MONOTONICITY IN LAG FOR NON-MONOTONE
MARKOV CHAINS

Rafał Kulik
Mathematical Institute
University of Wrocław
E-mail: rkuli@math.uni.wroc.pl
and
Department of Mathematics and Statistics
University of Ottawa
585 King Edward Av.
K1N 6N5 Ottawa, ON,
Canada
E-mail: Rafal.Kulik@science.uottawa.ca

March 12, 2004

To appear in Probability in Engineering and Information Sciences

Keywords: monotonicity in lag; Markov chains; supermodular ordering;

Short title: Monotonicity in lag for Markov chains

AMS 2000 Subject Classification: 60J10, 60E15

Abstract
It is well known that for a stochastically monotone Markov chain \( \{J_n\}_{n \geq 1} \) a function \( \gamma(n) = \text{Cov}[f(J_1), g(J_n)] \) is decreasing if \( f \) and \( g \) are increasing. We prove this property for a special subclass of non-monotone double stochastic Markov chains.

1 Introduction
Since the Daley’s [2] paper it is well known that stochastically monotone Markov chain \( \{J_n\}_{n \geq 1} \) has the following property. Let \( f : \mathbb{R} \to \mathbb{R}_+ \) be monotone and \( \gamma(n) = \text{Cov}[f(J_1), f(J_n)] \). Then a map \( n \to \gamma(n) \) is decreasing. Recall that a real-valued homogeneous discrete time Markov chain \( \{J_n\}_{n \geq 1} \) is stochastically monotone, if its one step transition probability function \( P(J_{n+1} > x | J_n = y) \) is non-decreasing in \( x \) for every fixed \( y \). Daley’s result was extended in Bergmann and Stoyan [1]. Hu and Joe [3], (see also Joe [5], Theorems 8.3, 8.4, 8.7) stated conditions for the concordance ordering of bivariates of Markov chains under some monotonicity assumptions. The most general result was given in Hu and Pan [4]. If both the Markov chain and its time-reversed counterpart are stochastically
monotone then \(\gamma(n_1, \ldots, n_m) = E[f(J_{n_1}, \ldots, J_{n_m})]\) is decreasing in \((n_1, \ldots, n_m)\) coordinatewise for each \(m \geq 2\) and supermodular \(f\). We refer to Müller and Stoyan [10] for a review of a recent work in this area. The monotonicity of a Markov chain is also sufficient for association of it (Lindqvist [7]). However, monotonicity is not necessary, see the Lindqvist’s paper for a counterexample.

In our note we define a family of stationary and homogeneous Markov chains which are not necessary monotone. For that family we derive monotonicity of bi-variates w.r.t. supermodular order and hence monotonicity of covariances. Therefore, we show that an answer to a question which appears in [9] is negative.

The paper is organized as follows. In Section 2 we collect some needed definitions and preliminary results. In Section 3 we define the family of Markov chains and prove the main result. Section 4 is devoted to examples and counterexamples. Some additional comparison results for our class are given in Section 5.

2 Preliminaries

Define for \(1 \leq i \leq m\) and \(\varepsilon > 0\) a difference operator \(\Delta_i^\varepsilon\) by

\[
\Delta_i^\varepsilon \varphi(u_1, \ldots, u_m) = \varphi(u_1, \ldots, u_{i-1}, u_i + \varepsilon, u_{i+1}, \ldots, u_m) - \varphi(u_1, \ldots, u_n)
\]

for given \(u_1, \ldots, u_m\). A function \(\varphi : \mathbb{R}^m \to \mathbb{R}\) is called supermodular if for all \(1 \leq i < j \leq m\) and \(\varepsilon_i, \varepsilon_j > 0\),

\[
\Delta_i^\varepsilon_i \Delta_j^\varepsilon_j \varphi(u) \geq 0
\]

for all \(u = (u_1, \ldots, u_m)\). For smooth functions the above condition is equivalent to

\[
\frac{d^2}{du_i du_j} \varphi(u) \geq 0
\]

for all \(1 \leq i < j \leq n\). The standard examples of supermodular functions are:

\(u_1 \times \cdots \times u_m\), \(-\max(u_1, \ldots, u_m)\), \(\min(u_1, \ldots, u_m)\), \(h(u_1 + \cdots + u_m)\), where \(h : \mathbb{R} \to \mathbb{R}\) is convex.

This class of functions induce so called supermodular order. For arbitrary random vectors \((Y_1, \ldots, Y_m), (\tilde{Y}_1, \ldots, \tilde{Y}_m)\) we write

\[
(Y_1, \ldots, Y_m) <_{\text{sm}} (\tilde{Y}_1, \ldots, \tilde{Y}_m)
\]

if \(E[\varphi(Y_1, \ldots, Y_m)] \leq E[\varphi(\tilde{Y}_1, \ldots, \tilde{Y}_m)]\) for all supermodular \(\varphi\) such that the respective expectations are finite. Similarly, for stationary random sequences \(\{Y_n\}_{n \geq 1}\) and \(\{\tilde{Y}_n\}_{n \geq 1}\) we write \((Y_n) <_{\text{sm}} (\tilde{Y}_n)\) if for all \(m \geq 1\),

\[
(Y_1, \ldots, Y_m) <_{\text{sm}} (\tilde{Y}_1, \ldots, \tilde{Y}_m).
\]

Supermodular ordering is a dependence ordering in the sense of Joe [5]. In particularly, if \((Y_1, \ldots, Y_m) <_{\text{sm}} (\tilde{Y}_1, \ldots, \tilde{Y}_m)\) then

- \(Y_i = \tilde{Y}_i\) for all \(i = 1, \ldots, m\),
\[ \text{Cov}[Y_i, Y_j] \leq \text{Cov}[\tilde{Y}_i, \tilde{Y}_j] \text{ for all } i, j = 1, \ldots, m. \]

Consider now two stationary homogeneous Markov chains \( \{J_n\}_{n \geq 1}, \{\tilde{J}_n\}_{n \geq 1} \) with the state space \( \{1, \ldots, N\} \), the same stationary distribution \( \pi = (\pi_1, \ldots, \pi_N) \) and transition matrices \( P = (p_{ij})_{i,j=1}^N, \tilde{P} = (\tilde{p}_{ij})_{i,j=1}^N \), respectively. We shall write
\[ P \prec_{sm} \tilde{P} \]
if \( (J_1, J_2) \prec_{sm} (\tilde{J}_1, \tilde{J}_2) \). Here in the sequel we shall assume that all matrices have dimension \( N \times N \).

For the supermodular ordering of \( (J_1, J_2) \) and \( (\tilde{J}_1, \tilde{J}_2) \) we have the following characterization (cf. Hu and Pan [4]).

**Lemma 2.1** Let \( \Pi \) be a diagonal matrix \( \text{diag}(\pi_1, \ldots, \pi_N) \). Then
\[ (J_1, J_2) \prec_{sm} (\tilde{J}_1, \tilde{J}_2) \]
if and only if
\[ \sum_{i=1}^r \sum_{j=1}^s (\Pi P)_{ij} \leq \sum_{i=1}^r \sum_{j=1}^s (\Pi \tilde{P})_{ij} \]
for all \( r, s \in \{1, \ldots, N\} \).

The condition (2.2) is equivalent to a concordance ordering of bivariates, i.e.
\[ P(J_1 \leq r, J_2 \leq s) \leq P(\tilde{J}_1 \leq r, \tilde{J}_2 \leq s) \]
for every \( r, s \) in the state space. Note that for the double stochastic matrices \( P \) and \( \tilde{P} \) the condition (2.2) is equivalent to
\[ \sum_{i=1}^r \sum_{j=1}^s p_{ij} \leq \sum_{i=1}^r \sum_{j=1}^s \tilde{p}_{ij}. \]

Moreover, supermodular ordering for the double stochastic matrices can be characterized in the following way. Let
\[ T = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \]

Note that
\[ T^d = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & 1 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \]

The matrix \( T \) was used in Keilson and Kester [6] for obtaining stochastic monotonicity. A matrix \( A \) is stochastic monotone if and only if \( T^{-1}AT \geq 0 \), where \( 0 \) is a matrix which consists of zeros.
Lemma 2.2 Assume that $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are double stochastic. Then $\mathbf{P} <_{\text{sm}} \tilde{\mathbf{P}}$ if and only if $T(\tilde{\mathbf{P}} - \mathbf{P})T' \geq 0$.

Proof. Inequality $T(\tilde{\mathbf{P}} - \mathbf{P})T' \geq 0$ is equivalent to $\sum_{i=1}^{r} \sum_{j=1}^{s} \tilde{p}_{ij} \geq \sum_{i=1}^{r} \sum_{j=1}^{s} p_{ij}$ for all $r$ and $s$, which means $\mathbf{P} <_{\text{sm}} \tilde{\mathbf{P}}$ by Eq. (2.3).

3 Main result

We introduce a class of Markov chains which are not necessary monotone.

Definition 3.1 We say that a matrix $\mathbf{A}$ belongs to a class $\mathcal{PS}$ if it is stochastic and $T\mathbf{A}T^{-1} \geq 0$. Equivalently, we say that a stationary homogeneous Markov chain belongs to $\mathcal{PS}$ if its transition probability matrix does.

Let $\mathbf{a}$ and $\mathbf{b}$ be $N$-dimensional vectors. We say that $\mathbf{a}$ precedes $\mathbf{b}$ in the partial sum ordering ($\mathbf{a} <_{\text{ps}} \mathbf{b}$) if $\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i$, $k = 1, \ldots, N$. Now, denote by $\mathbf{a}_i$, $i = 1, \ldots, N$, the columns of a matrix $\mathbf{A}$. Then $\mathbf{A} \in \mathcal{PS}$ if and only if

$$\mathbf{a}_N \prec_{\text{ps}} \mathbf{a}_{N-1} \prec_{\text{ps}} \cdots \prec_{\text{ps}} \mathbf{a}_1.$$ 

Denote now by $\mathcal{DS}$ a class of double stochastic matrices. It turns out that for the Markov chains driven by $\mathbf{P} \in \mathcal{PS} \cap \mathcal{DS}$ the following monotonicity property holds.

Theorem 3.1 Let $\{J_n\}_{n \geq 1}$ be a stationary homogeneous Markov chain with a transition probability matrix $\mathbf{P} \in \mathcal{PS} \cap \mathcal{DS}$. Then for any $n \geq 1$,

$$(J_1, J_{n+1}) <_{\text{sm}} (J_1, J_n).$$

Corollary 3.1 Under conditions of Theorem 3.1 we have for the functions $f$ and $g$ both either nonincreasing or nondecreasing,

$$\text{Cov}[f(J_0), g(J_{n+1})] \leq \text{Cov}[f(J_0), g(J_n)].$$

The proof of Theorem 3.1 consists of the sequence of Lemmas.

Lemma 3.1 1. If $\mathbf{A}, \mathbf{B} \in \mathcal{PS}$ then $\mathbf{A} \mathbf{B} \in \mathcal{PS}$;

2. If $\mathbf{A}(i) \in \mathcal{PS}$, $i = 1, \ldots, k$ and $\mathbf{w} = (w_1, \ldots, w_k)$ is a probability vector then $\sum_{i=1}^{k} w_i \mathbf{A}(i) \in \mathcal{PS}$.

Proof.

1. We have $T\mathbf{A}^{-1} \mathbf{T}^{-1} = T\mathbf{A}^{-1} \mathbf{T}^{-1} \mathbf{B}^{-1} \geq 0$ from the assumptions on $\mathbf{A}$ and $\mathbf{B}$.

2. Obvious.
Lemma 3.2 Assume that $B, \tilde{B} \in DS$ and $A \in PS$. If $B <_{sm} \tilde{B}$ then $A\tilde{B} <_{sm} AB$ and $BA <_{sm} \tilde{B}A$.

Proof. We have

$$T(A\tilde{B} - AB)T^t = TA(\tilde{B} - B)T^t = [TAT^{-1}][T(\tilde{B} - B)T^t].$$

The first term in brackets is greater than 0 due to assumptions on $A$, the second one has the same property because of Lemma 2.2. Therefore $T(A\tilde{B} - AB)T^t \geq 0$ and using Lemma 2.2 once more we obtained the comparison result.

Lemma 3.3 Assume that $A \in PS \cap DS$. Then $A^2 <_{sm} A$. Moreover, for every $n \geq 1$, $A^{n+1} <_{sm} A^n$.

Proof. Note that for any double stochastic matrix $A$ we have $A <_{sm} I$, where $I$ is an identity matrix. Moreover, $I \in DS$. From Lemma 3.2 we have $A^2 <_{sm} A$. Assume now that $A^n <_{sm} A^{n-1}$. Because of Lemma 3.1 $A^n$ and $A^{n-1}$ belong to $PS$. Moreover, they are double stochastic. Therefore, Lemma 3.2 applies and we have $AA^n <_{sm} AA^{n-1}$.

Proof of the Theorem 3.1. Denote by $P^{(k)} = (p_{ij}^{(k)})_{i,j=1}^N$, $k \geq 1$, the $k$-step probability matrix for $\{J_n\}_{n \geq 1}$. From Lemma 3.3 we have that

$$\sum_{i=1}^r \sum_{s=1}^s p_{ij}^{(n+1)} \leq \sum_{i=1}^r \sum_{s=1}^s p_{ij}^{(n)}$$

for all $r, s = 1, \ldots, N$. Therefore, $(J_1, J_{n+1}) <_{sm} (J_1, J_n)$.

The Corollary 3.1 follows from stationarity of a Markov chain and the fact that for the functions $f$ and $g$ both either nonincreasing or nondecreasing the function $\varphi(x, y) = f(x)g(y)$ is supermodular.

4 Examples and Counterexamples

Example 4.1 Let $p$ and $\varepsilon > 0$ be such that $p \geq 2\varepsilon$ and $p + \varepsilon \leq 1$. Then the following matrix belongs to $PS \cap DS$ and is not monotone.

$$P = \frac{1}{4p - 2\varepsilon} \begin{bmatrix} p + \varepsilon & p + \varepsilon & p - 2\varepsilon & p - 2\varepsilon \\ p - \varepsilon & p - \varepsilon & p & p \\ p & p & p - \varepsilon & p - \varepsilon \\ p - 2\varepsilon & p - 2\varepsilon & p + \varepsilon & p + \varepsilon \end{bmatrix}.$$
Example 4.2  The lemma 3.3 fails if one of the assumptions is removed. Let

\[
A = \frac{1}{36} \begin{bmatrix}
18 & 0 & 18 & 0 \\
6 & 12 & 6 & 12 \\
7 & 11 & 7 & 11 \\
5 & 13 & 5 & 13 \\
\end{bmatrix}.
\]

Then \(A \in DS, A \notin PS\) and it is not true that \(A^2 <_{sm} A\). The above matrix was used in Lindqvist [7] as a counterexample of a Markov chain which is associated but not stochastically monotone.

Let now

\[
A = \frac{1}{4} \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
2 & 1 & 1 & 0 \\
2 & 1 & 1 & 0 \\
\end{bmatrix}.
\]

This matrix belongs to \(PS\) but is neither double stochastic nor monotone. It is not true that \(A^2 <_{sm} A\).

5  Additional comparison results

Observe that for \(a \leq a'\) and stochastic matrices \(A, B\) with the same invariant distributions we have \(a'A + (1-a')B <_{sm} aA + (1-a)B\) if \(A <_{sm} B\). This can be interpreted as follows. Let \(\{J_n\}_{n \geq 1}, \{\tilde{J}_n\}_{n \geq 1}\) be the stationary homogeneous Markov chains with transition matrices \(A\) and \(B\), respectively. Define \(Z_n = \Theta J_n + (1-\Theta)\tilde{J}_n\) and \(\tilde{Z}_n = \Theta \tilde{J}_n + (1-\Theta)J_n\), where \(\Theta, \tilde{\Theta}\) are Bernoulli random variables with \(P(\Theta = 1) = a, P(\tilde{\Theta} = 1) = a'\), independent of everything else. Then \((\tilde{Z}_1, \tilde{Z}_2) <_{sm} (Z_1, Z_2)\) provided \((J_1, \tilde{J}_2) <_{sm} (\tilde{J}_1, J_2)\) and \(a \leq a'\). If we have some additional monotonicity properties we can have comparison of the whole sequences \(\{Z_n\}_{n \geq 1}\) and \(\{\tilde{Z}_n\}_{n \geq 1}\). However, the above consideration can not be rewritten for the random variables \(\Theta, \tilde{\Theta}\) assuming their values in \(\{1, \ldots, K\}\). This can be done if the transition matrices belong to \(PS \cap DS\). We refer to Marshall and Olkin [8] for the concept of majorization.

Proposition 5.1  Assume that

(a) \(P(k) := (p_{ij}(k))_{i,j=1}^N \in PS \cap DS, k=1, \ldots, K;\)

(b) \(P(1) <_{sm} P(2) <_{sm} \cdots <_{sm} P(K);\)

(c) \(a\) is majorized by \(b\) (we write \(a < b\)), where \(a, b\) are \(K\)-dimensional probability vectors with coordinates arranged in the decreasing order.

Then

\[
P := \sum_{k=1}^K b_k P(k) <_{sm} \sum_{k=1}^K a_k P(k) =: \tilde{P}.
\]
Proof. Define for \( l = 1, \ldots, N \) and \( r = 1, \ldots, K \) the vectors
\[
\mathbf{v}^{(l,r)} := (v_1^{(l,r)}, \ldots, v_N^{(l,r)}),
\]
where for \( j = 1, \ldots, N \), \( v_j^{(l,r)} \) are defined as
\[
v_j^{(l,r)} := \sum_{i=1}^{l} p_{ij}(r).
\]
According to assumption (a) these vectors have decreasing coordinates. Moreover, because of (b) we have
\[
\mathbf{v}^{(l,1)} \prec \mathbf{v}^{(l,2)} \prec \cdots \prec \mathbf{v}^{(l,K)}
\]
for each fixed \( l \). Take now
\[
\mathbf{w}^l := \sum_{r=1}^{K} b_r \mathbf{v}^{(l,r)}
\]
and
\[
\tilde{\mathbf{w}}^l := \sum_{r=1}^{K} a_r \mathbf{v}^{(l,r)}.
\]
From Marshall and Olkin [8, p. 125] we have \( \mathbf{w}^l \prec \tilde{\mathbf{w}}^l \) for every \( l = 1, \ldots, N \). Because \( \mathbf{w}^l \) and \( \tilde{\mathbf{w}}^l \) have decreasing coordinates majorization order implies \( \mathbf{w}^l \prec_{ps} \tilde{\mathbf{w}}^l \). Observe now that \( j \)th coordinates of \( \mathbf{w}^l \) and \( \tilde{\mathbf{w}}^l \) can be written as
\[
\mathbf{w}^l(j) = \sum_{i=1}^{l} p_{ij}
\]
and
\[
\tilde{\mathbf{w}}^l(j) = \sum_{i=1}^{l} \tilde{p}_{ij},
\]
respectively. It implies comparison result.

Acknowledgement The author is grateful to Ryszard Szekli for many helpful comments. This work was partially done during the author’s stay at University of Ottawa.

References


