The Decomposition-Separation Theorem for Finite Nonhomogeneous Markov Chains and Related Problems

Isaac M. Sonin[1]

University of North Carolina at Charlotte

Abstract: Let M be a finite set, P be a stochastic matrix and $U = \{(Z_n)\}$ be the family of all finite Markov chains (MC) (Z_n) defined by M, P, and all possible initial distributions. The behavior of a MC (Z_n) is a classical result of probability theory derived in the 1930s by A. N. Kolmogorov and W. Doeblin. If a stochastic matrix P is replaced by a sequence of stochastic matrices (P_n) and transitions at moment n are defined by P_n, then U becomes a family of *nonhomogeneous* MCs. There are numerous results concerning the behavior of such MCs given some specific properties of the sequence (P_n). But what if *there are no assumptions* about sequence (P_n)? Is it possible to say something about the behavior of the family U? The surprising answer to this question is Yes. Such behavior is described by a theorem which we call a decomposition-separation (DS) theorem, and which was initiated by a small paper of A. N. Kolmogorov (1936) and formulated and proved in a few stages in a series of papers including D. Blackwell (1945), H. Cohn (1971, 1989), and I. Sonin (1987, 1991, 1996).

1. Introduction

The notion of a (homogeneous) Markov chain (MC) is one of the key notions in the theory of stochastic processes and in probability theory. Its simplest version, the case of discrete time and finite state space, is specified by a pair (M, P), where M is a state space and $P = \{p(i,j)\}$ is a transition matrix indexed by the elements of M. We denote by $Z = (Z_n)$, $n \in \mathbf{N} = \{0, 1, \ldots\}$, a MC from a family U_0 of all MCs defined by M, P, and all initial distributions on M.

The classical Kolmogorov–Doeblin results describing the decomposition of a state space M into essential and nonessential (transient) states, into ergodic classes and cyclic subclasses, and the asymptotic behavior of MCs from U_0 can be found in most advanced books on probability theory as well as the monographs on MC (see for example Shiryayev [33], Kemeny and Snell [16], Isaacson and Madsen [15]).

If a MC is irreducible and aperiodic then an ergodic property holds, i.e., there exists a limit (invariant) distribution π such that

$$\lim_{n \to \infty} P(Z_n = j \mid Z_0 = i) = \pi(j) > 0, \tag{1}$$

which does not depend on the initial state i. If the number of cyclic subclasses exceeds one, then the MC is aperiodic when considered only at the times of visiting the given subclass, and (1) is true for corresponding n.

---
[1]Department of Mathematics, UNC at Charlotte, 9201 University Bld, Charlotte, NC 28223, e-mail: `imsonin@uncc.edu`

These results, of course, represent only the basic facts about the structure of MCs and many more detailed and subtle properties of MCs are contained in the rich and extensive theory of this subject.

A natural extension of this theory is a theory of *nonhomogeneous* MCs when the transitions at moment n are defined by a stochastic matrix P_n from a sequence of stochastic matrices (P_n). We denote by U the family of all nonhomogeneous MCs specified by M, (P_n), and all initial distributions on M now specified not only for an initial moment 0 but for all initial moments $k = 0, 1, 2, \ldots$. We again denote by $Z = (Z_n)$, $n \geq k$, a MC from this family. There is a substantial body of literature on nonhomogeneous MCs, though this is still a small fraction of the literature on homogeneous MCs. See e.g. the classical works of R. Dobrushin [8], D. Griffeath [10], J. Hajnal [11], D. Isaacson and R. Madsen [15], M. Iosifescu [14], J. Kingman [17], V. Maksimov [20], A. Mukherjea [22], E. Seneta [25], and others. A survey of results about the products of stochastic matrices together with his own contributions can be found in D. Hartfiel [12]. More recent publications are e.g. [7], [19], and [32]. In the last ten years interest in this area has surged because an important class of computational algorithms, so called simulated annealing, is based on nonhomogeneous MCs with very specific transition probabilities. Namely, the transition probabilities have the form $p_n(i, j) = c(i, j) \exp\{-q(i, j)/T_n\}$ for $i \neq j$, where $\{q(i, j)$ is a nonnegative matrix defined by an optimization problem, and T_n is a "temperature" that tends to zero. From the vast literature on this topic we mention only two papers, W. Niemiro and P. Pokarowski [23] and H. Cohn and M. Fielding [5], which are more closely related to our paper.

Almost all literature mentioned above follows a rather natural format—given some assumptions about the structure of the sequence (P_n), some results about the behavior of the corresponding family (Z_n) are obtained. A natural question, which was not asked for a long time, is as follows: Is it possible to say something about the behavior of the family U if *there are no assumptions* about the sequence (P_n)? At first sight, the answer seems to be negative, especially if we take into account that this question is equivalent to the question of how the products of matrices $P_1 P_2 \cdots P_n$ behave when n tends to infinity, if the only information available about these matrices is that they are stochastic. Nevertheless, surprisingly the answer to this question is affirmative: There is a fundamental theorem which describes such behavior. In particular this theorem generalizes the above mentioned results of Kolmogorov–Doeblin about homogeneous MCs.

We call this theorem a decomposition-separation (DS) theorem, and briefly, this theorem states that a decomposition with properties similar to that of homogeneous MCs does exist but now it is not a decomposition of the state space M, but a decomposition of the space-time representation of M, i.e., of the sequence $(M_n) = M \times \mathbf{N}$. The only assumption is that the set M is finite, $|M| = N < \infty$.

The DS theorem was initiated by a small paper of A. N. Kolmogorov [18] who analyzed the situation when a sequence of stochastic matrices (P_n) is given in *inverse* time, i.e., for $n = 0, -1, -2, \ldots$. This paper is known mainly by the reversibility criterion introduced there but besides this result Kolmogorov asked two questions. First, does there exist a MC (Z_n) governed by this sequence, i.e., satisfying equalities $P(Z_{n+1} = j \mid Z_n = i) = p_n(i, j)$ for all n, i, j. In a few lines Kolmogorov answered this question positively. The second question was: When is such a MC unique? Kolmogorov proved that a necessary and sufficient condition for the uniqueness is that the limits

$$(2) \qquad\qquad \lim_{n \to -\infty} P(Z_m = j \mid Z_n = i) = \pi_m(j),$$

exist for all m, j and do not depend on the initial point i when n tend to minus infinity.

A breakthrough step to the description of *all* MCs defined by a sequence (P_n) was made in 1945 by David Blackwell [2]. In our terms his description can be explained as follows. Let us introduce a sequence (M_n) of disjoint copies of the state space M, e.g., $M_n = (M, n)$, $n \leq 0$. Without loss of generality we can assume that the stochastic matrices (P_n) are indexed by the elements of these sets, i.e., $P_n = \{p_n(i, j), \ i \in M_n, \ j \in M_{n+1}\}$. A sequence $J = (J_n)$, $J_n \subset M_n$, $n \leq 0$, for brevity is called a *jet*. A tuple of jets $(J^1, ..., J^c)$ is called a *partition* of (M_n) if correspondingly $(J_n^1, \ldots, J_n^c)$ form a partition of M_n for every n. Blackwell proved that there is a partition $(T^0, T^1, \ldots, T^c)$ of (M_n) such that for any MC $Z \in U$ the trajectories of this MC with probability one will reach and stay eventually in one of the jets T^i, $i = 1, \ldots, c$, i.e., $(\limsup(Z_n \in T_n^i)) = (\liminf(Z_n \in T_n^i))$, and for MCs "inside" of one of these jets there are limits similar to (2). In T^0 such limits may not exist but $P(\limsup(Z_n \in T_n^0)) = 0$. The decisive point of his proof was the use of the existence of limits for almost all trajectories of bounded (sub)martingales, then a relatively new result of his Ph.D. advisor J. Doob. As Kolmogorov did, Blackwell considered MCs in reverse time.

The next step was made in the works of Harry Cohn (see [3], [4], and the expository paper [6]). Cohn considered *forward* time, proved that the tail σ-algebra of any nonhomogeneous MC consists of a finite number $c \leq N$ of atomic (indecomposable) sets, each of them related with an element T^k of Blackwell's decomposition, $k = 1, \ldots, c$. He also simplified Blackwell's proof, though it was still very complicated. Note that the jets (T_n^i) in the Blackwell–Cohn decomposition are defined up to jets (J_n) such that $P(\limsup(Z_n \in J_n)) = P(\liminf(Z_n \in J_n)) = 0$, so generally there is a continuum of such partitions.

The last step in the proof of the DS theorem was made by the author in a series of papers Sonin [26], [27], [28], [29], [30], where it was proved that among the Blackwell–Cohn partitions there are partitions into jets having the additional property that the *expected number of transitions* of trajectories of any MC (Z_n) between jets is *finite* on the infinite time interval. This additional separation property was not obvious and its existence was not noted or mentioned before. At the same time it played a crucial role in the initial problem that led the author to the formulation of the separation property, the problem of sufficiency of Markov strategies for the Dubins–Savage functional. An example of such a functional is the probability of visiting a given subset of the state space infinitely often. The study of the problem of sufficiency of Markov strategies led to the study of equivalent random sequences and to the proof of the so called Feinberg inequality (see [29]). In this paper the initial proof of sufficiency given for the finite case by T. Hill [13] was substantially simplified. Note also that the problem of sufficiency of Markov strategies is still an open problem for the countable case.

The DS theorem also has a simple *deterministic* interpretation in terms of the behavior of the simplest model of an *irreversible process* represented by a system of N cups filled with a liquid with initial concentration of a second substance and mixed at discrete moments of time in some proportions defined at each moment n by a stochastic matrix P_n. The irreversibility of this process manifests itself in the martingale property of some bounded random sequences defined by the family of MCs U. Since the state space M is finite, $|M| = N < \infty$, these (sub)(super)martingales take no more than N values at each moment of time. Such martingales have extra properties that do not follow from the well-known Doob's upcrossing lemma. The generalization of Doob's lemma for these martingales, the theorem about the

existence of "barriers," published in [26] played a crucial role in the proof of the separation property. A survey of corresponding results and their interrelationship was given in [30].

The main goal of this expository paper is to complement this survey by presenting a more refined version of the DS theorem, to give an answer to an open problem, and to give a sketch of the proofs of the DS theorem and "barriers" theorem.

The plan of this paper is: In Section 2 we present the deterministic version of DS theorem and its probabilistic counterpart; in Section 3 we discuss the third key ingredient—the martingale type (i.e., martingales or (sub)supermartingales) random sequences. In Section 4 we outline the sketch of the proof of DS theorem and we will show why the separation part needs the strengthening of Doob's up-crossing lemma—the theorem about existence of barriers. Section 5 will outline the construction of barriers. Section 6 is about open problems related to the DS theorem.

2. A simple model of an irreversible process, two formulations of the DS theorem

We assume that two sequences (M_n) and (P_n) are given, where a set M_n represents the state space at moment n. The stochastic matrices (P_n) are indexed by the elements of these sets, i.e., $P_n = \{p_n(i,j), \ i \in M_n, j \in M_{n+1}\}$. At this stage it is does not matter whether the sets M_n are countable or finite, though the DS theorem holds only under the assumption

$$(3) \qquad |M_n| \leq N < \infty, \quad n \in \mathbf{N}.$$

The following simple physical model and physical interpretation of the DS theorem for a particular Markov chain was introduced in [26]. Given a sequence (M_n), let M_n represent a set of "cups" containing a "liquid"—tea, schnapps, vodka, etc. A cup $i \in M_n$ is characterized at moment n by a *volume* of liquid in this cup, $m_n(i)$. The matrix P_n describes the redistribution of liquid from the cups M_n to the (initially empty) cups M_{n+1} at the time of the nth transition, i.e., $p_n(i,j)$ is the proportion of liquid transferred from cup i to cup j. The sequence (m_n), $m_n = (m_n(i), i \in M_n)$, $n \in \mathbf{N}$, for the sake of brevity called the (discrete) *flow*, satisfies the relations

$$(4) \qquad m_{n+1}(j) = \sum_i m_n(i) p_n(i,j).$$

We assume that for some $k \in \mathbf{N}$ an initial condition $m_k(i)$, $i \in M_k$, is given and without loss of generality $\sum_i m_k(i) = 1$, and hence a similar equality holds for any $n \geq k$. In Section 1 we introduced U, the family of all (nonhomogeneous) Markov chains (MC) $Z = (Z_n)$, $n \in \mathbf{N}$, specified by (M_n) and (P_n) and all possible initial distributions μ on M_k, $k = 0, 1, \ldots$. Given a MC (Z_n) we can define a flow (m_n), with $m_n(i) = P(Z_n = i)$, $i \in M_n$, $n \geq k$. Vice versa, a flow (m_n), $m_n = (m_n(i), i \in M_n)$, satisfying (4) defines a MC $Z \in U$. Thus we have

Proposition 1. *There is a one-to-one correspondence between MCs $(Z_n) \in U$ and flows (m_n).*

Let us assume additionally that each cup contains some material (substance, color), and let us denote $\alpha_n(i)$, $0 \leq \alpha \leq 1$, a "*concentration*" of this material at cup i at moment n. The sequence $(m_n, \alpha_n) = (m_n(i), \alpha_n(i))$, $i \in M_n$, $n \in \mathbf{N}$,

for the sake of brevity is called a *colored* (discrete) *flow*. Concentrations obviously satisfy the relations

$$(5) \qquad \alpha_{n+1}(j) = \sum_i m_n(i)\alpha_n(i)p_n(i,j)/m_{n+1}(j).$$

Note that we can replace the notion of concentration by *temperature* since it follows the same formula (5). One more interpretation is obtained if we consider $m_n(i)$ as masses and $\alpha_n(i)$ as their positions on a horizontal axis. The mixing is replaced by taking the center of gravity of corresponding subsystem. We can also use not one but many colors and so on.

The initial conditions $m_k(i), \alpha_k(i), i \in M_k$, for some $k \in \mathbf{N}$ are assumed given and the sequence $m_n(i), \alpha_n(i), i \in M_n$, evolve in time according to (4) and (5) for $n \geq k$. If we introduce $s_n(i) = m_n(i)\alpha_n(i)$, the amount of "substance" contained at cup i at moment n, and we denote by m_n and s_n the corresponding row vectors, then the equations (4) and (5) can be presented in the symmetrical form

$$(6) \qquad m_{n+1} = m_n P_n, \quad s_{n+1} = s_n P_n, \quad n = 0, 1, \ldots.$$

The colored flow described above is probably the simplest example of an *irreversible process*, i.e., a process such that no state can be repeated after a few steps, or in other words a process for which any sequence of states in reversed time is not an admissible sequence in a forward time. Such property holds of course except in the trivial cases when all concentrations become equal after some moment, or when the redistribution avoids any mixing. Intuitively the property of irreversibility seems obvious. The formal proof is based on the consideration e.g. of the following function describing the state of a system at moment n:

$$(7) \qquad F_n(m_n, \alpha_n) = \sum_i m_n(i)\alpha_n^2(i).$$

It is easy to prove that for any colored flow function F_n is nonincreasing and $F_n = F_{n+1}$ only if there no mixing at moment n. As we will see later this function is equivalent to a variance of some random sequence.

The colored flows also have a simple *probabilistic* interpretation. Let $(Z_n) \in U$, $n \geq k$, be a Markov chain and set $D_k \subset M_k$. Let us denote

$$(8) \qquad \alpha_n(i) = P(Z_k \in D_k \mid Z_n = i).$$

It is easy to check that the sequence $(m_n(i), \alpha_n(i))$, $n \geq k$, specifies a colored flow with initial values $\alpha_k(i) = 1$ for $i \in D_k$, $\alpha_k(i) = 0$ otherwise. Vice versa, for every colored flow (m_n, α_n) with initial data of concentrations equal to zero or one at the initial moment k, there is a pair $((Z_n), D_k)$, where $(Z_n) \in U$, $n \geq k$, $D_k \subseteq M_k$, such that $\alpha_n(i)$ coincide with values given by (8).

We will consider also a slightly more general colored flows which allows a jet (O_n), $O_n \subseteq M_n$, $n \in \mathbf{N}$, called an "*ocean*," where by definition for $i \in D_n = M_n \setminus O_n$,

$$(9) \qquad \alpha_n(i) = P(Z_s \in D_s, s = k, \ldots, n \mid Z_n = i),$$

and for $i \in O_n$, $n \in \mathbf{N}$, we define $\alpha_n(i) \equiv 0$. If $Z = (Z_n)$, $n \geq k$, is a MC and (D_n) is a sequence of sets, $D_n \subseteq M_n$, $n \geq k$, we call (Z_n, D_n) a *Markov pair*.

Proposition 2. *There is a one-to-one correspondence between Markov pairs (Z_n, D_n) and colored flows with ocean (m_n, α_n, O_n), $O_n = M_n \setminus D_n$, with initial values of concentrations equal 0 or 1.*

First we formulate the DS theorem as a theorem about the asymptotic behavior of colored flows. Denote also $r_n(i,j) = P(Z_n = i, Z_{n+1} = j) = m_n(i)p_n(i,j)$. In terms of flows, this is the amount of liquid transferred from cup i to cup j at moment n.

To prepare the reader for the general case we first consider the cases of two and three cups. Given a colored flow (m_n, α_n) we can relabel cups at each moment $n \geq k$ in such a way that $\alpha_n(1) \leq \alpha_n(2) \leq \cdots \leq \alpha_n(N)$. Then if $N = 2$ there are only two possibilities: $\lim_{n\to\infty} \alpha_n(1) = \lim_{n\to\infty} \alpha_n(2)$ or $\lim_{n\to\infty} \alpha_n(1) < \lim_{n\to\infty} \alpha_n(2)$. In the first case there is a complete mixing, i.e., concentrations in both cups have the same limit and then the sequences of volumes, $m_n(i)$, $i = 1, 2$, may have no limits. It is easy to prove that in the second case $\lim_{n\to\infty} m_n(i)$ always exist. But this a simple statement. A much less trivial fact (though still relatively simple) is the following statement

Proposition 3. *For the case of $N = 2$, if $\lim_{n\to\infty} \alpha_n(1) < \lim_{n\to\infty} \alpha_n(2)$, then the total amount of liquid transferred between cups 1 and 2 is finite, i.e., $\sum_{n=0}^{\infty}[r_n(1,2)+ r_n(2,1)] < \infty$.*

More than that, if this is true for one colored flow then it is true for any colored flow, i.e., a property of convergence of a sum in Proposition 3 is a property of a sequence (P_n).

Starting from a three cups situation becomes absolutely nontrivial. Let us again relabel the cups at each moment n so that $\alpha_n(1) \leq \alpha_n(2) \leq \alpha_n(3)$. If $\lim_{n\to\infty} \alpha_n(1) = \lim_{n\to\infty} \alpha_n(3)$ then again there is a complete mixing. If $\lim_{n\to\infty} \alpha_n(1) = \alpha_*(1) < \lim_{n\to\infty} \alpha_n(3) = \alpha_*(3)$ it can happen that the concentration in the middle cup may have no limit at all, i.e., $\alpha_n(2)$ will oscillate between $\alpha_*(1)$ and $\alpha_*(3)$. Such a situation is possible only if the volume in this cup tends to zero, $\lim_{n\to\infty} m_n(2) = 0$. The direct total exchange between cups 1 and 3 will be finite as in Proposition 3 but the cup number 2 can actively participate in the exchange between cups 1 and 3. Though its volume tends to zero, the series $\sum_n r_n(2,1)$, $\sum_n r_n(2,3)$ can be infinite. Thus the true analog of Proposition 3 will be a statement about the existence of two jets (J_n^1) and (J_n^2), such that at each moment $n = 0, 1, 2, \ldots$ they form a partition of $M_n = (\{1, 2, 3\}, n)$, $1 \in J_n^1$, $3 \in J_n^2$, and the total exchange of liquid between cups from these two jets is finite.

Note that the set of all such partitions has the power of the continuum and the existence of a partition with the finite exchange property cannot be obtained from the stabilization statement. But such decomposition does take place and it is universal with respect to the initial conditions of the colored flow. This is one of the main points of the DS theorem. The exact formulation is as follows.

Given a flow $m = (m_n)$ or equivalently a MC $Z = (Z_n)$, denote by $V(J^k, J^s \mid m)$ the total amount of liquid transferred between jets J^k and J^s,

$$(10) \qquad V(J^s, J^k | m) = \sum_{n=0}^{\infty} \left[\sum_{i \in J_n^k, j \in J_{n+1}^s} r_n(i,j) + \sum_{i \in J_n^s, j \in J_{n+1}^k} r_n(i,j) \right].$$

Note that if one of these sums is finite then the other is finite also. The equality $r_n(i,j) = P(Z_n = i, Z_{n+1} = j)$ implies also that $V(J^s, J^k \mid m)$ is the expected number of transitions of trajectories of (Z_n) between these two jets.

Theorem 1 (DS theorem, the elementary (deterministic) formulation)**.** *Let a sequence of disjoint sets (M_n) satisfying condition (3) and a sequence of stochastic*

matrices (P_n) be given. Then there exists an integer c, $1 \leq c \leq N$, and a decomposition of the sequence (M_n) into disjoint jets $J^0, J^1, \ldots, J^c$, $J^k = (J_n^k)$, such that for any colored flow (m_n, α_n, O_n),

(**a**) the stabilization of volume and concentration take place inside of any jet J^k, $k = 1, \ldots, c$, i.e., $\lim_{n \to \infty} \sum_{i \in J_n^k} m(i) = m_*^k$; $\lim_{n \to \infty} \alpha(i_n) = \alpha_*^k$, $i_n \in J_n^k$; the concentration in jet J^0 may oscillate; the total volume in this jet tends to zero, i.e., $\lim_{n \to \infty} \sum_{i \in J_n^0} m(i) = 0$;

(**b**) the total amount of liquid transferred between any two different jets is finite on the infinite time interval, i.e., $V(J^k, J^s \mid m) < \infty$, $s \neq k$;

(**c**) this decomposition is unique up to jets (J_n) such that for any flow (m_n) the relation $\lim_n m_n(J_n) = 0$ holds and the total amount of liquid transferred between (J_n) and $(M_n \backslash J_n)$ is finite.

The correspondence between (colored) flows and MCs (Markov pairs) allows us to reformulate the DS theorem as a statement about the behavior of MCs as follows:

Theorem 1 (probabilistic formulation). *Let a sequence of disjoint sets (M_n) satisfying condition (3) and a sequence of stochastic matrices (P_n) be given. Then there exists an integer c, $1 \leq c \leq N$, and a decomposition of the sequence (M_n) into disjoint jets $J^0, J^1, \ldots, J^c$, $J^k = (J_n^k)$, such that*

(**a_1**) *for any Markov chain $Z \in U$ with probability one its trajectory after a finite number of steps enters into one of the jets J^k, $k = 1, \ldots, c$, and stays there forever;*

(**a_2**) *each jet J^k, $k = 1, \ldots, c$, is mixing, i.e., for any two Markov chains $Z^1, Z^2 \in U$ such that $\lim_n P(Z_n^i \in J_n^k) > 0$, $i = 1, 2$, and any sequence of states $i_n \in J_n^k$, $n \in \mathbf{N}$,*

$$(11) \qquad \lim_n \frac{P(Z_n^1 = i_n \mid Z_n^1 \in J_n^k)}{P(Z_n^2 = i_n \mid Z_n^2 \in J_n^k)} = 1;$$

(**b**) *for any Markov chain $Z \in U$ the expected number of transitions of its trajectories between two different jets is finite on the infinite time interval, i.e.,*

$$(12) \qquad \sum_{n=0}^{\infty} [P(Z_n \in J_n^k, Z_{n+1} \notin J_{n+1}^k) + P(Z_n \notin J_n^k, Z_{n+1} \in J_{n+1}^k)] < \infty;$$

(**c**) *this decomposition is unique up to jets (J_n) such that for any Markov chain $Z \in U$ the expected number of transitions of Z between (J_n) and $(M_n \backslash J_n)$ is finite and $\lim_n P(Z_n \in J_n) = 0$.*

Property (b) combined with $\lim_n P(Z_n \in J_n^0) = 0$ implies (a_1), but we prefer to formulate (a_1) and (c) separately. We refer to the points (a_1), (a_2) as the decomposition part and (b) as the separation part.

It was proved in [29] that in the homogeneous case when all stochastic matrices P_n, $n \in \mathbf{N}$, are copies of the same matrix P, the above decomposition is nothing else than the space-time representation of the decomposition of M into ergodic classes and cyclic subclasses, where each subclass is represented by a sequence J^k, $k \neq 0$. Thus the DS theorem is a direct generalization of the Kolmogorov–Doeblin results.

3. Key elements of the proof of DS theorem

Given a MC $Z = (Z_n)$ a jet $J = (J_n)$ is called a *trap* if the event that "Z visits J infinitely often" coincides almost surely with the event that "Z stays in J forever," i.e., if $P(\limsup(Z_n \in J_n)) = P(\liminf(Z_n \in J_n))$. If $J = (J_n)$ is a trap then it is

easy to check that $\lim_n P(Z_n \in J)$ does exists and coincides with these limits. We denote this limit as $v(Z, J)$, the "volume" of J for Z.

Given a MC $Z = (Z_n)$ a jet $J = (J_n)$ is called a *strap* (strong trap) if the *expected number* of transitions of Z between (J_n) and its complement $(M_n \setminus J_n)$ is finite, i.e., a sum similar to (12) is finite. Obviously, each strap is a trap but not vice versa, because it is possible that a (random) number of exits from a given jet is finite with probability one, but the expected value is infinite.

In the language of flows, a jet (J_n) is a strap for a flow $m = (m_n)$ if the total "overflow" from jet J to all other jets is finite, i.e., a sum similar to (10) is finite.

Given a MC $Z = (Z_n)$ a strap (J_n) is called *indecomposable* if it cannot be partitioned into two straps (S_n) and (K_n) of positive volume.

The decomposition into straps $J^0, J^1, \ldots, J^c$ described in the DS theorem has two key features: First, it is universal, i.e., the same decomposition for all MCs in U, second, for any MC (any colored flow) there is a mixing inside of every jet J^k, $k > 0$. The first feature can be obtained if we consider a "universal" MC Z_*, i.e., a MC which coincides with positive probability with any MC from U, and prove the existence of decomposition for this MC. The construction of Z_* can be easily done, see details in [29]. The decomposition into indecomposable straps for this MC exists almost by definition. If (M_n) is indecomposable then $c = 1$ and there is only one jet. If (M_n) is decomposable then there are two straps of positive volume and if each of them is indecomposable then $c = 2$ and we obtained a decomposition. Otherwise we can continue this process and since every jet with positive volume for large n contains at least one point then in no more than N steps we obtain a decomposition into indecomposable jets. The only remaining question is: Why for any MC (colored flow) there is a mixing inside of an indecomposable jet? To answer this question we need to relate some martingales to a given colored flow (Markov pair (Z, D)) and to show that due to condition (3) they will have very specific properties.

Given two sequences (a_n) and (b_n) we say that they intersect at moment k if $a_k \leq b_k, a_{k+1} > b_{k+1}$, or $a_k > b_k, a_{k+1} \leq b_{k+1}$. Given a real valued r.s. $X = (X_n)$ and a nonrandom sequence $d = (d_n)$, we denote $R_T(X \mid d)$ the expected number of intersections of trajectories of X with (d_n) on the interval $(0, T)$,

$$(13) \quad R_T(X \mid d) = \sum_{n=0}^{T-1} [P(X_n \leq d_n, X_{n+1} > d_{n+1}) + P(X_n > d_n, X_{n+1} \leq d_{n+1})].$$

A nonrandom sequence (d_n) is called a *barrier* for the r.s. $X = (X_n)$ if the expected number of intersections of (d_n) by the trajectories of X on the infinite time interval is finite, i.e., $\lim_T R_T(X \mid d) < \infty$. If additionally $d_n = d$ for all n, we call (d_n) a *level barrier*.

To prove the existence of barriers and relate them to the separation part of the DS theorem we introduce a r.s. (Y_n) as follows. Suppose a colored flow (m_n, α_n), or equivalently a Markov pair (Z_n, D_n) is given, where $\alpha_n(i)$ are as in (5). Then define

$$(14) \quad\quad\quad\quad\quad\quad Y_n = \alpha_n(Z_n), \quad n \in \mathbf{N},$$

Lemma 1. *A random sequence (Y_n) specified by (14) is a submartingale in reverse time.*

To see why this lemma is true it is sufficient to notice that if we denote $q_n(j, i) = m_n(i)p_n(i, j)/m_{n+1}((j)$, the transition probabilities of a MC (Z_n) in inverse time,

then formula (5) will represent the condition of a r.s. (Y_n) to be a martingale in reverse time. Since earlier we introduced colored flows with an "ocean," where $\alpha_n(i)$ are not calculated by formula (5) but are defined as $\alpha_n(i) = 0$ for i from the "ocean," sometimes the equality in (5) is replaced by an inequality representing the submartingale property.

This simple lemma is a bridge between the DS theorem and Theorem 2 about the existence of barriers. This theorem implies that r.s. (Y_n) has barriers inside of any interval (a, b). We will discuss Theorem 2 in the next section. Note also that the reversed martingales are used extensively e.g. in [34] and [35].

Proposition 4. *Let (Z, D) be a Markov pair and (Y_n) be a corresponding submartingale, i.e., $Y_n = \alpha_n(Z_n)$. Then a sequence (d_n) is a barrier for (Y_n) iff a jet (J_n), $J_n = \{i \in M_n : \alpha_n(i) \le d_n\}$, is a strap for (Z_n).*

The validity of Proposition 4 follows from the definition of barriers and straps.

Now we are able to explain heuristically why the existence of barriers is equivalent to the mixing property inside of any indecomposable strap $J = (J_n)$. Suppose that there is a colored flow such that there are two disjoint jets (S_n) and (K_n) such that $\liminf_n \alpha_n(i_n) \ge b > a \ge \limsup_n \alpha_n(j_n)$ for $i_n \in S_n$ and $j_n \in K_n$, and $\lim_n \sum_{i \in S_n} m_n(i) > 0$, $\lim_n \sum_{j \in K_n} m_n(j) > 0$. Then r.s. (Y_n) will have values above b and below a with positive probability. By Theorem 2 from the next section there is a barrier for (Y_n) inside of an interval (a, b) and therefore the indecomposable strap (J_n) can be decomposed into two straps of positive volume, a contradiction.

4. Doob's lemma and existence of barriers

Generally barriers or level-barriers may not exist for a given r.s. (X_n). The closest statement about intersections of a level or an interval is the well known Doob's upcrossing lemma (Doob's inequality). This lemma implies one of the central results in the theory of stochastic processes—a theorem of Doob about the existence of the limits of trajectories of a (sub)(super)martingale when time tends to infinity.

We recall that given two numbers a and b, with $a < b$ and a sequence (x_n), the *number of upcrossings* of an interval (a, b) by (x_n) on the interval $(0, T)$ is the maximal number of disjoint intervals $(n_i, n_{i+1}) \subset (0, T)$ such that $x_{n_i} \le a$ and $x_{n_{i+1}} \ge b$. Given a random sequence (r.s.) $X = (X_n)$ we denote the expected number of upcrossings of (a, b) by trajectories of X as $U_T(X, (a, b))$. Note that when an interval (a, b) is replaced by a sequence (d_n), then the notion of an (up)(down)crossing transforms naturally into a notion of *intersection*. We have obviously $U_T(X, (a, b)) \le R_T(X, d)$ for any $(d_n) \subset (a, b)$.

Doob's lemma. *Let $X = (X_n)$ be a (sub)martingale. Then, for every T,*

$$(15) \qquad U_T(X, (a, b)) \le \frac{E(X_T - a)^+}{b - a}.$$

In particular, if $\sup_n EX_n^+ < \infty$, then Doob's lemma implies that the expected number of upcrossings of every fixed interval (a, b) on the infinite time interval, $\lim_T U_T(X, (a, b))$, is *finite*. The previously mentioned Doob's theorem follows immediately. The inequalities similar to (15) hold for downcrossings and crossings, for the supermartingales and for sub(super)martingales in reversed time. We call all such r.s. a *martingale type* r.s. and we denote the class of all bounded

(sub)(super)martingales in forward or inverse time by $\mathcal{M}$. For simplicity we will further consider only bounded r.s. $0 \leq X_n \leq 1$ for all n.

The width of the interval $(b-a)$ is in the denominator of the estimate (15) so Doob's lemma *does not imply* that: **1)** inside of the interval (a, b) there exists a *level c* such that the expected number of intersections of this level is finite, or, a weaker statement, that **2)** there exists a *nonrandom sequence* $d = (d_n)$ with a similar property.

In [31] an example shows that not only the level barriers but even the barriers may not exist inside of a given interval if a bounded martingale (X_n) takes a countable number of values. At the same time if (X_n) at each moment n takes only a bounded number of values, i.e. if there exists a sequence of finite sets (G_n) such that $P(X_n \in G_n) = 1$ for each n and $|G_n| \leq N < \infty$, $n \in \mathbf{N}$, then Doob's lemma can be substantially strengthened. Denote the class of all such r.s. by $\mathcal{G}^N$. The following theorem holds.

Theorem 2. *Let a and b be two numbers, $a < b$, and $X = (X_n) \in \mathcal{M} \cap \mathcal{G}^N$. Then inside of the interval (a, b) there exists a barrier $d = (d_n)$.*

This theorem follows from a more general theorem in [26] about the existence of barriers for processes with finite variation and a bounded number of values.

Let $\varphi(s)$, $s \geq 0$, be a nondecreasing function, $\varphi(0) = 0$. Denote by $V_T^\varphi(X)$ the φ-variation of X on the time interval $(0, T)$, defined as

$$(16) \qquad V_T^\varphi(X) = \sup_{0 \leq n_1 < n_2 < \cdots < n_k \leq T} \sum_{i=1}^{k-1} E\varphi(|X_{n_{i+1}} - X_{n_i}|),$$

where the sup is taken over all possible partitions $(n_1 < n_2 < \cdots < n_k)$, $k = 1, 2, \ldots$, of the interval $(0, T)$. Let us denote by $\mathcal{F}^\varphi$ the class of all r.s. with finite variation, i.e., with $\lim_T V_T^\varphi(X) < \infty$. Using notation similar to the previously defined $\mathcal{G}^N$, let us denote by $\mathcal{G}^N(a, b)$ the class of all r.s. which take no more than N values inside of an interval (a, b).

Theorem 3. *If $(X_n) \in \mathcal{G}^N(a, b) \cap \mathcal{F}^\varphi$ then there is a number $h > 0$ and a sequence of intervals (Δ_n), such that $(\Delta_n) \in (a, b)$, $|\Delta| \geq h$, and any sequence (d_n), $d_n \in (\Delta_n)$, is a barrier for (X_n).*

Theorem 2 immediately follows from Theorem 3 because it is well known that any r.s. $X = (X_n) \in \mathcal{M}$ belongs to $\mathcal{F}^\varphi$ for $\varphi(x) = x^2$.

We will sketch the proof of Theorem 3 in the next section.

The statement of Theorem 2 left an open question, whether under its assumptions inside of any interval (a, b) there exist *level barriers*, i.e., nonrandom sequences (d_n) with constant values of $d_n = d$. It happens that the level barriers exist only if $N = 2, 3$ and for $N \geq 4$ the negative answer is given by the following theorem. For a r.s. $X = (X_n)$, denote by $N(X, d)$ the function counting the expected number of intersections of level d by r.s. (X_n) on an infinite time interval.

Theorem 4 (joint with Alexander Gordon). *There is a nonhomogeneous MC $X = (X_n)$ that is also a martingale such that*
a) X takes no more than four values $\{a_n, r_n, s_n, 1\}$ at each moment $n = 1, 2, \ldots$, $0 < a_n < r_n < s_n \leq 1$, and
b) $N(X, d) = \infty$ for each $d \in (0, 1)$.

Such a MC martingale is constructed explicitly but has a rather complicated structure and this example is not published yet. Theorem 4 of course does not

contradict Doob's theorem. In this example every trajectory will tend to 0 or 1 and will have only a finite number of intersections with any level d, but nevertheless the expected number of intersections for every d is infinite.

5. Sketch of the proof of Theorem 3

Theorem 3 follows easily from the following estimate.

Lemma 2 (basic estimate). *Let (a, b) be an interval, $G = (G_n)$ be a sequence of sets, $|G_n \cap (a, b)| \le N$. Then there is a sequence of intervals (Δ_n), constructed recursively, $\Delta_{n+1} = f(\Delta_n, G_{n+1})$, such that*

$$(17) \qquad (A) \qquad \Delta_n \subset (a, b), \ |\Delta_n| \ge h > 0, \ \Delta_n \cap G_n = \emptyset, \ n = 1, 2, \dots$$

and for any sequence $d = (d_n)$, $d_n \in \Delta_n$, any $X = (X_n) \in \mathcal{G}^N(a, b)$, any nondecreasing function φ, and any T,

$$(18) \qquad\qquad R_T(X, d) \le c V_T^\varphi(X),$$

for a constant $c = c(\varphi, N, b - a)$.

Lemma 2 immediately implies that barriers do exist for any r.s. $X = (X_n)$ from $\mathcal{G}^N(a, b)$ for which there is a φ such that $\lim_T V_T^\varphi(X) < \infty$.

Note that the sequence of intervals (Δ_n) in Lemma 2 has a universal character, i.e., it depends only on a structure of a deterministic sequence of sets (G_n) and that function $f(\Delta, G)$ in Lemma 2 is constructed in an explicit form.

Now we explain heuristically what kind of sequence of intervals (Δ_n) we need to construct to ensure that (18) will hold for any d, X, φ, and T.

Suppose initially that given a sequence (G_n), there is a sequence of intervals (Δ_n) that satisfy, in addition to (A), the condition

$$(19) \qquad (B) \qquad |\Delta_n \cap \Delta_{n+1}| \ge h > 0, \quad n = 1, 2, \dots.$$

The following elementary lemma holds.

Lemma 3. *Let Δ_1 and Δ_2 be intervals, X_1, X_2 be random variables such that $|\Delta_1 \cap \Delta_2| \ge h > 0$, and $P(X_i \in \Delta_i) = 0$, $i = 1, 2$. Then for any numbers $d_1 \in \Delta_1$, $d_2 \in \Delta_2$, and nondecreasing function φ,*

$$(20) \qquad\qquad P(X_1 \le d_1, X_2 \ge d_2) \le E\varphi(|X_1 - X_2|)/\varphi(h).$$

The assertion of Lemma 3 follows immediately from the implications $(X_1 \le d_1, X_2 \ge d_2) \subset (|X_1 - X_2| \ge h)$ and Chebyshev's inequality, $P(|Y| \ge h) \le E\varphi(|Y|)/\varphi(h)$ for any r.v. Y.

The existence of a sequence (Δ_n) satisfying the conditions (A) and (B) implies an estimate (18). Indeed, in this case by Lemma 3 for any sequence (d_n), $d_n \in \Delta_n$, we have $P(X_n \le d_n, X_{n+1} > d_{n+1}) \le E\varphi(|X_{n+1} - X_n|)/\varphi(h)$ and therefore $R_T(X, d) \le \sum_{n=0}^{T-1} E\varphi(|X_{n+1} - X_n|)/\varphi(h) \le c V_T^\varphi(X, d)$ with $c = 1/\varphi(h)$.

Since $|G_n \cap (a, b)| \le N$ there are of course sequences (Δ_n) satisfying (A) for every h, $0 < h \le 1/(N + 1)$, but generally we cannot expect that for such (G_n) there is a sequence of intervals that satisfy both (A) and (B). We will show that an estimate (18) still holds if condition (B) is replaced by a weaker condition (C).

(C) (a) For any n there is $r(n)$, $1 \le r(n) \le n$, such that $|\Delta_{r(n)} \cap \Delta_n| \ge h > 0$, $|\Delta_{r(n)} \cap \Delta_{n+1}| \ge h$,

(b) every n is covered by no more than M intervals of the form $[r(k), k]$.

First we formulate Lemma 4 which is an analog of Lemma 2 when the condition (B) is replaced by condition (C).

Lemma 4. *Let Δ_1, Δ_2, and Δ_3 be intervals and Y_1, Y_2, and Y_3 be random variables such that $|\Delta_1 \cap \Delta_2| \geq h > 0$, $|\Delta_1 \cap \Delta_3| \geq h$, and $P(Y_i \in \Delta_i) = 0$, $i = 1, 2, 3$. Then for any numbers $d_2 \in \Delta_2$, $d_3 \in \Delta_3$, and nondecreasing function φ,*

$$(21) \qquad P(Y_2 \leq d_2, Y_3 \geq d_3) \leq E[\varphi(|Y_1 - Y_2|) + \varphi(|Y_1 - Y_3|)]/\varphi(h).$$

Note that for any $d_1 \in \Delta_1$ we have a trivial inequality $P(Y_2 \leq d_2, Y_3 \geq d_3) \leq P(Y_1 \geq d_1, Y_2 \leq d_2) + P(Y_1 \leq d_1, Y_3 \geq d_3)$. Now the assertion of Lemma 4 can be easily obtained if we apply Lemma 2 to each of the terms in the right side of the last inequality using pairs (Y_2, Y_1) and (Y_1, Y_3). Then using (21) for $Y_1 = X_{r(n)}$, $Y_2 = X_n$, $Y_3 = X_{n+1}$, we obtain $P(X_n \leq d_n, X_{n+1} \geq d_{n+1}) \leq E[\varphi(|X_{r(n)} - X_n|) + \varphi(|X_{r(n)} - X_{n+1}|)]/\varphi(h)$. Summing this inequality over all n, and taking into account point (b) of condition (C), we obtain (18) with $c = M/\varphi(h)$.

Thus to prove Theorem 2 it remains to prove the pure combinatorial lemma

Lemma 5. *Let (a, b) be an interval, $G = (G_n)$ be a sequence of sets, $|G_n \cap (a, b)| \leq N$. Then there is a sequence of intervals (Δ_n), calculated by a recursive formula $\Delta_{n+1} = f(\Delta_n, G_{n+1})$ satisfying conditions (A) and (C) for some $h > 0$.*

The formal proof is rather complicated but the idea of the construction can be explained using cases $N = 1, 2$. Note that the statement of Lemma 2 is not quite trivial even in the case of $N = 1$.

Without loss of generality we can assume that $|G_n| = N$, $n = 1, 2, \ldots$ and $(a, b) = (0, 1)$. Initially we will construct a sequence of intervals (s_n), which will serve as a "frame" for intervals (Δ_n), $s_n \subset \Delta_n$, $n = 1, 2, \ldots$.

The case $N = 1$. Let us divide the interval $(0, 1)$ into three equal intervals and let us denote $(0, 1/3) = (0)$ and $(2/3, 1) = (1)$. To explain our construction we can use the following informal interpretation. There is a "hunter" and a "game" which tries to avoid the hunter hiding in one of the intervals $(s_n) \in \{(0), (1)\}$. The position of the hunter at moment n is one element set G_n. The game knows the position of the hunter so it always can avoid the hunter but its goal is to spend the minimal amount of "energy" switching from one of the possible hiding locations to the other. We define the sequences of intervals (s_n) and (Δ_n) as follows: $s_{n+1} = s_n$ if $s_n \cap G_{n+1} = \emptyset$, otherwise $s_{n+1} = (0)$ if $s_n = (1)$ and $s_{n+1} = (1)$ if $s_n = (0)$. The intervals Δ_n are defined as $\Delta_{n+1} = s_{n+1}$ if $s_{n+1} = s_n$ and $\Delta_{n+1} = (0, 1) \setminus s_n$, otherwise, $n = 1, 2, \ldots$. Thus, $\Delta_{n+1} = f(s_n, G_{n+1})$, where $f(s, G) = s$ if $s \cap G = \emptyset$, and $f(s, G) = (0, 1) \setminus s$ otherwise.

For $N = 2$ we divide the vertical interval $(0, 1)$ into three equal intervals and each of the intervals $(0, 1/3) = (0)$ and $(2/3, 1) = (1)$ we again divide into three equal intervals and denote the lower and upper of the corresponding intervals as (00), (01), (10), and (11). Now there are two hunters with positions at two points from (G_n), and a game who can use any of these four intervals for hiding. The minimizing energy strategy of a game is now as follows. First, $s_{n+1} = s_n$ if $s_n \cap G_{n+1} = \emptyset$, otherwise if s_n is e.g. in (0) and only one hunter is in (0), then game avoids the second hunter changing s_n inside of (0), i.e., from (00) to (01) and back. Only when the second hunter is also in (0), is the hiding position in (1), i.e., $s_n = (10)$ or (11). A sequence (Δ_n) is defined by $\Delta_{n+1} = s_{n+1}$ if $s_{n+1} = s_n$, and otherwise as follow: $\Delta_{n+1} = (0) \setminus s_n$ if a switching occurs inside of (0), $\Delta_{n+1} = (1) \setminus s_n$ if a

switching occurs inside of (1), and e.g. $\Delta_{n+1} = (0,1)\backslash(0)$ if a switching occurs from an interval in (0) to an interval in (1), $n > 1$.

For general $N \geq 2$ our construction proceeds as follows. An interval $(0,1)$ is divided as in the first N steps of Cantor's construction of a perfect set, i.e., at each step a middle interval of the three equal intervals is eliminated. The process of switching the form of the hiding intervals to the other depends on how many hunters are close to the game.

6. Open Problems

1. The DS theorem is an *existence* theorem. The value c and the structure of the decomposition depends naturally on the structure and assumption about the sequence (P_n). Most of the literature on nonhomogeneous MCs in general and on simulated annealing is in fact a study of such decompositions, without paying any attention to the difference between traps and straps (see as an example an interesting paper on simulated annealing [5]). At the same time such a distinction plays a very important practical role. The statement that some algorithm or computational procedure converges with probability one leaves a question whether such convergence on average requires a finite or infinite time. Thus a general open problem is to describe necessary and sufficient conditions on a certain structure of decomposition. In fact, in many papers, starting from the pioneering paper of Kolmogorov [18] the conditions for complete mixing, i.e., for $c = 1$ are well established.

2. The idea of using stochastic and especially doubly stochastic matrices for the description of ordering in the space of finite-dimensional vectors is the key idea of the so-called *theory of majorization*. We refer the reader to the monograph of Marshall and Olkin [21] for the theory of majorization and to Sonin [27], where the relation between the DS theorem and majorization theory is briefly described. There is an unpublished (in English) paper of the author about the economic interpretation of the DS theorem. The idea of using the theory of majorization in the description of irreversible physical processes was elaborated in some papers following the pioneering work of Ruch ([24]), see also [1]. The possible analog of the DS theorem on general irreversible processes should replace formula (6) by a general transformation of a system.

3. The analog of the DS theorem for the countable case. The main result of [31] is the following:

Theorem 5. *There exist a sequence of finite sets (M_n), $|M_n| \to \infty$, a sequence of stochastic matrices (P_n) indexed by (M_n), a Markov chain (Z_n), and a sequence of sets (D_n), $D_n \subseteq M_n$, $n \in \mathbf{N}$, such that the submartingale (in a reversed time) (Y_n) specified by (14) has no barriers inside of some interval (a,b).*

Note that while the above statement shows that the DS theorem is not true in the form presented in Section 2, it is nevertheless possible that its analog may exist in the countable case if the expected *number* of *intersections* is replaced by other characteristics of the transitions of trajectories. The analog of the DS theorem for the countable case would give a possibility to consider the case of continuous space. Even the finite case can serve as a basis for the generalization of the DS theorem to continuous time.

Acknowledgments

References

[1] ALBERTI, P. and UHLMANN, A. (1982). *Stochasticity and Partial Order: Doubly stochastic maps and unitary mixing.* Mathematics and Its Applications **9**. D. Reidel Publishing Co., Dordrecht, Boston.

[2] BLACKWELL, D. (1945). Finite nonhomogeneous Markov chains. *Ann. Math.* **46** 594–599.

[3] COHN, H. (1970). On the tail σ-algebra of the finite Markov chains. *Ann. Math. Statist.* **41** 2175–2176.

[4] COHN, H. (1976). Finite nonhomogeneous Markov chains: Asymptotic behavior. *Adv. Appl. Prob.* **8** 502–516.

[5] COHN, H. and FIELDING, M. (1999). Simulated annealing: Searching for an optimal temperature schedule. *SIAM J. Optim.* **9** (3) 779–802.

[6] COHN, H. (1989). Products of stochastic matrices and applications. *Int. J. Math. Sci.* **12** 209–333.

[7] DIETZ, Z. and SETHURAMAN, S. (2005). Large deviations for a class of nonhomogeneous Markov chains. *Ann. Appl. Probab.* **15** (1A) 421–486.

[8] DOBRUSHIN, R. (1956). Central limit theorem for non-stationary Markov chains. I. *Theory Probab. Appl.* **1** (1) 65–80.

[9] ETHIER, S. and KURTZ, T. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[10] GRIFFEATH, D. (1975). Uniform coupling of non-homogeneous Markov chains. *J. Appl. Probability* **12** (4) 753–762.

[11] HAJNAL, J. (1956). The ergodic properties of non-homogeneous finite Markov chains. *Proc. Cambridge Philos. Soc.* **52** 67–77.

[12] HARTFIEL, D. (2002). *Nonhomogeneous Matrix Products.* World Scientific Publishing Co., Inc., River Edge, NJ.

[13] HILL, T. (1979). On the existence of good Markov strategies. *Trans. Am. Math. Soc.* **247** 157-176.

[14] IOSIFESCU, M. (1966). On the uniform ergodicity of a class of nonhomogeneous random systems with complete connections. *Rev. Roumaine Math. Pures Appl.* **11** 763–772.

[15] ISAACSON, D. and MADSEN, R. (1976). *Markov Chains: Theory and Applications.* Wiley, New York.

[16] KEMENY, J. and SNELL, J. (1976). *Finite Markov chains.* Reprinting of the 1960 original. Springer-Verlag, New York, Heidelberg.

[17] KINGMAN, J. F. C. (1975). Geometrical aspects of the theory of nonhomogeneous Markov chains. *Math. Proc. Cambridge Philos. Soc.* **77** 171–183.

[18] KOLMOGOROFF, A. N. (1936). Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112** 155–160. *Selected Works of A. N. Kolmogorov* **2** (A. N. Shiryaev, ed.). Probability Theory and Math. Statistics. Kluwer Acad. Publ.

[19] LIU, W. and YANG, W. (1996). An extension of Shannon–McMillan theorem and some limit properties for nonhomogeneous Markov chains. *Stoch. Process. Appl.* **61** (1) 129–145.

[20] MAKSIMOV, V. M. (1970). The convergence of nonhomogeneous doubly stochastic Markov chains. *Theory Probab. Appl.* **15** 622–636.

[21] MARSHALL, A. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

[22] MUKHERJEA, A. (1983). Nonhomogeneous Markov chains: Tail idempotents, tail sigma-fields and basis. *Math. Z.* **183** (3) 293–309.

[23] NIEMIRO, W. and POKAROWSKI, P. (1995). Tail events of some nonhomogeneous Markov chains. *Ann. Appl. Probab.* **5** (1) 261–293.

[24] RUCH, E. and MEAD, A. (1976). The principle of increasing mixing character and some of its consequences. *Journ. Theor. Chem. A.* **41** (2) 95–117.

[25] SENETA, E. (1973). On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Cambridge Philos. Soc.* **74** 507–513.

[26] SONIN, I. (1987). Theorem on separation of jets and some properties of random sequences. *Stochastics* **21** 231–250.

[27] SONIN, I. M. (1988). The separation of jets and some asymptotic properties of random sequences. Discrete Event Systems: Models and Applications. *Lecture Notes in Contr. Inf.* **103** 275–282. Springer.

[28] SONIN, I. M. (1991a). On an extremal property of Markov chains and sufficiency of Markov strategies in Markov decision processes with the Dubins–Savage criterion. *Ann. of Oper. Res.* **29** 417–426.

[29] SONIN, I. (1991b). An arbitrary nonhomogeneous Markov chain with bounded number of states may be decomposed into asymptotically noncommunicating components having the mixing property. *Theory Probab. Appl.* **36** 74–85.

[30] SONIN, I. (1996). The asymptotic behaviour of a general finite nonhomogeneous Markov chain (the decomposition-separation theorem). In *Statistics, Probability and Game Theory, papers in Honor of David Blackwell* (T. S. Ferguson, L. S. Shapley and J. B. MacQueen, eds.) Lecture Notes-Monograph Series, **30** 337–346. Inst. of Math. Stat.

[31] SONIN, I. (1997). On some asymptotic properties of nonhomogeneous Markov chains and random sequences with countable number of values. In *Stat. and Control of Stoch. Proc., The Liptser Festschrift, Proceedings of Steklov Math. Inst. Seminar* (Y. Kabanov, B. Rozovskii, A. Shiryaev, eds.) 297–313. World Sci. Publ., River Edge, NJ.

[32] SETHURAMAN, S. and VARADHAN, S. R. S. (2005). A martingale proof of Dobrushin's theorem for non-homogeneous Markov chains. *Electron. J. Probab.* **10** (36) 1221–1235.

[33] SHIRYAEV, A. N. (1996). *Probability.* Springer-Verlag, New York.

[34] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration.* Probability and its Applications. Springer-Verlag, New York.

[35] VERSHIK, A. and KACHUROVSKI, A. (1999). Rates of convergence in ergodic theorems for locally finite groups, and reversed martingales. *Differ. Equat. and Contr. Proc.* (1) 19–26.