Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

THE ASYMPTOTIC BEHAVIOUR OF A GENERAL FINITE NONHOMOGENEOUS MARKOV CHAIN (THE DECOMPOSITION-SEPARATION THEOREM)[1]

ISAAC SONIN
University of North Carolina at Charlotte

Abstract

The Decomposition-Separation Theorem generalizing the classical Kolmogorov-Doeblin results about the decomposition of finite homogeneous Markov chains to the nonhomogeneous case is presented. The ground-breaking result in this direction was given in the work of David Blackwell in 1945. The relation of this theorem with other problems in probability theory and Markov Decison Processes is discussed.

Dedicated to David Blackwell in deep respect for his many wonderful mathematical achievements.

1. Introduction. Let M be a finite set, $P = \{p(i,j)\}$ be a stochastic matrix, $i, j \in M$, U_0 be the family of all (homogeneous) Markov chains (MC) $X = (X_n), n \in \mathbf{N} = \{0, 1, \ldots\}$, specified by M and P and all possible initial distributions μ. The classical Kolmogorov-Doeblin results describing the asymptotic behavior of MC from U_0 can be found in most advanced books on probability theory as well as the monographs on MC (see for example Kemeny and Snell (1960), Isaacson and Madsen (1976), Shiryayev (1984)).

According to these results the state space M can be decomposed into the set of nonessential states and the classes of essential communicating states. Furthermore, the following are true:

(**A**) With probability one, each trajectory of a MC X from U_0 will reach one of these classes and never leave it.

Each class S can be decomposed into cyclical subclasses. If the number of subclasses is equal to one (an aperiodic class), then

(**B**) every MC X from U_0 has a mixing property inside such a class, i.e. there exists a limit distribution π

$$\lim_n P(X_n = x | X_n \in S) = \pi(x) > 0, x \in S, \tag{1}$$

which does not depend on the initial distribution μ and such that π is invariant with respect to the matrix P.

[1] *Key words:* Finite nonhomogeneous Markov chain, submartingale, irreversible process. AMS 1991 Subject classification. Primary: 60J10.

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 those n which are comparable modulo the number of cyclical subclasses.

Let us now assume that instead of a stochastic matrix P we have a *sequence* (P_n) of stochastic matrices, $P_n = \{p_n(i,j)\}$ and let U be the corresponding family of all *nonhomogeneous* Markov chains. What can be said about the behavior of MC from U? At first glance the natural answer is "Nothing can be said until some assumptions on the sequence (P_n) are made".

But, though it may seem surprising, there is a theorem describing the asymptotic behavior of Markov chains in U *without any assumptions* on the sequence of stochastic matrices P_n. (The situation in fact is similar to the homogeneous case where we only assume that P is a stochastic matrix).

Briefly, this theorem states that a decomposition with the properties similar to (A), (B) 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 $(M_n) = M \times \mathbf{N}$.

Note that this space-time decomposition and the corresponding formulation of this theorem is not a notational convenience but is the heart of the problem. In the general nonhomogeneous case without any specific assumptions about the structure of (P_n), the label of a state is in a sense meaningless without the reference to time. To stress this point and to use more compact and unified notations, we will assume that there is no fixed state set M at all and that we are given a sequence (M_n) of countable disjoint sets and that (P_n) is a sequence of stochastic matrices indexed by the elements of these sets, i.e. $P_n = \{p_n(i,j)\}, i \in M_n, j \in M_{n+1}, n \in \mathbf{N}$. Denote by U the family of all *nonhomogeneous* Markov chains, referred to below simply as Markov chains, $Z = (Z_n)$, $Z_n \in M_n$, $n \in \mathbf{N}$, specified by these two sequences and all possible "initial" distributions μ defined on all $M_k, k \in \mathbf{N}$. The assumption that M is finite is now replaced by the assumption

$$|M_n| \leq N < \infty, n \in \mathbf{N}. \tag{2}$$

2. The Decomposition-Separation Theorem. Formulation.

Theorem 1. *Let a sequence of disjoint sets (M_n), satisfying condition (2) and a sequence of stochastic matrices (P_n) be given. Then there exists a decomposition of the sequence (M_n) into disjoint sequences $J^0, J^1, \ldots, J^c, 1 \leq c \leq N, J^k = (J_n^k), J_n^k \cap J_n^s = \emptyset, k \neq s, \bigcup_k J_n^k = M_n, n \in \mathbf{N}$ such that*

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

(**b**) *each sequence* $J^k, k = 1, \ldots, c$ *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* $i_n \in J_n^k, n \in \mathbf{N}$

$$\lim_n \frac{P(Z_n^1 = i_n | Z_n^1 \in J_n^k)}{P(Z_n^2 = i_n | Z_n^2 \in J_n^k)} = 1; \tag{3}$$

(**c**) *the expected number of transitions of trajectories for any Markov chain* $Z \in U$ *between any sequence* J^k *and its complement is finite on the infinite time interval, i.e.*

$$\sum_{n=0}^{\infty} [P(Z_n \in J_n^k, Z_{n+1} \notin J_{n+1}^k) + P(Z_n \notin J_n^k, Z_{n+1} \in J_{n+1}^k)] < \infty; \tag{4}$$

and

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

Property (c) combined with $\lim_n P(Z_n \in J_n^0) = 0$ implies (a), but we prefer to formulate (a) and (c) separately.

We call this theorem the Decomposition-Separation (DS) theorem, referring to the points (a), (b) as the decomposition part and (c) as the separation part.

It can be proved, 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. Brief History. The formulation and the proof of the DS theorem are associated with the names of A. Kolmogorov, D. Blackwell, H. Cohn and the author of this paper.

The starting point for the whole topic was a small paper of Kolmogorov (1936), who asked and answered the following question. When, given a set M and a sequence (P_n), defined for all $n = \ldots, -1, 0, 1, \ldots$, is there a unique corresponding MC specified for all such n? The answer is that this is true if and only if the limits

$$\lim_{n \to -\infty} P(Z_m = j | Z_n = i) = (\prod_{k=n}^{m-1} P_k)(i, j)$$

exist for all m and j and does not depend on i.

The ground-breaking step was made in 1945 by David Blackwell who proved that for any sequence (P_n) there is a decomposition of (M_n) into sequences $(T_n^0), (T_n^1), \dots, (T_n^c)$ with properties (a) and (b) of Theorem 1. The decisive point of his proof was the use of a then relatively new result of Doob about the existence of the limit for almost all trajectories of bounded (sub)martingales and an elaborate construction of $(T_n^k), k = 0, 1, \dots, c$ to eliminate the states where the limits in (3) do not exist. As Kolmogorov did, Blackwell considered MC in reverse time.

The next step was made in the works of Harry Cohn (1970), (1976) and other of his papers, (see his expository paper, Cohn (1989)), who considered the 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 the decomposition, $k = 1, \dots, c$. He also simplified Blackwell's proof, though it was still very complicated. Note also that the papers of Cohn contain many other results for the finite and countable cases when some additional assumptions about the structure of (P_n) are made. Briefly, the Blackwell-Cohn results can be described as the DS theorem without property (c), i.e. the decomposition part. Such decomposition lacks a transparent physical interpretation and this probably is one of the reasons why the work of Blackwell (1945) and its generalization by Cohn are not referenced in monographs on stochastic processes or probability theory despite its general character.

The last step in the proof of the DS theorem was made by the author in a series of papers Sonin (1987, 1988, 1991a, 1991b), where it was proved that among the Blackwell-Cohn decompositions there are decompositions into sequences having the additional property (c). These sequences for a particular Markov chain were called *traps* and for the family of Markov chains correspondingly *universal traps*. The property (c) and the existence of universal traps were not obvious and they were not noted or mentioned before. The list of problems that have led the author to the formulation of point (c) is as follows: the problem of sufficiency of Markov strategies for the Dubins-Savage functional; the equivalent random sequences and Feinberg inequality; the deterministic model of the family of MC (colored flows); and Doob's upcrossing lemma and its strengthening to the case of bounded (sub)martingales which take on only a bounded number of values. The last result, published in Sonin (1987) plays a crucial role in the proof of point (c), and we discuss it in the next section.

Note also that there exists a substantial body of literature on nonhomogeneous Markov chains with some special assumptions on the transition matrices (P_n). (See the works of R. Dobrushin, D. Griffeath, J. Hainal, D. Isaacson, M. Iosifescu, J. Kingman, R. Madsen, V. Maksimov, A. Mukherjea, E. Seneta and others who have contributed in this area). The study of backwards limits in the context of nonhomogeneous regenerative processes was continued by H. Thorisson (see Thorisson (1988) and his other works).

4. Doob's Lemma and Its Modification. One of the most remarkable and widely used results in the theory of stochastic processes is the theorem of Doob about the existence of the limits of trajectories of bounded (sub)martingale when time tends to infinity. In particular Doob's theorem played an important role in Blackwell's paper. This theorem is based on Doob's upcrossing lemma.

Doob's Lemma. *If $X = (X_n)$ is a bounded (sub)martingale then the expected number of intersections of every fixed interval (a, b) by the trajectories of X is finite on the infinite time interval.*

The width of the interval $(b-a)$ is in the denominator of the corresponding estimate so Doob's lemma does not imply for example that inside the interval there exists a *level* such that the expected number of intersections of this level is finite. (Sonin (1994) gave an example to show that this is not true in countable case).

If (X_n) takes values in (M_n) and condition (2) holds, then Doob's lemma can be substantially strengthened. Let us call a nonrandom sequence (d_n) a *barrier* for the random sequence $X = (X_n)$ if the expected number of intersections of (d_n) by the trajectories of X is finite, i.e.

$$\sum_{n=0}^{\infty}[P(X_n \leq d_n, X_{n+1} > d_{n+1}) + P(X_n > d_n, X_{n+1} \leq d_{n+1})] < \infty.$$

Theorem 3 in Sonin (1987) about the existence of barriers for processes with finite variation and which take only a bounded number of values implies the following:

Theorem 2. *Let (X_n) be a bounded (sub)martingale with values in (M_n) and assume that condition (2) holds. Then inside of each interval (a, b) there exists a barrier (d_n), $(d_n \in (a, b), n \in \mathbf{N})$.*

Now we will describe the path that leads to the formulation and proof of Theorem 2 and point (c) of DS theorem.

5. Markov Decision Processes and Sufficiency of Markov Strategies. One of the classical problem in the general theory of Markov Decision Processes is: when do Markov (or any other specific) strategies ensure the same payoff as general strategies depending on the whole past? Probably the most difficult functional in this regard is the Dubins-Savage functional $E_x^\pi \limsup f(x_n)$. The simplest example of such a functional is when f is the characteristic function of a subset G of a state set. In this case we have a problem of maximizing the probability of visiting the set G infinitely often. It is enough to mention that the sufficiency of Markov strategies for this functional is still an open problem though the statement seems obvious and very few doubt that it is true. The proof of this statement even for a finite state set (for a slightly more general functional) was given in Hill (1979) and requires a more than fifteen pages. A simple proof of Hill's theorem given in Sonin (1991a) follows easily from the Feinberg's inequality, which is presented in the next section.

6. The Equivalent Random Sequences. Let us call the random sequences $X = (X_n)$ and $Y = (Y_n)$ with values in discrete disjoint sets (M_n) equivalent $(X \sim Y)$ if for all $A \subseteq M_n, B \subseteq M_{n+1}, n \in \mathbf{N}$

$$P(X_n \in A, X_{n+1} \in B) = P(Y_n \in A, Y_{n+1} \in B). \tag{5}$$

It is obvious that every class of equivalent random sequences contains a nonhomogeneous Markov chain and, vice versa, every Markov chain defines some class of equivalent random sequences. Let $X = (X_n)$ be a random sequence with values in $(M_n), D = (D_n)$ be a sequence of sets, $D_n \subseteq M_n, n \in \mathbf{N}$. Denote by

$$P(X_n \in D_n \ ult.) \equiv P(\liminf(X_n \in D_n)) \equiv P(\cup_k \cap_{n \geq k} (X_n \in D_n))$$

the probability that X does not leave (D_n) after some random time. The following theorem was stated in Sonin (1987) and a proof was given in Sonin (1991a).

Theorem 3. (E.A. Feinberg's inequality). *Let $Z = (Z_n)$ be a Markov chain with values in (M_n), (M_n) satisfies the condition* (2) *and (D_n) be a sequence of sets, $D_n \subseteq M_n, n \in N$. Then for every random sequence $X = (X_n)$ equivalent to Z.*

$$P(Z_n \in D_n \ ult.) \leq P(X_n \in D_n \ ult.) \tag{6}$$

E. Feinberg first conjectured inequality (6) and suggested how it could be used in Markov Decision Models, so we labeled (6) with his name. Note that an example in Sonin (1994) shows the Feinberg's inequality is not true in the countable case but this fact does not contradict the sufficiency of Markov strategies for the Dubins-Savage functional.

7. The Simple Model of Irreversible Process. A simple physical model and physical interpretation of the DS Theorem for a particular Markov chain was also introduced in Sonin (1987). For each moment of time let M_n represent a set of "vessels" containing a "solution" of a given concentration of some substance. Then a vessel $i \in M_n$ can be characterized by a *volume* of solution $m_n(i)$ and its *concentration* $\alpha_n(i), 0 \leq \alpha \leq 1$. The matrix P_n describes the redistribution of the solution from the vessels M_n to the (initially empty) vessels M_{n+1} at the time of the nth transition. In other words the sequence $(m_n(i), \alpha_n(i)), i \in M_n, n \in \mathbf{N}$, for the sake of brevity called *colored* (discrete) *flow*, satisfies the relations

$$m_{n+1}(j) = \sum_i m_n(i)p_n(i,j), \alpha_{n+1}(j) = \sum_i \alpha_n(i)p_n(i,j)/m_{n+1}((j), \qquad (7)$$

where $j \in M_{n+1}$ and the sum is taken over M_n.

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 (7) for $n > k$.

We will also consider slightly more general colored flows, allowing for a sequence of vessels $(O_n), O_n \subseteq M_n, n \in \mathbf{N}$, called an *"ocean"*, where by definition $\alpha_n(i) \equiv 0, i \in O_n$, for all $n \in \mathbf{N}$, (instead of being defined by the second of formulas (7)). If all $\alpha_n(i)$ in a colored flow are constant, it is called a *flow*. It is obvious that every Markov chain Z specifies a flow $m_n(i) = P(Z_n = i), i \in M_n, n \in \mathbf{N}$ and vice versa. The colored flows also have a simple interpretation. Let $Z \in U$ be a Markov chain and let (D_n) be a sequence of sets, $D_n \subseteq M_n, n \geq k$. Let us denote

$$\alpha_n(i) = \begin{cases} P(Z_s \in D_s, s = k, \ldots, n | Z_n = i) & \text{if } P(Z_n = i) > 0, n \geq k, \\ 0 & \text{otherwise.} \end{cases} \qquad (8)$$

It is easy to check that the sequence $(m_n(i), \alpha_n(i)), n \geq k$ specifies a colored flow with an ocean $(O_n), O_n = M_n \backslash D_n, n \in N$ and initial values for $\alpha_k(i) = 1$ for $i \in D_k, \alpha_k(i) = 0$ otherwise. Vice versa, for every colored flow $(m_n(i), \alpha_n(i))$ with initial data of concentrations equal to zero or one, there is a pair $((Z_n), (D_n)), (Z_n) \in U, D_n \subseteq M_n, n \in \mathbf{N}$, such that $\alpha_n(i)$ given by (8) coincide with $\alpha_n(i)$.

It is also easy to check that for each such pair, or equivalently for a colored flow, that a random sequence (Y_n) specified by

$$Y_n = \alpha_n(Z_n), n \in \mathbf{N}, \qquad (9)$$

where $\alpha_n(i)$ *s'* are given by (8), is a *submartingale in reverse time.* This simple fact is the bridge between DS theorem and Theorem 2.

The DS theorem for Markov chains can be reformulated as a theorem about the asymptotic behavior of colored flows. It is intuitively clear that the colored flow described above is probably the simplest example of an *irreversible process*, i.e. a process whose sequence of states in reversed time is not an admissible sequence in a forward time. It is obviously true for all colored flows except the trivial cases when the initial concentration is constant or there is no mixing at all. Thus the DS theorem can be presented as a statement about the decomposition of irreversible processes. The irreversibility is strongly related to the notion of ordering or ordered structures. 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 Marshall and Olkin (1972) for the theory of majorization and to Sonin (1988), where the relation between the DS theorem and majorization theory is briefly described.

8. Countable Case. The main result of the unpublished paper Sonin (1994) is the following:

Theorem 4. *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*

a) the submartingale (in reverse time) (Y_n) *specified by* (8) *and* (9) *has no barriers inside of some interval* (a, b),

b) there exists a random sequence (X_n) *with values in* (M_n) *equivalent to* (Z_n) *and such that Feinberg's inequality* (6) *is violated.*

Note that while the above statement shows that the DS theorem is not true in the form presented in Section 2, it is none the less possible that its analog may exists in the countable case if the expected *number* of intersections is replaced by other characteristics of the transitions between elements (J_n^k) of the decomposition.

9. "0-1" Law for Nonhomogeneous Markov Chains. This result was presented in different variants and proved in Sonin (1991a). The statements and proof are very simple, so they were referred to as "may be known but we know of no reference."

Let us remind the reader of Kolmogorov's "0-1" law for a sequence of independent random variables. Let (ξ_n) be a sequence of independent random variables, and let Ψ be the "tail" σ-algebra, i.e. $\Psi = \cap_{n=1}^{\infty} F_{n\infty}$, where $F_{n\infty}$ is the σ-algebra generated by $(\xi_n, \xi_{n+1}, \ldots), n \in \mathbf{N}$. Then if $A \in \Psi$, we have that $P(A) = 0$ or $P(A) = 1$.

The heuristic formulation of the "0-1" Law for nonhomogeneous Markov chain is the following: let $Z = (Z_n)$ be a nonhomogeneous Markov chain with values in discrete spaces (M_n) (not necessarily with a bounded number of elements), Ψ be the corresponding "tail" σ-algebra, and $A \in \Psi$. Then for large n with probability near one the trajectories of Z_n are in states i, where $P(A|Z_n = i)$ is near 0 or near 1.

The precise formulation is the following. Denote by $P(A|i) \equiv P(A|Z_n = i)$, $i \in M_n, n \in \mathbf{N}, M_n(p,q) = \{i \in M_n : p \leq P(A|i) \leq q\}, B_n(p,q) = \{Z_n \in M_n(p,q)\}, 0 \leq p < q \leq 1$.

Lemma 1. ("0-1" Law for nonhomogeneous Markov chains). *Let $Z = (Z)_n$ be a nonhomogeneous Markov chain with values in (M_n) and $A \in \Psi$. Then for any $0 < p < q < 1$*

a) $\lim_n P(B_n(p,1)) = \lim_n P(AB_n(p,1)) = P(A)$,

b) $\lim_n P(B_n(0,p)) = 1 - P(A)$,

c) $\lim_n P(B_n(p,q)) = 0$.

10. The DS Theorem in Backward Time. Let the sequences (M_n) and (P_n) be defined for $n \in \mathbf{N}_- = \{\ldots,-1,0\}$ or $n \in \mathbf{N}_\infty = \{\ldots,-1,0,1,\ldots\}$ and let us assume that condition (2) is satisfied. Denote by $U_-, (U_\infty)$ the corresponding family of nonhomogeneous Markov chains $Z = (Z_n)$ (in forward time). Note that in contrast with the family U, a priori it is not clear that such Markov chains exist at all. It can be proved that the full analog of the DS in forward time is valid but these results are not published yet.

11. The DS Theorem and Simulated Annealing. Another interesting and important field of application of the DS theorem is the study of the simulated annealing algorithms, where the asymptotic behavior of nonhomogeneous Markov chains plays a crucial role. A successful attempt in this direction was undertaken in Cohn (1995).

REFERENCES

BLACKWELL, D. (1945). Finite non-homogeneous chains. *Ann. Math.*. **46** 594-599.

CHUNG, K. L. (1960). *Markov chains with stationary transition probabilities.* Springer, Berlin.

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

COHN, H. (1976). Finite nonhomogeneous Markov chains: asymptotic behavior. *Adv. Appl. Prob.* **8** 502-516.

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

COHN, H. (1995). On the simulated annealing chain. Preprint. Department of Statistics, University of Melbourne.

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

ISAACSON, D. L. AND MADSEN, R. W. (1976). *Markov chains: Theory and Applications.* Wiley, New York.

KEMENY, J. G. AND SNELL, J. L. (1960). *Finite Markov chains.* Van Nostrand Reinhold, Princeton, NJ.

KEMENY, J. G, SNELL, J. L. AND KNAPP, A. W. (1976). *Denumerable Markov chains.* Springer-Verlag, Berlin.

KOLMOGOROFF, A. N. (1936). Zur Theorie der Markoffschen Ketten. *Math. Ann. Bd.* **112**.

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

SHIRYAYEV, A. N. (1984). *Probability.* Springer-Verlag, Berlin.

SONIN, I. M. (1987). A theorem on separation of jets and some properties of random sequences. *Stochastics* **21** 231-250.

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** (Springer, 1988) 275-282.

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.

SONIN, I. M. (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.

SONIN, I. M. (1994). Counterexample to the Decomposition-Separation theorem in countable case. Manuscript.

THORISSON, H. (1988). Backward limits. *Ann. Prob.* **16** 914-924.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
CHARLOTTE, NC 28223
fma00is1@unccvm.uncc.edu