to adapt any good set of independence tests for finite valued i.i.d. processes to produce an independence test for general processes. In the second section we shall give a specific example of such a good test. Similar questions can be asked about classes of Markov processes and we have answered some of these in earlier work. In [8] we showed that even when we restrict to binary processes the class of all finite order Markov chains cannot be distinguished from its complement by any weakly consistent test. On the other hand for any k the order-k Markov chains can be distinguished from their complement by a consistent test even in the setting of countable alphabets (see [7] and [9]).

2. A general test for independence

Let $\{\mathcal{X}, \mathcal{F}\}$ be a countably generated measurable space, such as $\mathbb{R}$ or $\mathbb{R}^d$ or any Polish space with its Borel σ-algebra. Our process $\{X_n\}_{n=-\infty}^{\infty}$ will be stationary and ergodic taking values in $\mathcal{X}$. (Note that all stationary time series $\{X_n\}_{n=0}^{\infty}$ can be thought to be a two sided time series, that is, $\{X_n\}_{n=-\infty}^{\infty}$.) The reader can keep in mind real valued processes without loss of generality.

For notational convenience, let $X_m^n = (X_m, \dots, X_n)$, where $m \le n$. Note that if $m > n$ then X_m^n is the empty string.

Let $\mathcal{A}_k$ be a refining sequence of finite partitions of $\mathcal{X}$ and denote by $\hat{\mathcal{A}}_k$ the corresponding algebras of sets. Assume that $\bigcup_{k=1}^{\infty} \hat{\mathcal{A}}_k$ generates the whole σ-algebra $\mathcal{F}$. For a countably generated measurable space such sequences always exist. For example $\mathcal{A}_k$ can be the partition of $\mathbb{R}$ into intervals of the form $[j/2^k, (j+1)/2^k)$ for all integers $|j| < k2^k$ and their complement. We will denote by

$$Q_k(x) = A_i \quad \text{if } x \in A_i \in \mathcal{A}_k$$

the function which assigns to x the set of the partition to which it belongs. For any stationary and ergodic $\mathcal{X}$-valued process $\{X_n\}$, let

$$Y_n^{(k)} = Q_k(X_n)$$

denote the process obtained by quantizing at level k.

Lemma 1. *A stationary and ergodic $\mathcal{X}$-valued process $\{X_n\}$ is independent if and only if the corresponding $\mathcal{A}_k$-valued process $Y_n^{(k)}$ is independent for each k.*

Proof. If the $\{X_n\}$ is independent then as functions of the X_n's so are all the $Y_n^{(k)}$. In the opposite direction assume that for all k the $Y_n^{(k)}$'s are independent. Denote by $\mathcal{A}$ the union of the finite algebras $\bigcup_{k=1}^{\infty} \hat{\mathcal{A}}_k$ and by $\mathcal{A}^{\mathbb{Z}}$ the product σ-algebra generated by $\mathcal{A}$. On this σ-algebra the probability measure defined by the $\{X_n\}$ process is a product measure. Since the algebra $\bigcup_{k=1}^{\infty} \hat{\mathcal{A}}_k$ generates the full σ-algebra $\mathcal{F}$ and since measures that agree on an algebra agree also on the σ-algebra it generates this implies that the probability measure defined by the $\{X_n\}$ process is a product measure on $\mathcal{F}^{\mathbb{Z}}$ as was to be shown. □

Let $\mathcal{Y}$ be a finite set. A sequence of functions g_n defined on $\mathcal{Y}^{n+1}$ is a consistent test for independence if for any $\mathcal{Y}$-valued stationary and ergodic process $\{Y_n\}$, eventually almost surely

$$g_n(Y_0, \dots, Y_n) = \begin{cases} \mathit{IND} & \text{if the process is independent,} \\ \mathit{DEP} & \text{otherwise.} \end{cases}$$

Define

$$E_n = \sup P\bigl(g_n(Y_0, \dots, Y_n) = \mathit{DEP}\bigr),$$

where the sup is over all $\mathcal{Y}$-valued independent and identically distributed processes.

If in addition

$$\sum_{n=0}^{\infty} E_n < \infty$$

then we will say that g_n is a *strongly consistent* test. In the next section we will show how to construct such tests.

Now let $\mathcal{Y}_k = \mathcal{A}_k$ as above. We will show how to use any family of strongly consistent tests $\{g_n^{(k)}\}$ to devise a consistent test for the class of all $\mathcal{X}$-valued processes. For an increasing sequence $\mathbf{b} = \{b_n\}$ such that $b_n \to \infty$ define

$$TEST_n^{\mathbf{b}}(X_0^n) = \begin{cases} IND & \text{if } g_n^{(k)}(Y_0^{(k)}, \dots, Y_n^{(k)}) = IND \text{ for all } 1 \le k \le b_n, \\ DEP & \text{otherwise.} \end{cases}$$

Theorem 1. *Let $\{g_n^{(k)}\}$ be a family of strongly consistent tests and define m_k be so large that $\sum_{k=1}^{\infty} \sum_{n=m_k}^{\infty} E_n^{(k)} < \infty$ (e.g. $\sum_{n=m_k}^{\infty} E_n^{(k)} < 2^{-k}$). For $n \ge m_1$ let $b_n = \max\{1 \le k\colon m_l \le n \text{ for all } 1 \le l \le k\}$. Then $TEST_n^{\mathbf{b}}(X_0^n)$ is consistent for all stationary and ergodic $\mathcal{X}$-valued processes $\{X_n\}$.*

Proof. If the $\mathcal{X}$-valued process $\{X_n\}$ is dependent then by Lemma 1 for some k, $g_n^{(k)}(Y_0^{(k)}, \dots, Y_n^{(k)}) = DEP$ eventually almost surely and so $TEST_n^{\mathbf{b}}(X_0^n) = DEP$ eventually almost surely. Now assume that $\{X_n\}$ is an independent $\mathcal{X}$-valued process. Then by Lemma 1, for each k, $\{Y_n^{(k)}\}$ is independent and so

$$\sum_{n=m_1}^{\infty} \sum_{k=1}^{b_n} E_n^{(k)} = \sum_{k=1}^{\infty} \sum_{n=m_k}^{\infty} E_n^{(k)} < \infty.$$

Thus

$$\begin{aligned} &\sum_{n=m_1}^{\infty} P\big(TEST_n^{\mathbf{b}}(X_0^n) = DEP\big) \\ &\quad = \sum_{n=m_1}^{\infty} P\big(g_n^{(k)}\big(Y_0^{(k)}, \dots, Y_n^{(k)}\big) = DEP \text{ for some } 1 \le k \le b_n\big) \\ &\quad \le \sum_{n=m_1}^{\infty} \sum_{k=1}^{b_n} P\big(g_n^{(k)}\big(Y_0^{(k)}, \dots, Y_n^{(k)}\big) = DEP\big) \\ &\quad \le \sum_{n=m_1}^{\infty} \sum_{k=1}^{b_n} E_n^{(k)} < \infty. \end{aligned}$$

By the Borel–Cantelli Lemma, $g_n^{(k)}(Y_0^{(k)}, \dots, Y_n^{(k)}) = IND$ for all $1 \le k \le b_n$ eventually almost surely. The proof of Theorem 1 is complete. □

3. Construction of strongly consistent tests

In this section we will show how to construct a strongly consistent test for independence for the class of all processes taking values in a fixed finite set $\mathcal{Y}$. Assume $\{Y_n\}$ is a stationary and ergodic process taking values in $\mathcal{Y}$. First let us define a number which will measure the degree of independence of the process. It will be zero if and only if the process is independent.

For convenience let $p(y_{-k+1}^0)$ and $p(y|y_{-k+1}^0)$ denote the distribution $P(Y_{-k+1}^0 = y_{-k+1}^0)$ and the conditional distribution $P(Y_1 = y|Y_{-k+1}^0 = y_{-k+1}^0)$, respectively.

Define

$$\Gamma = \sup_{1 \le k < \infty} \sup_{\{z_{-k+1}^0 \in \mathcal{Y}^k, y \in \mathcal{Y}:\ p(z_{-k+1}^0, y) > 0\}} \left| p(y) - p\left(y | z_{-k+1}^0\right) \right|.$$

We proceed to define an empirical version of this based on the observation of a finite data segment Y_0^n. To this end first define the empirical version of $p(y)$ and $p(y|w_{-k+1}^0)$ as

$$\hat{p}_n(y) = \frac{\#\{0 \le t \le n-1:\ Y_{t+1} = y\}}{n}$$

and

$$\hat{p}_n\left(y|w_{-k+1}^0\right) = \frac{\#\{k-1 \le t \le n-1:\ Y_{t-k+1}^{t+1} = (w_{-k+1}^0, y)\}}{\#\{k-1 \le t \le n-1:\ Y_{t-k+1}^{t} = w_{-k+1}^0\}}.$$

These empirical distributions, as well as the sets we are about to introduce are functions of Y_0^n, but we suppress the dependence to keep the notation manageable.

For a fixed $0 < \gamma < 1$ let $\mathcal{L}_k^n$ denote the set of strings with length k which appear more than $n^{1-\gamma}$ times in Y_0^n. That is,

$$\mathcal{L}_k^n = \left\{ y_{-k+1}^0 \in \mathcal{Y}^k:\ \#\left\{k-1 \le t \le n-1:\ Y_{t-k+1}^t = y_{-k+1}^0\right\} > n^{1-\gamma} \right\}.$$

Finally, define the empirical version of Γ as follows:

$$\hat{\Gamma}_n = \max_{y \in \mathcal{Y}} \max_{1 \le k < \infty} \max_{z_{-k+1}^0 \in \mathcal{L}_k^n} \left| \hat{p}_n(y) - \hat{p}_n\left(y|z_{-k+1}^0\right) \right|.$$

Let us agree by convention that if the set over which we are maximizing is empty then the maximum is zero.

Observe, that by ergodicity, the ergodic theorem implies that almost surely the empirical distributions $\hat{p}$ converge to the true distributions p and so

$$\liminf_{n\to\infty} \hat{\Gamma}_n \ge \Gamma \quad \text{almost surely.}$$

If the process is dependent then clearly $\Gamma > 0$. If the process is independent then $\Gamma = 0$ and we will show that not just $\hat{\Gamma}_n \to 0$ almost surely but it converges to zero at a certain rate.

Let $0 < \beta < \frac{1-\gamma}{2}$ be arbitrary. Let

$$g_n = \begin{cases} \mathit{IND} & \text{if } \hat{\Gamma}_n \le n^{-\beta}, \\ \mathit{DEP} & \text{otherwise.} \end{cases}$$

Note that g_n depends on Y_0^n.

Theorem 2. *Let $\{Y_n\}$ be an arbitrary stationary and ergodic process taking values from a finite set $\mathcal{Y}$. Assume that $0 < \gamma < 1$ and $0 < \beta < \frac{1-\gamma}{2}$. If the process is independent then eventually almost surely, $g_n = \mathit{IND}$ and if the process is not independent then eventually almost surely, $g_n = \mathit{DEP}$. Furthermore, if the process is independent and identically distributed then for $n > 2^{1/(1-\gamma-2\beta)}$*

$$P(g_n = \mathit{DEP}) \le 14|\mathcal{Y}|n^4 \mathrm{e}^{-n^{-2\beta+1-\gamma}/2}$$

and the right-hand side is summable.

Proof. If the process is not independent then there is a string $z_{-k+1}^0 \in \mathcal{Y}^k$ and a letter $y \in \mathcal{Y}$ such that $p(y) \ne p(y|z_{-k+1}^0)$ and $p(z_{-k+1}^0 y) > 0$. By ergodicity, $\hat{p}_n(y) \to p(y)$ and $\hat{p}_n(y|z_{-k+1}^0) \to p(y|z_{-k+1}^0)$ and so $\liminf_{n\to\infty} \hat{\Gamma}_n > 0$ almost surely which in turn implies that $g_n = \mathit{DEP}$ eventually almost surely.

Assume that the process is independent and identically distributed. Then

$$
\begin{aligned}
&P(\hat{\Gamma}_n > n^{-\beta}) \\
&\quad = P\Big(\max_{y\in\mathcal{Y}} \max_{1\le k<\infty} \max_{(z^0_{-k+1},y)\in\mathcal{L}^n_k} \big|\hat{p}_n(y) - \hat{p}_n(y|z^0_{-k+1})\big| > n^{-\beta}\Big) \\
&\quad \le P\Big(\max_{y\in\mathcal{Y}} \big|\hat{p}_n(y) - p(y)\big| > n^{-\beta}/2\Big) \\
&\qquad + P\Big(\max_{y\in\mathcal{Y}} \max_{1\le k<n} \max_{z^0_{-k+1}\in\mathcal{L}^n_k} \big|p(y|z^0_{-k+1}) - \hat{p}_n(y|z^0_{-k+1})\big| > n^{-\beta}/2\Big) \\
&\quad \le P\big(\text{For some } y\in\mathcal{Y}\text{: } \big|\hat{p}_n(y) - p(y)\big| > n^{-\beta}/2\big) \\
&\qquad + P\big(\text{For some } y\in\mathcal{Y}, 1\le k<n, k-1\le l\le n-1\text{: } Y^l_{l-k+1}\in\mathcal{L}^n_k, \\
&\qquad \big|\hat{p}_n(y|Y^l_{l-k+1}) - p(y|Y^l_{l-k+1})\big| > n^{-\beta}/2\big).
\end{aligned}
$$

By the union bound this in turn is:

$$
\begin{aligned}
P(\hat{\Gamma}_n > n^{-\beta}) \le{}& \sum_{y\in\mathcal{Y}} P\big(\big|\hat{p}_n(y) - p(y)\big| > n^{-\beta}/2\big) \\
&+ \sum_{y\in\mathcal{Y}} \sum_{k=1}^{n-1} \sum_{l=k-1}^{n-1} P\big(Y^l_{l-k+1}\in\mathcal{L}^n_k, \big|\hat{p}_n(y|Y^l_{l-k+1}) - p(y|Y^l_{l-k+1})\big| > n^{-\beta}/2\big).
\end{aligned}
$$

By Hoeffding's inequality (cf. Theorem 2 in [4]) for sums of bounded independent random variables, for a given $y\in\mathcal{Y}$,

$$P\big(\big|\hat{p}_n(y) - p(y)\big| > n^{-\beta}/2\big) \le 2\mathrm{e}^{-0.5n^{-2\beta}n}.$$

Summing over y we get that

$$\sum_{y\in\mathcal{Y}} P\big(\big|\hat{p}_n(y) - p(y)\big| > n^{-\beta}/2\big) \le 2|\mathcal{Y}|\mathrm{e}^{-0.5n^{-2\beta}n}.$$

Now for a given $0\le l\le n-1$, $1\le k\le n-1$ and $y\in\mathcal{Y}$ we will give an upper bound on the probability

$$P\big(Y^l_{l-k+1}\in\mathcal{L}^n_k, \big|\hat{p}_n(y|Y^l_{l-k+1}) - p(y|Y^l_{l-k+1})\big| > n^{-\beta}/2\big).$$

Let $l+\theta^+(l,j,i)$ and $l-\theta^-(l,j,i)$ denote the position of the ith occurrence of the pattern Y^l_{l-j} (with length j and position l) going in positive and negative directions respectively. Formally, set $\theta^+(l,j,0)=0$, $\theta^-(l,j,0)=0$ and define

$$\theta^+(l,j,i) = \theta^+(l,j,i-1) + \min\big\{t>0\text{: } Y^{l+\theta^+(l,j,i-1)+t}_{l+\theta^+(l,j,i-1)-j+1+t} = Y^{l+\theta^+(l,j,i-1)}_{l+\theta^+(l,j,i-1)-j+1}\big\}$$

and

$$\theta^-(l,j,i) = \theta^-(l,j,i-1) + \min\big\{t>0\text{: } Y^{l-\theta^-(l,j,i-1)-t}_{l-\theta^-(l,j,i-1)-j+1-t} = Y^{l-\theta^-(l,j,i-1)}_{l-\theta^-(l,j,i-1)-j+1}\big\}.$$

For a given $0\le l\le n-1$, $1\le k\le n-1$, $y\in\mathcal{Y}$, $r\ge 0$ and $s\ge 0$, by Hoeffding's inequality (cf. Theorem 2 in [4]) for sums of bounded independent random variables,

$$
\begin{aligned}
&P\bigg(\bigg|\frac{\sum_{h=1}^r 1_{\{Y_{l-\theta^-(l,k,h)+1}=y\}} + \sum_{h=0}^s 1_{\{Y_{l+\theta^+(l,k,h)+1}=y\}}}{r+s+1} - p(y|Y^l_{l-k+1})\bigg| \ge 0.5n^{-\beta}\bigg| Y^l_{l-k+1} = y^0_{-k+1}\bigg) \\
&\quad \le 2\mathrm{e}^{-0.5n^{-2\beta}(r+s+1)}.
\end{aligned}
$$

Multiplying both sides by $P(Y_{l-k+1}^l = y_{-k+1}^0)$ and summing over all possible $y_{-k+1}^0 \in \mathcal{Y}^k$ we get that

$$P\left(Y_{l-k+1}^l \in \mathcal{L}_k^n, \left|\frac{\sum_{h=1}^r 1_{\{Y_{l-\theta^-(l,k,h)+1}=y\}} + \sum_{h=0}^s 1_{\{Y_{l+\theta^+(l,k,h)+1}=y\}}}{r+s+1} - p\big(y|Y_{l-k+1}^l\big)\right| > 0.5n^{-\beta}\right)$$
$$\leq P\left(\left|\frac{\sum_{h=1}^r 1_{\{Y_{l-\theta^-(l,k,h)+1}=y\}} + \sum_{h=0}^s 1_{\{Y_{l+\theta^+(l,k,h)+1}=0\}}}{r+s+1} - p\big(y|Y_{l-k+1}^l\big)\right| > 0.5n^{-\beta}\right)$$
$$\leq 2\mathrm{e}^{-0.5n^{-2\beta}(r+s+1)}.$$

Summing over all $0 \leq l \leq n-1$ and over all pairs (r, s) such that $r \geq 0$, $s \geq 0$, $r+s+1 \geq \lceil n^{1-\gamma} \rceil$ we get that

$$\sum_{l=0}^{n-1} P\big(Y_{l-k+1}^l \in \mathcal{L}_k^n, \big|\hat{p}_n\big(y|Y_{l-k+1}^l\big) - p\big(y|Y_{l-k+1}^l\big)\big| > n^{-\beta}/2\big) \leq n \sum_{h=\lceil n^{1-\gamma}\rceil}^{\infty} h2\mathrm{e}^{-0.5n^{-2\beta}h}.$$

Thus

$$\sum_{y\in\mathcal{Y}}\sum_{k=0}^{n-1}\sum_{l=k-1}^{n-1} P\big(Y_{l-k+1}^l \in \mathcal{L}_k^n, \big|\hat{p}_n\big(y|Y_{l-k+1}^l\big) - p\big(y|Y_{l-k+1}^l\big)\big| > n^{-\beta}/2\big) \leq 2n^2|\mathcal{Y}| \sum_{h=\lceil n^{1-\gamma}\rceil}^{\infty} h\mathrm{e}^{-n^{-2\beta}h/2}.$$

Now we give an upper bound on the sum on the right-hand side. Observe that $h\mathrm{e}^{-n^{-2\beta}h/2}$ is monotone decreasing in h as soon as the derivative $\mathrm{e}^{-n^{-2\beta}h/2} - h0.5n^{-2\beta}h\mathrm{e}^{-n^{-2\beta}h/2}$ is negative for $h > n^{1-\gamma}$ which is the case for $n > 2^{1/(1-\gamma-2\beta)}$. Using this fact, we bound the sum by the integral

$$\sum_{h=\lceil n^{1-\gamma}\rceil}^{\infty} h\mathrm{e}^{-n^{-2\beta}h/2} \leq \int_{n^{1-\gamma}}^{\infty} h\mathrm{e}^{-n^{-2\beta}h/2}\,\mathrm{d}h.$$

Integrating by parts we get that

$$\begin{aligned}\int_{n^{1-\gamma}}^{\infty} h\mathrm{e}^{-n^{-2\beta}h/2}\,\mathrm{d}h &= \left[h\frac{-1}{n^{-2\beta}/2}\mathrm{e}^{-n^{-2\beta}h/2}\right]_{n^{1-\gamma}}^{\infty} - \int_{n^{1-\gamma}}^{\infty}\frac{-1}{n^{-2\beta}/2}\mathrm{e}^{-n^{-2\beta}h/2}\,\mathrm{d}h\\ &= \frac{n^{1-\gamma}}{n^{-2\beta}/2}\mathrm{e}^{-n^{-2\beta}n^{1-\gamma}/2} - \left[\frac{1}{(n^{-2\beta}/2)^2}\mathrm{e}^{-n^{-2\beta}h/2}\right]_{n^{1-\gamma}}^{\infty}\\ &= \frac{n^{1-\gamma}}{n^{-2\beta}/2}\mathrm{e}^{-n^{1-\gamma-2\beta}/2} + \frac{1}{(n^{-2\beta}/2)^2}\mathrm{e}^{-n^{1-\gamma-2\beta}/2}\\ &= \big(2n^{1-\gamma+2\beta} + 4n^{4\beta}\big)\mathrm{e}^{-n^{1-\gamma-2\beta}/2}\\ &\leq \big(2n^2 + 4n^2\big)\mathrm{e}^{-n^{1-\gamma-2\beta}/2}\end{aligned}$$

since by assumption $0 < \gamma < 1$ and $0 < \beta < \frac{1-\gamma}{2}$.

Combining all these we get that for $n > 2^{1/(1-\gamma-2\beta)}$

$$\begin{aligned}P(g_n = DEP) &= P\big(\hat{\Gamma}_n > n^{-\beta}\big)\\ &\leq 2|\mathcal{Y}|\mathrm{e}^{-0.5n^{1-2\beta}} + 12n^4|\mathcal{Y}|\mathrm{e}^{-n^{1-\gamma-2\beta}/2}\\ &\leq 14n^4|\mathcal{Y}|\mathrm{e}^{-n^{1-\gamma-2\beta}/2}.\end{aligned}$$

The right-hand side is summable provided $2\beta + \gamma < 1$ and the Borel–Cantelli Lemma yields that

$$P\big(\hat{\Gamma}_n \leq n^{-\beta}\text{ eventually}\big) = 1$$

and so $g_n = IND$ eventually almost surely. The proof of Theorem 2 is complete. $\square$

This test can be used as the input needed for Theorem 1. Now let $(\mathcal{X}, \mathcal{F})$ be a countably generated measurable space. As is well known any countably generated measurable space has a refining sequence of finite partitions $\mathcal{A}_k$ such that $\bigcup_{k=1}^{\infty} \hat{\mathcal{A}}_k$ generates the whole σ-algebra $\mathcal{F}$ on $\mathcal{X}$ where $\hat{\mathcal{A}}_k$ denotes the algebra which is generated by partition $\mathcal{A}_k$. Let $g_n^{(k)}$ be the test constructed above for partition $\mathcal{A}_k$. Then choosing $\mathbf{b}$ so as to satisfy the condition in Theorem 1 we have the following corollary.

Corollary 1. *For all stationary and ergodic $\mathcal{X}$-valued process $\{X_n\}$ eventually almost surely, $TEST_n^{\mathbf{b}}(X_0^n) = IND$ if the process is independent and eventually almost surely $TEST_n^{\mathbf{b}}(X_0^n) = DEP$ if the process is not independent.*

References

[1] D. H. Bailey. Sequential schemes for classifying and predicting ergodic processes. Ph.D. thesis, Stanford Univ., 1976. MR2626644

[2] A. Berger. On uniformly consistent tests. *Ann. Math. Statist.* **22** (1951) 289–293. MR0042653

[3] A. Dembo and Y. Peres. A topological criterion for hypothesis testing. *Ann. Statist.* **22** (1994) 106–117. MR1272078

[4] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1963) 13–30. MR0144363

[5] W. Hoeffding and J. Wolfowitz. Distinguishability of sets of distributions. *Ann. Math. Statist.* **29** (1958) 700–718. MR0095555

[6] Ch. Kraft. Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** (1955) 125–141. MR0073896

[7] G. Morvai and B. Weiss. Order estimation of Markov chains. *IEEE Trans. Inform. Theory* **51** (2005) 1496–1497. MR2241507

[8] G. Morvai and B. Weiss. On classifying processes. *Bernoulli* **11** (2005) 523–532. MR2146893

[9] G. Morvai and B. Weiss. Estimating the lengths of memory words. *IEEE Transactions on Information Theory* **54** (2008) 3804–3807. MR2451043

[10] A. Nobel. Hypothesis testing for families of ergodic processes. *Bernoulli* **12** (2006) 251–269. MR2218555

[11] D. Ornstein and B. Weiss. How sampling reveals a process. *Ann. Probab.* **18** (1990) 905–930. MR1062052

[12] B. Weiss. Some remarks on filtering and prediction of stationary processes. *Israel J. Math.* **149** (2005) 345–360. MR2191220