



HAL
open science

An impossibility result for process discrimination

Daniil Ryabko

► **To cite this version:**

Daniil Ryabko. An impossibility result for process discrimination. [Research Report] 2008. inria-00319076v1

HAL Id: inria-00319076

<https://inria.hal.science/inria-00319076v1>

Submitted on 5 Sep 2008 (v1), last revised 10 Jul 2009 (v7)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An impossibility result for process discrimination.

Daniil Ryabko

INRIA, Lille, France. daniil@ryabko.net

Abstract

Two series of binary observations x_1, x_2, \dots and y_1, y_2, \dots are presented: at each time $n \in \mathbb{N}$ we are given x_n and y_n . It is assumed that the sequences are generated independently of each other by two B-processes. We are interested in the question of whether the sequences represent a typical realization of two different processes or of the same one. We demonstrate that this is impossible to decide, in the sense that every discrimination procedure is bound to err with non-negligible frequency when presented with sequences from some B-processes. This contrasts earlier positive results on B-processes, in particular those showing that there are consistent \bar{d} -distance estimates for this class of processes.

1 Introduction

Two series of binary observations x_1, x_2, \dots and y_1, y_2, \dots are presented sequentially. A *discrimination procedure* D is a family of mappings $D_n : X^n \times X^n \rightarrow \{0, 1\}$, $n \in \mathbb{N}$, that maps a pair of samples $(x_1, \dots, x_n), (y_1, \dots, y_n)$ into a binary (“yes” or “no”) answer: the samples are generated by different distributions, or they are generated by the same distribution.

A discrimination procedure D is *asymptotically correct for a set \mathcal{C} of process distributions* if for any two distributions $\rho_x, \rho_y \in \mathcal{C}$ independently generating the sequences x_1, x_2, \dots and y_1, y_2, \dots correspondingly the expected output converges to the correct answer: the following limit exists and the equality holds

$$\lim_{n \rightarrow \infty} \mathbf{E} D_n((x_1, \dots, x_n), (y_1, \dots, y_n)) = \begin{cases} 0 & \text{if } \rho_x = \rho_y \\ 1 & \text{otherwise} \end{cases} .$$

Note that one can consider other notions of asymptotic correctness, for example one can require the output to stabilize on the correct answer with probability 1. The notion of correctness that we consider is perhaps one of the weakest. Clearly, asymptotically correct discriminating procedures exist for many classes of processes, for example for the class of all i.i.d. processes.

Ornstein and Weiss [5] and Ornstein and Shields [4] show that consistent estimates of \bar{d} -distance for B -processes (see definitions below) exist, while it is impossible to estimate this distance outside this class (i.e. in general for stationary ergodic processes). We show that discrimination between B -processes

is impossible, in the sense that any discrimination procedure is bound to err on some processes (the expected answer does not converge to the correct one). This demonstrates that discrimination is harder than distance estimation.

The class of B -processes is sufficiently wide to include, for example, k -order Markov processes and functions of them, but, on the other hand, it is a strict subset of the set of stationary ergodic processes. B -processes play important role in such fields as information theory and ergodic theory [7, 8, 3]. Discrimination procedures for smaller classes of processes, such as the set of i.i.d. processes or various parametric families, exist and are widely studied (see e.g. [1]).

Next we define the \bar{d} distance and B -processes, mainly following [5] in our formulations. For two finite-valued stationary processes ρ_x and ρ_y the \bar{d} -distance $\bar{d}(\rho_x, \rho_y)$ is said to be less than ε if there exists a single stationary process ν_{xy} on pairs (x_n, y_n) , $n \in \mathbb{N}$, such that x_n , $n \in \mathbb{N}$ are distributed according to ρ_x and y_n are distributed according to ρ_y while

$$\nu_{xy}(x_1 \neq y_1) \leq \varepsilon. \quad (1)$$

The infimum of the ε 's for which a coupling can be found such that (1) is satisfied is taken to be the \bar{d} -distance between ρ_x and ρ_y .

A process is called a B -process (or a Bernoulli process) if it is in the \bar{d} -closure of the set of all aperiodic stationary ergodic k -step Markov processes, where $k \in \mathbb{N}$. For more information on \bar{d} -distance and B -processes (including a more conventional ergodic-theory definition and its equivalence to the one above) the reader is referred to [3].

2 The main result

The main result of this work is the following theorem; the construction used in the proof is based on the same ideas as that employed in [6] to demonstrate that consistent prediction for stationary ergodic processes is impossible.

Theorem 1. *There is no asymptotically correct discrimination procedure for the set of all B -processes.*

All the processes in the proof are of the form of partially observable stationary ergodic Markov chains with countably many states. The processes are defined as a deterministic binary function applied to the states of such a chain. Therefore, the result can be formulated as that discrimination between such processes is impossible; such a result may be of its own interest. It is also worth noting that, by slightly modifying the construction in the proof below, one can make all the processes finitarily Markovian, as defined in [2]; therefore consistent discrimination for this class of processes is also impossible.

Proof. We will assume that such a procedure D exists and will construct a B -process ρ such that if both sequences x_i and y_i , $i \in \mathbb{N}$ are generated by ρ then $\mathbf{E}D_n$ diverges; this contradiction will prove the theorem. The (binary) processes that we construct are of the form of a stationary Markov chain with a countably

infinite set of states, with a (deterministic) function mapping each state to $\{0, 1\}$. In other words, the constructions are based on partially observable Markov processes, where the observed variables are from $\{0, 1\}$ and the (non-observable) states are from the set \mathbb{N} . In the course of the proof we will show that the Markov chains and hence the resulting processes are stationary ergodic, and in the end of the proof we explain that all the processes are in fact B -processes.

Assume that there exists an asymptotically correct discriminating procedure D .

Construct the process ρ_0 as follows. A Markov chain m_0 is defined on the set \mathbb{N} of states. From each state $i \in \mathbb{N}$ the chain passes to the state 0 with probability δ and to the state $i + 1$ with probability $1 - \delta$, where $\delta \in [1/2, 1)$ is a parameter to be defined later. With transition probabilities so defined, the chain possesses a unique stationary distribution M_0 on the set \mathbb{N} of states such that $M_0(i) > 0$ for all $i \in \mathbb{N}$ (see e.g. [9]). Take this distribution as the initial distribution over the states. The resulting Markov process m_0 is a stationary ergodic process.

The function f_0 maps the states to the output alphabet $\{0, 1\}$ as follows: $f_0(i) = 1$ for every $i \in \mathbb{N}$. Let s_t be the state of the chain at time t . The process ρ_0 is defined as $\rho_0 = f_0(s_t)$ for $t \in \mathbb{N}$. Since the process m_0 is stationary ergodic, so is the process ρ . In fact, as a result of our definition, the process ρ_0 simply outputs 1 with probability 1 on every time step (however by using different functions f we will have less trivial processes in the sequel). So, we have defined the chain m_0 (and the process ρ_0) up to a parameter δ .

Fix some (small) $\varepsilon > 0$. Since the test is asymptotically correct we will have

$$\mathbf{E}_{\rho_0 \times \rho_0} D_{t_0}((x_1, \dots, x_{t_0}), (y_1, \dots, y_{t_0})) < \varepsilon,$$

for some t_0 , where both samples x_i and y_i are generated by ρ_0 (that is, both samples consist of 1s only).

Let k_0 be such an index that the chain m_0 starting from the state 0 with probability 1 does not reach the state $k_0 - 1$ by time t_0 (we can take $k_0 = t_0 + 2$). Construct two processes ρ_{u1} and ρ_{d1} as follows. They are also based on the Markov chain m_0 , but the functions f are different. The function $f_{u1} : \mathbb{N} \rightarrow \{0, 1\}$ is defined as follows: $f_{u1}(i) = f_0(i) = 1$ for $i \leq k_0$ and $f_{u1}(i) = 0$ for $i > k_0$. The function f_{d1} is identically 1 ($f_{d1}(i) = 1, i \in \mathbb{N}$). The processes ρ_{u1} and ρ_{d1} are defined as $\rho_{u1} = f_{u1}(s_t)$ and $\rho_{d1} = f_{d1}(s_t)$ for $t \in \mathbb{N}$. Thus the process ρ_{d1} will again produce only 1s, but the process ρ_{u1} will occasionally produce 0s.

Being run on two samples generated by ρ_{u1} and ρ_{d1} the test D_n on the first t_0 steps with high probability (that is, at least if both processes start at the state 0) produces many 0s, since on these first k_0 states all the functions f , f_{u1} and f_{d1} coincide. However since the processes are different and the test is asymptotically correct (by assumption), the test starts producing 1s, until by a certain time step t_1 almost all answers are 1s. Next we will construct the process ρ_2 by “gluing” together ρ_{u1} and ρ_{d1} and continuing them in such a way that, being run on two samples produced by ρ_2 the test first produces 0s (as

if the samples were drawn from ρ_0), then, with probability close to $1/2$ it will produce many 1s (as if the samples were from ρ_{u1} and ρ_{d1}) and then again 0s. Furthermore, we will continue this construction eventually obtaining a process on which the expected answer of the test will diverge.

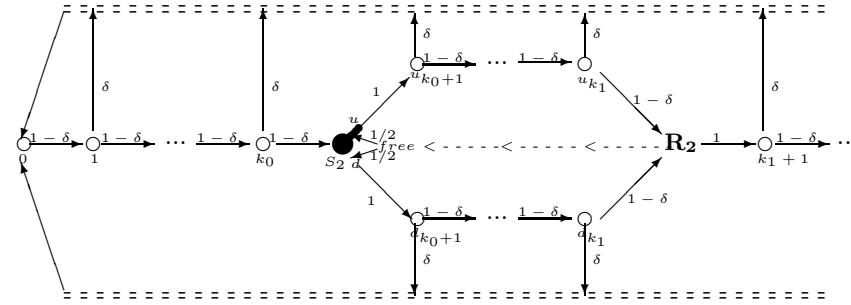
Let $t_1 > t_0$ be such a time index that

$$\mathbf{E}_{\rho_{u1} \times \rho_{d1}} D_k((x_1, \dots, x_{t_1}), (y_1, \dots, y_{t_1})) > 1 - \varepsilon,$$

where the samples x_i and y_i are generated by ρ_{u1} and ρ_{d1} correspondingly (the samples are generated independently; that is, the process are based on two independent copies of the Markov chain m). Let k_1 be such an index that the chain m starting from the state 0 with probability 1 does not reach the state $k_1 - 1$ by time t_1 .

Construct the process ρ_2 as follows (see fig. 1). It is based on a chain m_2 on

Figure 1: The processes m_2 and ρ_2 . The states are depicted as circles, the arrows symbolize transition probabilities: from every state the process returns to 0 with probability δ or goes to the next state with probability $1 - \delta$. The function f_2 is 1 on all states except $u_{k_0+1}, \dots, u_{k_1}$ where it is 0; f_2 applied to the states output by m_2 defines ρ_2 .

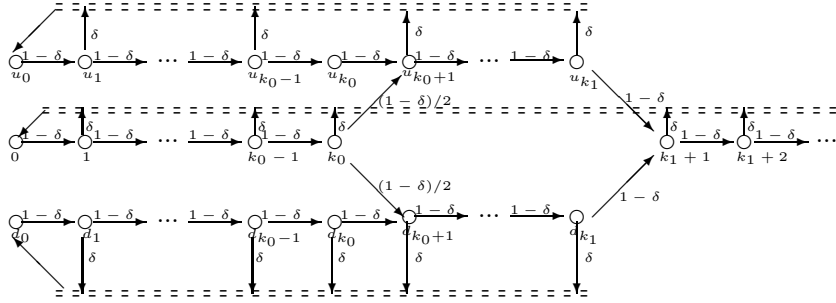


which Markov assumption is violated, but we will show that m_2 is equivalent to a Markov chain. For now assume the process starts at the state 0. The process m_2 is constructed as follows. The transition probabilities on states $0, \dots, k_0$ are the same as for the Markov chain m (from each state return to 0 with probability δ or go to the next state with probability $1 - \delta$). There are two “special” states: the “switch” S_2 and the “reset” R_2 . From state k_0 the chain passes with probability $1 - \delta$ to the “switch” state S_2 . The switch S_2 can itself have 3 values: u, d or $free$. If S_2 has the value u then from S_2 the chain passes to the state u_{k_0+1} , while if $S_2 = d$ the chain goes to d_{k_0+1} , with probability 1. In these cases S_2 does not change its value. If $S_2 = free$ then S_2 takes the value u or d with equal probabilities and passes either to u_{k_0+1} or to d_{k_0+1} accordingly. If the chain reaches the state R_2 then the value of S_2 is set to $free$. For now assume that the initial value of S_2 is $free$. In other words, the first transition from S_2 is random (either to u_{k_0+1} or to d_{k_0+1} with equal probabilities) and then this decision is remembered until the “reset” state R_2 is visited. From each

state u_i , $k_0 \leq i \leq k_1$ the chain passes to the state 0 with probability δ and to the next state u_{i+1} with probability $1 - \delta$. From the state u_{k_1} the process goes with probability δ to 0 and with probability $1 - \delta$ to the “reset” state R_2 . The same with states d_i : for $k_0 < i \leq k_1$ the process returns to 0 with probability δ or goes to the next state d_{i+1} with probability $1 - \delta$, where the next state for d_{k_1} is the “reset” state R_2 . From R_2 the process goes with probability 1 to the state $k_1 + 1$ where from the chain continues ad infinitum: to the state 0 with probability δ or to the next state $k_1 + 2$ etc. with probability $1 - \delta$. The function f_2 is defined as follows: $f_2(i) = 1$ for $0 \leq i \leq k_0$ and $i > k_1$ (before the switch and after the reset); $f_2(u_i) = 0$ for all i , $k_0 < i \leq k_1$ and $f_2(d_i) = 1$ for all i , $k_0 < i \leq k_1$. The function f_2 is undefined on S_2 and R_2 , therefore there is no output on these states (we also assume that passing through S_2 and R_2 does not increment time). As before, the process ρ_2 is defined as $\rho_2 = f_2(s_t)$ where s_t is the state of m_2 at time t , omitting the states S_2 and R_2 . The resulting process is illustrated on fig. 1.

Before we define the initial distribution on the states of the process m_2 , we will show that it is equivalent to a Markov chain. Indeed, construct the Markov chain m'_2 as follows (see fig. 2). This chain has states $0, \dots, k_0, k_1 + 1, \dots$ and also $u_0, \dots, u_{k_0}, u_{k_0+1}, \dots, u_{k_1}$ and $d_0, \dots, d_{k_0}, d_{k_0+1}, \dots, d_{k_1}$. Transitions on

Figure 2: The process m'_2 . The function f_2 is 1 everywhere except the states $d_{k_0+1}, \dots, d_{k_1}$ where it is 0.



the states 0 to $k_0 - 1$ are defined in the same way as for all the chains described before. From the state k_0 the chain passes with probability $(1 - \delta)/2$ to the state u_{k_0+1} and with probability $(1 - \delta)/2$ to the state d_{k_0+1} , while with probability δ it returns to 0 ; thus the state k_0 corresponds to the *free* state of the switch S_2 . From the states u_i , $i = 0, \dots, k_1$ the chain passes with probability $1 - \delta$ to the next state u_{i+1} , where the next state for u_{k_1} is $k + 1$ and with probability δ returns to the state u_0 (and not to the state 0). Transitions for the state d_0, \dots, d_{k_1-1} are defined analogously. Thus the state u_{k_0} corresponds to the state u of the switch S_2 and the state d_{k_0} — to the state d of the switch. Transitions for the states $k + 1, k + 2, \dots$ are defined as before: with probability δ to 0 and with probability $1 - \delta$ to the next state. Thus the state $k_1 + 1$ corresponds to the reset R_2 . Clearly, the chain m'_2 as defined possesses a unique

stationary distribution M_2 over the set of states and $M_2(i) > 0$ for every state i ; take this distribution as its initial distribution. The resulting process m'_2 is stationary ergodic. The initial distribution M_2 can be used to define the initial distribution on the original (non-Markov) process m_2 , with the initial value of the switch S_2 random: if m'_2 starts at a state u_i where $i \leq k_0$ this corresponds to the process m'_2 starting at the state i with S_2 set to u , and analogously for d . If m'_2 starts at a state j where $j \leq k_0$ or $j > k_1$ this corresponds to the process m_2 starting at the state j with S_2 set to *free*; the initial probabilities for S_2 and R_2 themselves are 0. Thus, we have defined (in two equivalent ways) the process ρ_2 , which is stationary ergodic.

As before, we can continue the construction of the processes ρ_{u3} and ρ_{d3} , that start with a segment of ρ_2 . Let $t_2 > t_1$ be a time index such that

$$\mathbf{E}_{\rho_2 \times \rho_2} D_{t_2} < \varepsilon,$$

where both samples are generated by ρ_2 . Let k_2 be such an index that when starting from the state 0 the process m_2 with probability 1 does not reach $k_2 - 1$ by time t_2 (equivalently: the process m'_2 does not reach $k_2 - 1$ when starting from either 0, u_0 or d_0). The processes ρ_{u3} and ρ_{d3} are based on the same process m_2 as ρ_2 . The functions f_{u3} and f_{d3} coincide with f_2 on all states up to the state k_2 (including the states u_i and d_i , $k_0 < i \leq k_1$). After k_2 the function f_{u3} outputs 0s while f_{d3} outputs 1s: $f_{u3}(i) = 0$, $f_{d3}(i) = 1$ for $i > k_2$. Furthermore, we find an index k_3 such that the process m_2 does not reach $k_3 - 1$ before the time $t_3 > t_2$ by which we have $\mathbf{E}_{\rho_{u3} \times \rho_{d3}} D_{t_3} > 1 - \varepsilon$, where the samples are generated by ρ_{u3} and ρ_{d3} . We then construct the process ρ_4 based on a (non-Markovian) process m_4 by “gluing” together ρ_{u3} and ρ_{d3} after the step k_3 with a switch S_4 and a reset R_4 exactly as was done when constructing the process ρ_2 . The process m_4 is illustrated on fig. 3 a). The process m_4 can be shown to be equivalent to a Markov chain m'_4 (cf. fig. 3 b)), which consists of 3 copies of the process m_2 (which was shown to be equivalent to a Markov chain) truncated at step k_2 and linked as follows. From the step k_2 the process passes with probability $1/2$ to the state u_{k_2+1} with probability $(1 - \delta)/2$ and to the state d_{k_2+1} with the same probability, while with probability δ it goes to the origin 0. From states u_i , $k_2 < i \leq k_3$ the process returns to u_0 with probability δ or goes to the next state $u_i + 1$ with probability $1 - \delta$, and analogously for d_i . For the states u_{k_3} and d_{k_3} the next state is $k_3 + 1$, from which the process returns to the state 0 with probability δ or continues to the next state with probability $1 - \delta$. The definition of f_4 is analogous to the previous definitions of f_i .

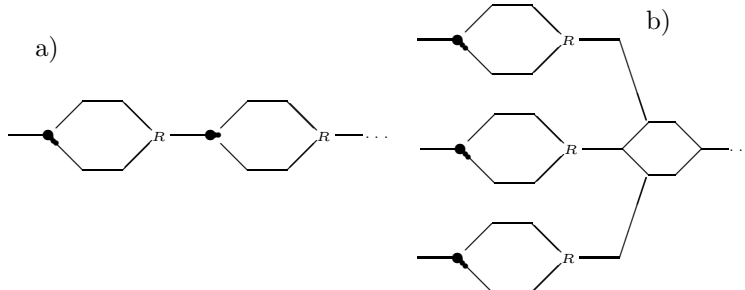
Proceeding this way we can construct the processes ρ_{2j} , ρ_{u2j+1} and ρ_{d2j+1} , $j \in \mathbb{N}$ choosing the time steps $t_j > t_{j-1}$ so that the test converges to 0 by t_j being run on two samples produced by ρ_j for even j , and converges to 1 by t_j being run on samples produced by ρ_{u_j} and ρ_{d_j} for odd j :

$$\mathbf{E}_{\rho_{2j} \times \rho_{2j}} D_{t_{2j}} < \varepsilon \tag{2}$$

and

$$\mathbf{E}_{\rho_{2j+1} \times \rho_{2j+1}} D_{t_{2j+1}} > (1 - \varepsilon). \tag{3}$$

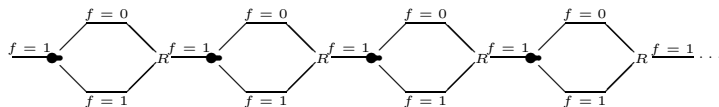
Figure 3: The processes m_4 and m'_4



For each j the number k_j is selected in a such a way that the state $k_j - 1$ is not reached (with probability 1) by the time t_j when starting from the state 0.

Having defined k_j , $j \in \mathbb{N}$ we can define the process ρ . It is defined as a limiting structure that would result by constructing ρ_{2j} , $j \in \mathbb{N}$ (see fig. 4). More

Figure 4: The processes m_ρ and ρ . The states are on horizontal lines. The function f that defines the process ρ takes value 0 on the states on the upper lines and 1 on the rest of the states.



precisely, it is based on the process m_ρ that has states $0, \dots, k_0, k_{2j+1}, \dots, k_{2j+2}, u_{k_{2j}}, \dots, u_{k_{2j+1}}$ and $d_{k_{2j}}, \dots, d_{k_{2j+1}}$ for $j \in \mathbb{N}$, along with switch states S_{2j} and reset states R_{2j} . Each switch S_{2j} diverts the process to the state $u_{k_{2j+1}}$ if the switch has value u and to $d_{k_{2j+1}}$ if it has the value d . If the switch has the value *free* it assumes one of the values u or d with equal probabilities. The reset R_{2j} sets S_{2j} to *free*. From the states between a reset and the next switch, that is, from the states $k_{2j+1}, \dots, k_{2j+2}$, the process returns to the state 0 with probability δ or continues to the next state. The function f is defined as 1 everywhere except for the states u_j (for all $j \in \mathbb{N}$ for which u_j is defined) on which f takes the value 0. The process ρ is defined at time t as $f(s_t)$, where s_t is the state of m_ρ at time t . As before, we can show that the process m_ρ is equivalent to a Markov chain, by introducing copies of all preceding states for each new switch S_j ; since the return state after each new reset R_j is the original state 0, the Markov chain so defined possesses a unique stationary distribution M such that $M(i) > 0$ for all states i . This in turn defines a distribution on the states of m_ρ , including the states of the switches. Taken as the initial state distribution this completes the definition of m_ρ and ρ . Since ρ is defined as an application of a (deterministic) function to the output of (a process that is equivalent to) a stationary ergodic Markov chain, it is itself stationary ergodic.

Finally, we have to show that the expected output of test D_k diverges if the

test is run on two independent samples produced by ρ .

Let M_{2j} denote the initial state distribution of the process m_{2j} , $j \in \mathbb{N}$, and M that of the process m_ρ . Since each of the processes m_{2j} , $j \in \mathbb{N}$ and the process m_ρ from each state returns to the state 0 with probability δ the limiting (and hence initial) probability of this state is δ : $M_{2j}(0) = M(0) = \delta$. By construction, if the process m_ρ starts at the state 0 then up to time k_{2j} it behaves exactly as ρ_{2j} that has started at state 0; more formally, we have

$$E_{\rho \times \rho}(D_{t_{2j}} | s_0^x = 0, s_0^y = 0) = E_{\rho_{2j} \times \rho_{2j}}(D_{t_{2j}} | s_0^x = 0, s_0^y = 0) \quad (4)$$

for $j \in \mathbb{N}$, where s_0^x and s_0^y denote the states of the processes generating the samples x and y correspondingly.

Using a simple decomposition

$$\mathbf{E}(D_{t_j}) = \delta^2 \mathbf{E}(D_{t_j} | s_0^x = 0, s_0^y = 0) + (1 - \delta^2) \mathbf{E}(D_{t_j} | s_0^x \neq 0 \text{ or } s_0^y \neq 0), \quad (5)$$

(4), and (2) we have

$$\begin{aligned} \mathbf{E}_{\rho \times \rho}(D_{t_{2j}}) &\leq \delta^2 \mathbf{E}_{\rho \times \rho}(D_{t_{2j}} | s_0^x = 0, s_0^y = 0) + (1 - \delta^2) \\ &= \delta^2 \mathbf{E}_{\rho_{2j} \times \rho_{2j}}(D_{t_{2j}} | s_0^x = 0, s_0^y = 0) + (1 - \delta^2) \\ &\leq \mathbf{E}_{\rho_{2j} \times \rho_{2j}} + (1 - \delta^2) < \varepsilon + (1 - \delta^2). \end{aligned} \quad (6)$$

For odd indices, if the process ρ starts at the state 0 then (from the definition of t_{2j+1}) by the time t_{2j+1} it will have passed the switch S_{2j} not more than once. Since the definition of m_ρ is symmetric with respect to the values u and d of each switch, the probability that two samples $x_1, \dots, x_{t_{2j+1}}$ and $y_1, \dots, y_{t_{2j+1}}$ generated independently by (two runs of) the process ρ produced different values of the switch S_{2j} when passing through it for the first time is $1/2$. In other words, with probability $1/2$ two samples generated by ρ starting at the state 0 will look by the time t_{2j+1} as two samples generated by $\rho_{u_{2j+1}}$ and $\rho_{d_{2j+1}}$ that has started at state 0. Thus

$$E_{\rho \times \rho}(D_{t_{2j+1}} | s_0^x = 0, s_0^y = 0) \geq \frac{1}{2} E_{\rho_{u_{2j+1}} \times \rho_{d_{2j+1}}}(D_{t_{2j+1}} | s_0^x = 0, s_0^y = 0) \quad (7)$$

for $j \in \mathbb{N}$. Using this, (5), and (3) we obtain

$$\begin{aligned} \mathbf{E}_{\rho \times \rho}(D_{t_{2j+1}}) &\geq \delta^2 \mathbf{E}_{\rho \times \rho}(D_{t_{2j+1}} | s_0^x = 0, s_0^y = 0) \\ &\geq \frac{1}{2} \delta^2 \mathbf{E}_{\rho_{2j+1} \times \rho_{2j+1}}(D_{t_{2j+1}} | s_0^x = 0, s_0^y = 0) \\ &\geq \frac{1}{2} (\mathbf{E}_{\rho_{2j+1} \times \rho_{2j+1}}(D_{t_{2j+1}}) - (1 - \delta^2)) > \frac{1}{2} (\delta^2 - \varepsilon). \end{aligned} \quad (8)$$

Taking δ large and ε small (e.g. $\delta = 0.9$ and $\varepsilon = 0.1$), we can make the bound (6) close to 0 and the bound (8) close to $1/2$, and the expected output of the test will cross values infinitely often.

It remains to show that the process ρ constructed in the proof is a B -processes. Indeed, this process is obtained by applying the function f to the

states of a chain m_ρ , that has a property that the sum of probabilities (in the initial stationary distribution) of all of the states that follow the state k is not greater than 2^{-k} . Let g_k be a function that coincides with f on all states up to k and is 0 for all states that follow k . The process μ_k obtained by applying g_k to the states of m_ρ is equivalent (that is, the distributions are the same) as of that which would be obtained by applying g_k to a k' -order Markov chain (for some $k' < 2^k$) constructed by replacing all states of m'_ρ greater than k by a single state, from which the chain passes to the state 0 with probability 1/2 and with probability 1/2 remains in this state. Therefore, each g_k is a function of an aperiodic stationary ergodic k -order Markov chain, and therefore it is a B -process (see e.g. [3]). Moreover, ρ is a limit in \bar{d} distance of the processes μ_k ; indeed, if we couple ρ with μ_k in an obvious way by requiring that the underlying Markov chains always take state transitions together, then the probability of observing a different output is not greater than the probability that the chains are in one of the states that follow the state k , which is bounded by 2^{-k} . Since the set of B -processes is closed in the \bar{d} -distance, the process ρ is a B -process. \square

References

- [1] Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. Springer.
- [2] G. Morvai and B. Weiss. *On classifying processes*, Bernoulli 11(3), pp. 523–532, 2005.
- [3] Ornstein, D. S. (1974). *Ergodic theory, randomness, and dynamical systems*. Yale Mathematical Monographs 5, Yale Univ. Press, New Haven, CT.
- [4] Ornstein, D. S., Shields, P.(1994). *The \bar{d} -recognition of processes*, Advances in Mathematics, 104, pp. 182-224.
- [5] Ornstein, D. S. and Weiss, B.(1990). *How Sampling Reveals a Process*. Annals of Probability 18(3), pp. 905–930.
- [6] Ryabko, B.(1988). *Prediction of random sequences and universal coding*. Problems of Information Transmission, Vol. 24, pp. 87–96.
- [7] Shields, P.(1998). *The Interactions Between Ergodic Theory and Information Theory*. IEEE Trans. on Information Theory, vol. 44, no. 6 , pp. 2079–2093.
- [8] Shields, P. (1996) *The Ergodic Theory of Discrete Sample Paths*. AMS Bookstore.
- [9] Shiryaev, A. (1996). *Probability*. Springer.