Probability in the Engineering and Informational Sciences, **18**, 2004, 485–492. Printed in the U.S.A.

MONOTONICITY IN LAG FOR NONMONOTONE MARKOV CHAINS

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

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

1. INTRODUCTION

Since Daley's [2] article, it is well known that the stochastically monotone Markov chain $\{J_n\}_{n\geq 1}$ has the following property. Let $f: \mathbb{R} \to \mathbb{R}_+$ be monotone and $\gamma(n) = \mathbf{Cov}[f(J_1), f(J_n)]$. 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 $P(J_{n+1} > x \mid J_n = y)$ is nondecreasing in x for every fixed y. Daley's result was extended in Bergmann and Stoyan [1]. Hu and Joe [3] (see also Joe [5, Thms. 8.3, 8.4, and 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

 0269-9648/04 $16.00

also sufficient for association of it (Lindqvist [7]). However, monotonicity is not necessary; see Lindqvist's article for a counterexample.

In our article, we define a family of stationary and homogeneous Markov chains which are not necessary monotone. For that family, we derive the monotonicity of bivariates 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 article 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 $1 \le i \le m$ and $\varepsilon > 0$ a difference operator Δ_i^ε 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 $1 \le i < j \le m$ and $\varepsilon_i, \varepsilon_j > 0$,

$$\Delta_i^{\varepsilon_i} \Delta_j^{\varepsilon_j} \varphi(\mathbf{u}) \ge 0$$

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

$$\frac{d^2}{du_i\, du_j} \varphi(\mathbf{u}) \ge 0$$

for all $1 \le i < j \le n$. The standard examples of supermodular functions are $u_1 \times \cdots \times u_m$, $-\max(u_1,\ldots,u_m)$, $\min(u_1,\ldots,u_m)$, and $h(u_1 + \cdots + u_m)$, where $h:\mathbb{R} \to \mathbb{R}$ is convex.

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

$$(Y_1,\ldots,Y_m) <_{\text{sm}} (\tilde{Y}_1,\ldots,\tilde{Y}_m) \tag{2.1}$$

if $\mathbb{E}[\varphi(Y_1,\ldots,Y_m)] \le \mathbb{E}[\varphi(\tilde{Y}_1,\ldots,\tilde{Y}_m)]$ for all supermodular φ such that the respective expectations are finite. Similarly, for stationary random sequences $\{Y_n\}_{n\ge1}$ and $\{\tilde{Y}_n\}_{n\ge1}$, we write $\{Y_n\} <_{\text{sm}} \{\tilde{Y}_n\}$ if for all $m \ge 1$,

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

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

- $Y_i \overset{\text{d}}{=} \tilde{Y}_i$ for all $i = 1,\ldots,m$,
- $\mathbf{Cov}[Y_i,Y_j] \le \mathbf{Cov}[\tilde{Y}_i,\tilde{Y}_j]$ for all $i,j = 1,\ldots,m$.

Consider now two stationary homogeneous Markov chains $\{J_n\}_{n\geq 1}$ and $\{\tilde{J}_n\}_{n\geq 1}$ with the state space $\{1,\dots,N\}$, the same stationary distribution $\pi = (\pi_1,\dots,\pi_N)$, and transition matrices $\mathbf{P} = (p_{ij})_{i,j=1}^N$ and $\tilde{\mathbf{P}} = (\tilde{p}_{ij})_{i,j=1}^N$, respectively. We will write

$$\mathbf{P} <_{\mathrm{sm}} \tilde{\mathbf{P}}$$

if $(J_1, J_2) <_{\mathrm{sm}} (\tilde{J}_1, \tilde{J}_2)$. Here, in the sequel, we will 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* $\mathbf{\Pi}$ *be a diagonal matrix* $\mathrm{diag}(\pi_1,\dots,\pi_N)$. *Then,*

$$(J_1, J_2) <_{\mathrm{sm}} (\tilde{J}_1, \tilde{J}_2)$$

if and only if

$$\sum_{i=1}^{r}\sum_{j=1}^{s} (\mathbf{\Pi P})_{ij} \leq \sum_{i=1}^{r}\sum_{j=1}^{s} (\mathbf{\Pi}\tilde{\mathbf{P}})_{ij} \qquad \textbf{(2.2)}$$

for all $r, s \in \{1,\dots,N\}$.

Condition (2.2) is equivalent to a *concordance ordering* of bivariates; that is,

$$P(J_1 \leq r, J_2 \leq s) \leq P(\tilde{J}_1 \leq r, \tilde{J}_2 \leq s)$$

for every r and s in the state space. Note that for the double stochastic matrices $\mathbf{P}$ and $\tilde{\mathbf{P}}$, condition (2.2) is equivalent to

$$\sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} \leq \sum_{i=1}^{r}\sum_{j=1}^{s} \tilde{p}_{ij}. \qquad \textbf{(2.3)}$$

Moreover, supermodular ordering for the double stochastic matrices can be characterized in the following way. Let

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 1 & 1 & 0 & 0 & \dots \\ 1 & 1 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Note that

$$\mathbf{T}^t = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots \\ 0 & 1 & 1 & 1 & \dots \\ 0 & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad \mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

https://doi.org/10.1017/S0269964804184040 Published online by Cambridge University Press

The matrix $\mathbf{T}$ was used in Keilson and Kester [6] for obtaining stochastic monotonicity. A matrix $\mathbf{A}$ is stochastic monotone if and only if $\mathbf{T}^{-1}\mathbf{AT} \geq \mathbf{0}$, where $\mathbf{0}$ 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* $\mathbf{T}(\tilde{\mathbf{P}} - \mathbf{P})\mathbf{T}^t \geq \mathbf{0}$.

PROOF: Inequality $\mathbf{T}(\tilde{\mathbf{P}} - \mathbf{P})\mathbf{T}^t \geq \mathbf{0}$ 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* $\mathbf{TAT}^{-1} \geq \mathbf{0}$. *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 $\mathbf{a}_{\cdot i}$, $i = 1,\ldots,N$, the columns of a matrix $\mathbf{A}$. Then, $\mathbf{A} \in \mathcal{PS}$ if and only if

$$\mathbf{a}_{\cdot N} <_{\text{ps}} \mathbf{a}_{\cdot N-1} <_{\text{ps}} \cdots <_{\text{ps}} \mathbf{a}_{\cdot 1}.$$

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,*

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

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* $\mathbf{AB} \in \mathcal{PS}$.
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 $\mathbf{TABT}^{-1} = \mathbf{TAT}^{-1}\mathbf{TBT}^{-1} \geq \mathbf{0}$ from the assumptions on $\mathbf{A}$ and $\mathbf{B}$.
2. Obvious. ■

LEMMA 3.2: *Assume that* $\mathbf{B}, \tilde{\mathbf{B}} \in \mathcal{DS}$ *and* $\mathbf{A} \in \mathcal{PS}$. *If* $\mathbf{B} <_{\text{sm}} \tilde{\mathbf{B}}$, *then* $\mathbf{AB} <_{\text{sm}} \mathbf{A}\tilde{\mathbf{B}}$ *and* $\mathbf{BA} <_{\text{sm}} \tilde{\mathbf{B}}\mathbf{A}$.

https://doi.org/10.1017/S0269964804184040 Published online by Cambridge University Press

PROOF: We have

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

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

LEMMA 3.3: *Assume that* $\mathbf{A} \in \mathcal{PS} \cap \mathcal{DS}$. *Then,* $\mathbf{A}^2 <_{\text{sm}} \mathbf{A}$. *Moreover, for every* $n \geq 1$, $\mathbf{A}^{n+1} <_{\text{sm}} \mathbf{A}^n$.

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

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

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

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

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 $\mathcal{PS} \cap \mathcal{DS}$ and is not monotone:

$$\mathbf{P} = \frac{1}{4p - 2\varepsilon} \begin{bmatrix} p+\varepsilon & p+\varepsilon & p-2\varepsilon & p-2\varepsilon \\ p-\varepsilon & p-\varepsilon & p & p \\ p & p & p-\varepsilon & p-\varepsilon \\ p-2\varepsilon & p-2\varepsilon & p+\varepsilon & p+\varepsilon \end{bmatrix}.$$

Example 4.2: Lemma 3.3 fails if one of the assumptions is removed. Let

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

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

Now, let

$$\mathbf{A} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 0 \\ 2 & 1 & 1 & 0 \end{bmatrix}.$$

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

5. ADDITIONAL COMPARISON RESULTS

Observe that for $a \le a'$ and stochastic matrices $\mathbf{A}$ and $\mathbf{B}$ with the same invariant distributions, we have $a'\mathbf{A} + (1-a')\mathbf{B} <_{\text{sm}} a\mathbf{A} + (1-a)\mathbf{B}$ if $\mathbf{A} <_{\text{sm}} \mathbf{B}$. This can be interpreted as follows. Let $\{J_n\}_{n\ge1}$ and $\{\tilde{J}_n\}_{n\ge1}$ be the stationary homogeneous Markov chains with transition matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. Define $Z_n = \Theta J_n + (1-\Theta)\tilde{J}_n$ and $\tilde{Z}_n = \tilde{\Theta} J_n + (1-\tilde{\Theta})\tilde{J}_n$, where Θ and $\tilde{\Theta}$ are Bernoulli random variables with $P(\Theta = 1) = a$ and $P(\tilde{\Theta} = 1) = a'$, independent of everything else. Then, $(\tilde{Z}_1, \tilde{Z}_2) <_{\text{sm}} (Z_1, Z_2)$, provided $(J_1, J_2) <_{\text{sm}} (\tilde{J}_1, \tilde{J}_2)$ and $a \le a'$. If we have some additional monotonicity properties, we can have a comparison of the whole sequences $\{Z_n\}_{n\ge1}$ and $\{\tilde{Z}_n\}_{n\ge1}$. However, the above consideration cannot be rewritten for the random variables Θ and $\tilde{\Theta}$ assuming their values in $\{1, \ldots, K\}$. This can be done if the transition matrices belong to $\mathcal{PS} \cap \mathcal{DS}$. We refer to Marshall and Olkin [8] for the concept of majorization.

PROPOSITION 5.1: *Assume the following:*

(a) $\mathbf{P}(k) := (p_{ij}(k))_{i,j=1}^N \in \mathcal{PS} \cap \mathcal{DS}$, $k = 1, \ldots, K$.
(b) $\mathbf{P}(1) <_{\text{sm}} \mathbf{P}(2) <_{\text{sm}} \cdots <_{\text{sm}} \mathbf{P}(K)$.
(c) $\mathbf{a}$ *is majorized by* $\mathbf{b}$ *(we write* $\mathbf{a} \prec \mathbf{b}$*), where* $\mathbf{a}$ *and* $\mathbf{b}$ *are K-dimensional probability vectors with coordinates arranged in the decreasing order.*

Then,

$$\mathbf{P} := \sum_{k=1}^{K} b_k \mathbf{P}(k) <_{\text{sm}} \sum_{k=1}^{K} a_k \mathbf{P}(k) =: \tilde{\mathbf{P}}.$$

PROOF: Define for $l = 1, \ldots, N$ and $r = 1, \ldots, K$ the vectors

$$\mathbf{v}^{(l,r)} := (v_1^{(l,r)}, \ldots, v_N^{(l,r)}),$$

https://doi.org/10.1017/S0269964804184040 Published online by Cambridge University Press

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 assumption (b), we have

$$\mathbf{v}^{(l,1)} \prec \mathbf{v}^{(l,2)} \prec \cdots \prec \mathbf{v}^{(l,K)}$$

for each fixed l. Now, take

$$\mathbf{w}^l := \sum_{r=1}^{K} b_r \mathbf{v}^{(l,r)}$$

and

$$\widetilde{\mathbf{w}}^l := \sum_{r=1}^{K} a_r \mathbf{v}^{(l,r)}.$$

From Marshall and Olkin [8, p. 125], we have $\mathbf{w}^l \prec \widetilde{\mathbf{w}}^l$ for every $l = 1, \ldots, N$. Because $\mathbf{w}^l$ and $\widetilde{\mathbf{w}}^l$ have decreasing coordinates, majorization order implies $\mathbf{w}^l \prec_{\text{ps}} \widetilde{\mathbf{w}}^l$. Observe now that the jth coordinates of $\mathbf{w}^l$ and $\widetilde{\mathbf{w}}^l$ can be written as

$$w^l(j) = \sum_{i=1}^{l} p_{ij}$$

and

$$\widetilde{w}^l(j) = \sum_{i=1}^{l} \tilde{p}_{ij},$$

respectively. Since $\mathbf{w} \prec_{\text{ps}} \widetilde{\mathbf{w}}$, we have for each k, l

$$\sum_{i=1}^{l} \sum_{j=1}^{kp_{ij}} \leq \sum_{i=1}^{l} \sum_{j=1}^{k} \tilde{p}_i$$

which implies comparison of matrices by (2.3). ■

Acknowledgments

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

1. Bergmann, R. & Stoyan, D. (1978). Monotonicity properties of second order characteristics of stochastically monotone Markov chains. *Mathematische Nachrichten* 82: 99–102.
2. Daley, D.J. (1968). Stochastically monotone Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 10: 305–317.

3. Hu, T. & Joe, H. (1995). Monotonicity of positive dependence with time for stationary reversible Markov chains. *Probability in the Engineering and Informational Sciences* 9: 227–237.
4. Hu, T. & Pan, X. (2000). Comparisons of dependence for stationary Markov processes. *Probability in the Engineering and Informational Sciences* 14: 299–315.
5. Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman & Hall.
6. Keilson, J. & Kester, A. (1977). Monotone matrices and monotone Markov processes. *Stochastic Processes and Their Applications* 5: 231–241.
7. Lindqvist, B.H. (1987). Monotone and associated Markov chains, with applications to reliability theory. *Journal of Applied Probability* 24: 679–695.
8. Marshall, A.W. & Olkin, I. (1979). *Inequalities: Theory of majorization and its applications*. New York: Academic Press.
9. Miyoshi, N. & Rolski, T. (2004). Ross type conjectures on monotonicity of queues. *Australian & New Zealand Journal of Statistics* 46: 121–131.
10. Müller A. & Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. New York: Wiley.