

DEPENDENCE ORDERING FOR MARKOV PROCESSES ON PARTIALLY ORDERED SPACES

HANS DADUNA,* *Hamburg University*

RYSZARD SZEKLI,** *Wrocław University*

Abstract

We compare dependence in stochastically monotone Markov processes with partially ordered Polish state spaces using the concordance and supermodular orders. We show necessary and sufficient conditions for the concordance order to hold both in terms of the one-step transition probabilities for discrete-time processes and in terms of the corresponding infinitesimal generators for continuous-time processes. We give examples showing that a stochastic monotonicity assumption is not necessary for such orderings. We indicate relations between dependence orderings and, variously, the asymptotic variance-reduction effect in Monte Carlo Markov chains, Cheeger constants, and positive dependence for Markov processes.

Keywords: Concordance order; supermodular order; stochastic monotonicity; monotone Markov process; dependence order; Cheeger constant; association; MCMC

2000 Mathematics Subject Classification: Primary 60J25; 60K25

1. Introduction

The aim of this paper is to study dependence ordering for Markov processes with general partially ordered state spaces. This study is a step in extending the dependence ordering theory of finite random vectors to that of infinite random vectors. Functionals of the form $\int \psi dF$, considered for those ψ in a specified cone of functions, lead in a natural way to stochastic (pre)orderings; e.g. we say that two distributions on \mathbb{R}^d , F and G , are supermodular stochastically ordered if $\int \psi dF \leq \int \psi dG$ for all supermodular functions ψ for which the integrals exist. If F and G are thus ordered then their one-dimensional marginal distributions are necessarily equal and the covariances between their respective coordinates are ordered. Therefore, this ordering is called a *dependence ordering*. For infinite stationary sequences, i.e. where all one-dimensional marginal distributions are equal, it is natural to study random vectors which are finite segments of the sequence. A special interest in the comparison of dependence in two stationary (ergodic) sequences is motivated by examples where a stationary sequence represents an input to a complex system (e.g. a queueing or reliability system) in which the replacement of an input by one which has the same one-dimensional stationary distribution but is more dependent implies a dramatic change in the basic performance measures of the system (see, e.g. Szekli *et al.* (1994) for such an example in a queueing system). A special

Received 13 April 2005; revision received 8 May 2006.

* Postal address: Department of Mathematics, Hamburg University, Bundesstrasse 55, 20146 Hamburg, Germany.

Supported by Deutscher Akademischer Austauschdienst.

** Postal address: Mathematical Institute, Wrocław University, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland.

Email address: szekli@math.uni.wroc.pl

Supported by the KBN, grant number 2P03A02023.

case – that of a stationary Markov chain with the state space being a subset of the real line – was studied by Hu and Pan (2000). There, stochastic monotonicity of the processes makes it possible to obtain necessary and sufficient conditions for supermodular ordering and concordance ordering for finite-dimensional distributions of stationary Markov chains in discrete and continuous time. Dependence orderings of some derived stationary sequences were studied by Kulik and Szekli (2004).

Application of the above-described ideas to Markov chains technically leads to *stochastically monotone* Markov chains, introduced by Daley (1968). The property of stochastic monotonicity is very often the explanation underlying the successful application of comparison techniques in specific problems, e.g. in random walks, epidemic processes, genetics processes, and queuing processes. Keilson and Kester (1977) gave necessary and sufficient conditions for the stochastic monotonicity of Markov chains with finite state spaces, Massey (1987) provided them for countable partially ordered spaces, and the most general solution was given by Chen (2004, pp. 211–216) for Markov jump processes (and classes of diffusions as well). Classical examples of stochastically monotone processes are, among others, attractive particle systems (see Griffeath (1979)), Jackson queueing networks (see Daduna and Szekli (1995)), birth–death processes (see van Doorn (1981)), population processes, and diffusions (see Chen and Wang (1993)).

In this paper we show necessary and sufficient conditions for the concordance ordering of two stationary Markov processes under stochastic monotonicity assumptions. These conditions are given in terms of transition kernels (for discrete time) or generators (for continuous time). We elaborate on the problem of the necessity of stochastic monotonicity assumptions for dependence ordering of Markov processes. We indicate connections of this study to comparison results on asymptotic variances from the theory of Monte Carlo Markov chains, and we relate our criteria to the definition of Cheeger constants. Finally, we discuss positive dependence for Markov processes.

For a review of dependence orderings, we refer the reader to Müller and Stoyan (2002, pp. 107–121), Joe (1997, pp. 55–57), and Shaked and Shanthikumar (1994, Chapter 4.G). For applications in the modelling of multivariate portfolios and financial risks, see, e.g. Rüschendorf (2004), and for Kolmogorov-type Hajek–Renyi inequalities for negatively associated random variables and Rosenthal-type inequalities, see, e.g. Christofides and Vaggelatou (2004).

2. Definitions and preliminary results

Before introducing general definitions of dependence orderings, we recall some classical results which give a proper perspective on our extensions. Consider the class $\mathcal{H}(F_1, \dots, F_d)$ of probability measures on \mathbb{R}^d with fixed one-dimensional marginal distribution functions F_1, \dots, F_d (the so-called Fréchet class). A natural problem studied in the literature is to evaluate and compare values of $\int \psi dF$ for $F \in \mathcal{H}(F_1, \dots, F_d)$, with ψ possessing specific properties. In particular, upper and lower bounds in this class are of interest. For example, fix an $n \in \mathbb{N}$, let $d = 2$, let $F = (1/n) \sum_{i=1}^n \delta_{(a_i, b_i)}$ for arbitrary $a_i, b_i \in \mathbb{R}_+$, $i = 1, \dots, n$, and let $\psi(x_1, x_2) = x_1 x_2$. (Here δ_x is the (distribution function) of the one-point distribution in $x \in \mathbb{R}^d$.) Denote by $(a_{(i)}, i = 1, \dots, n)$ and $(b_{(i)}, i = 1, \dots, n)$ increasing rearrangements of (a_i) and (b_i) , respectively. Then, for $G = (1/n) \sum_{i=1}^n \delta_{(a_{(i)}, b_{(i)})}$, the inequality $\int \psi dF \leq \int \psi dG$ is equivalent to the inequality

$$\sum_{i=1}^n a_i b_i \leq \sum_{i=1}^n a_{(i)} b_{(i)},$$

which is exactly the Hardy *et al.* (1952, Chapter 10) rearrangement theorem. Of course, F and G have the same one-dimensional distributions,

$$F_1 = G_1 = \frac{1}{n} \sum_{i=1}^n \delta_{a_i} \quad \text{and} \quad F_2 = G_2 = \frac{1}{n} \sum_{i=1}^n \delta_{b_i},$$

but G is maximally concordant in $\mathcal{H}(F_1, F_2)$ (note that the corresponding distribution function is $G(x_1, x_2) = \min_{i=1,2} G_i(x_i)$; see Müller and Stoyan (2002, p. 109)).

Intuitively, for measures on \mathbb{R}^d which concentrate their mass on finitely many atoms, a distribution with distribution function F is *more concordant* than the distribution with distribution function G if F can be obtained from G by making a finite number of mass transpositions which add a fixed amount of mass at $\min(x, y)$ and $\max(x, y)$ while subtracting it at x and y , $x, y \in \mathbb{R}^d$, so that large and small values of selected coordinates are more often associated with large and, respectively, small values of other coordinates. Hoeffding (1940) studied related inequalities for $d = 2$ with $\psi(x_1, x_2) = f(x_1 - x_2)$ and f convex. Lorentz (1953) showed that in \mathbb{R}^d (for arbitrary d) rearrangement inequalities hold for a much larger class of functions ψ , namely all supermodular functions. Taking, e.g. $\psi(x_1, \dots, x_d) = -\log(x_1 + \dots + x_d)$, he obtained Ruderman’s inequality,

$$\prod_{i=1}^n \sum_{j=1}^d a_{ij} \geq \prod_{i=1}^n \sum_{j=1}^d a_{ij}^*,$$

where $a_{ij} \geq 0$ and a_{ij}^* , $i = 1, \dots, n$, is the decreasing rearrangement of a_{ij} , $i = 1, \dots, n$. Research done by Hoeffding and Lorentz was unified by Whitt (1976) and developed by Tchen (1980) and Rüschendorf (1980), who studied the case of arbitrary d with ψ Δ -monotone. The class of supermodular functions is rich enough to allow for comparison of maxima, minima, spans, covariances and other dependency measures for two ordered random vectors with the same one-dimensional marginal distributions.

For each given pair of concordantly ordered distributions on \mathbb{R}^d , comparing integrals of supermodular functions leads to a number of interesting inequalities. Therefore, it is natural to introduce the corresponding definitions of dependence orderings on general partially ordered spaces.

We shall consider probability measures on a partially ordered Polish space \mathbb{E} endowed with a closed partial order ‘ $<$ ’ and the Borel σ -algebra \mathcal{E} , along with random elements $X: (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{E}, \mathcal{E}, <)$. We denote by $\mathcal{I}^*(\mathbb{E})$ the set of all real-valued, increasing, measurable, bounded functions on \mathbb{E} (that f is increasing means that, for all x and y , $x < y$ implies that $f(x) \leq f(y)$), by $\mathcal{I}_+^*(\mathbb{E})$ the set of such functions that are nonnegative, and by $\mathcal{I}(\mathbb{E})$ the set of all increasing sets (i.e. sets whose indicator functions are increasing). The decreasing analogues are denoted by $\mathcal{D}^*(\mathbb{E})$, $\mathcal{D}_+^*(\mathbb{E})$, and $\mathcal{D}(\mathbb{E})$, respectively. For $A \subseteq \mathbb{E}$, we define $A^\uparrow := \{y \in \mathbb{E} : y > x \text{ for some } x \in A\}$ and $A^\downarrow := \{y \in \mathbb{E} : y < x \text{ for some } x \in A\}$. We further define $\mathcal{I}_p(\mathbb{E}) = \{\{x\}^\uparrow : x \in \mathbb{E}\}$ and $\mathcal{D}_p(\mathbb{E}) = \{\{x\}^\downarrow : x \in \mathbb{E}\}$, the classes of one-point-generated increasing and, respectively, decreasing sets.

For a $(\mathbb{E}, \mathcal{E}, <)$ which is a lattice (i.e. such that for any $x, y \in \mathbb{E}$ there exist a largest lower bound, $x \wedge y \in \mathbb{E}$, and a smallest upper bound, $x \vee y \in \mathbb{E}$, both uniquely determined), we denote by $\mathcal{L}_{sm}(\mathbb{E})$ the set of all real-valued, bounded, measurable supermodular functions on \mathbb{E} , i.e. functions which satisfy

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y) \quad \text{for all } x, y \in \mathbb{E}.$$

For product spaces we shall use the notation $\mathbb{E}^{(n)} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_n$, where $\mathbb{E}_i, i = 1, \dots, n$, are partially ordered Polish spaces. If $\mathbb{E}_i \equiv \mathbb{E}$ for all i then we write \mathbb{E}^n instead of $\mathbb{E}^{(n)}$. Analogously, we write $\mathbb{E}^{(\infty)}$ and \mathbb{E}^∞ for infinite products. Product spaces will be considered with the product topology. Elements of $\mathbb{E}^{(n)}$ will be denoted by $x^{(n)} = (x_1, \dots, x_n)$ and elements of $\mathbb{E}^{(\infty)}$ by $x^{(\infty)}$. For random elements we use capital letters in this notation. We denote the coordinatewise ordering on $\mathbb{E}^{(n)}$ by ‘ \prec^n ’.

Definition 2.1. We say that two random elements X and Y of $(\mathbb{E}, \mathcal{E}, \prec)$ are stochastically ordered (and write $X \prec_{st} Y$ or $Y \succ_{st} X$) if $E[f(X)] \leq E[f(Y)]$ for all $f \in \mathcal{I}^*(\mathbb{E})$ for which the expectations exist.

We say that two random elements X and Y of $(\mathbb{E}, \mathcal{E}, \prec)$ are supermodular stochastically ordered (and write $X \prec_{sm} Y$ or $Y \succ_{sm} X$) if $E[f(X)] \leq E[f(Y)]$ for all $f \in \mathcal{L}_{sm}(\mathbb{E})$ for which the expectations exist.

A simple sufficient criterion for ‘ \prec_{sm} ’ ordering for an \mathbb{E} which is a discrete (countable) lattice is as follows.

Lemma 2.1. Let P_1 be a probability measure on a discrete-lattice ordered space $(\mathbb{E}, \mathcal{E}, \prec)$ and assume that, for incomparable points $x \neq y \in \mathbb{E}$, we have $P_1(x) \geq \alpha$ and $P_1(y) \geq \alpha$ for some $\alpha > 0$. Define a new probability measure, P_2 , on $(\mathbb{E}, \mathcal{E}, \prec)$ by

$$\begin{aligned} P_2(x) &= P_1(x) - \alpha, & P_2(x \vee y) &= P_1(x \vee y) + \alpha, \\ P_2(y) &= P_1(y) - \alpha, & P_2(x \wedge y) &= P_1(x \wedge y) + \alpha, \\ P_2(z) &= P_1(z), & z &\neq x, y, x \vee y, x \wedge y. \end{aligned} \tag{2.1}$$

Then $P_1 \prec_{sm} P_2$.

If some probability measure P_2 on $(\mathbb{E}, \mathcal{E}, \prec)$ can be obtained from P_1 by a finite sequence of transformations of the form (2.1), then $P_1 \prec_{sm} P_2$.

Proof. For any $f \in \mathcal{L}_{sm}(\mathbb{E})$, the supermodularity of f yields

$$\int_{\mathbb{E}} P_2(dz) f(z) - \int_{\mathbb{E}} P_1(dz) f(z) = \alpha(f(x \wedge y) + f(x \vee y) - f(x) - f(y)) \geq 0.$$

The second statement follows because ‘ \prec_{sm} ’ is transitive.

Remark 2.1. If, in Lemma 2.1, the state space \mathbb{E} is the set of all subsets of a finite set (which is a lattice under set inclusion) then the transformation described in (2.1) is what Li and Xu (2000) called a *pairwise g^+ transform* and our Lemma 2.1 specializes to Proposition 5.5 of Li and Xu (2000). For similar transforms see Mira (2001).

We shall use the following dependence orderings.

Definition 2.2. For arbitrary random elements X and Y of $(\mathbb{E}, \mathcal{E}, \prec)$, we say that X and Y are concordant weakly stochastically ordered, and write $X \prec_{cc-wk} Y$, if

$$P(X \in A) \leq P(Y \in A)$$

for all $A \in \mathcal{I}_p(\mathbb{E})$ and all $A \in \mathcal{D}_p(\mathbb{E})$.

Note that $X \prec_{cc-wk} Y$ is equivalent to the requirement that $X \leq_{\mathcal{J}_{wk}} Y$ and $X \geq_{\mathcal{J}_{wk}^*} Y$ both hold, in the notation used by Massey (1987).

Another ordering weaker than ‘ \prec_{sm} ’ can be defined on product spaces.

Definition 2.3. Let $X^{(n)}$ and $Y^{(n)}$ be random elements of $\mathbb{E}^{(n)}$. We say that $X^{(n)}$ and $Y^{(n)}$ are concordant stochastically ordered (written as $X^{(n)} \prec_{cc}^n Y^{(n)}$ or $Y^{(n)} \succ_{cc}^n X^{(n)}$) if

$$E \left[\prod_{i=1}^n f_i(X_i) \right] \leq E \left[\prod_{i=1}^n f_i(Y_i) \right] \tag{2.2}$$

for all $f_i \in \mathcal{I}_+^*(\mathbb{E}_i)$ and all $f_i \in \mathcal{D}_+^*(\mathbb{E}_i)$, $i = 1, \dots, n$, such that the expectations exist.

For $\mathbb{E}^{(n)} = \mathbb{R}^n$, $X^{(n)} \prec_{cc}^n Y^{(n)}$ is equivalent to the requirement that $X \leq_{uo} Y$ and $X \geq_{lo} Y$ both hold, in the notation used by Müller and Stoyan (2002). Note that concordant ordering is defined only for product spaces, while concordant weak ordering is defined on arbitrary partially ordered spaces. Definitions of concordant orderings do not require a lattice structure on the space, but this is required for supermodular ordering.

On product lattices, Lemma 2.1 reduces to the concept of concordance order investigated by Joe and can be illustrated as follows (see Joe (1990, Example 2.2)). Let $(\mathbb{E}, \mathcal{E}, \prec)$ be a finite lattice and let P_1 and P_2 be probability measures on \mathbb{E}^2 . Let $x, y \in \mathbb{E}^2$ be two points which are not comparable under coordinatewise ordering. If P_1 is obtained from P_2 by adding the probability mass $\alpha > 0$ at points $x \wedge y$ and $x \vee y$ and by subtracting the probability mass $\alpha > 0$ at points x and y , then $P_1 \prec_{cc}^2 P_2$. We can drop the assumption that \mathbb{E} is a lattice, and obtain the following.

Corollary 2.1. Let P_1 be a probability measure on \mathbb{E}^2 , where $(\mathbb{E}, \mathcal{E}, \prec)$ is a discrete, partially ordered space. Assume that, for $a, b, c, d \in \mathbb{E}$, we have $a \prec c$ and $b \prec d$ such that $(a, d) \in \mathbb{E}^2$ and $(c, b) \in \mathbb{E}^2$ are not comparable in the coordinatewise ordering on \mathbb{E}^2 . Assume that $P_1(a, d) \geq \alpha$ and $P_1(c, b) \geq \alpha$ for some $\alpha > 0$. (Throughout, we abbreviate $P_i(\{(x, y)\})$ to $P_i(x, y)$.)

Define a new probability measure, P_2 , on \mathbb{E}^2 by

$$\begin{aligned} P_2(a, d) &= P_1(a, d) - \alpha, & P_2(a, b) &= P_1(a, b) + \alpha, \\ P_2(c, b) &= P_1(c, b) - \alpha, & P_2(c, d) &= P_1(c, d) + \alpha, \\ P_2(x, y) &= P_1(x, y), & x \neq a, c, y \neq b, d. \end{aligned} \tag{2.3}$$

Then $P_1 \prec_{cc}^2 P_2$.

If some probability measure P_2 on \mathbb{E}^2 can be obtained from P_1 by a finite sequence of transformations of the form (2.3), then $P_1 \prec_{cc}^2 P_2$.

In the special case of product spaces $(\mathbb{E}^{(n)}, \prec^n)$, concordant weak ordering is characterized in the following way: $X^{(n)} \prec_{cc-wk}^n Y^{(n)}$ if and only if

$$P(X_1 \in A_1, \dots, X_n \in A_n) \leq P(Y_1 \in A_1, \dots, Y_n \in A_n)$$

for all $A_i \in \mathcal{I}_p(\mathbb{E}_i)$ and all $A_i \in \mathcal{D}_p(\mathbb{E}_i)$, $i = 1, \dots, n$. Since $\mathcal{I}_p(\mathbb{E}_i) \subset \mathcal{I}(\mathbb{E}_i)$ and $\mathcal{D}_p(\mathbb{E}_i) \subset \mathcal{D}(\mathbb{E}_i)$, $i = 1, \dots, n$, it is clear that ‘ \prec_{cc}^n ’ implies ‘ \prec_{cc-wk}^n ’. For products of linearly ordered sets, the above inclusion relations can be replaced by equalities and ‘ \prec_{cc}^n ’ and ‘ \prec_{cc-wk}^n ’ are therefore equivalent.

In the next example we show that ‘ \prec_{cc-wk}^n ’ does not imply ‘ \prec_{cc}^n ’ in general.

Example 2.1. If we consider $\mathbb{E} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ with the usual partial ordering, $(0, 0) < (0, 1) < (1, 1)$ and $(0, 0) < (1, 0) < (1, 1)$, and let $P_1((1, 1)) = P_1((0, 0)) = \frac{1}{2}$ and $P_2((0, 1)) = P_2((1, 0)) = \frac{1}{2}$, then, directly from the definition, for product measures on \mathbb{E}^2 we obtain $P_1 \times P_1 \prec_{\text{cc-wk}}^2 P_2 \times P_2$ but not $P_1 \times P_1 \prec_{\text{cc}}^2 P_2 \times P_2$.

The proof of the following lemma can be given using arguments similar to those proving Lindqvist (1988, Theorem 3.1).

Lemma 2.2. *The following statements are equivalent for random elements $X^{(n)}$ and $Y^{(n)}$ of $\mathbb{E}^{(n)}$.*

1. $X^{(n)} \prec_{\text{cc}}^n Y^{(n)}$.
2. $P(X_1 \in A_1, \dots, X_n \in A_n) \leq P(Y_1 \in A_1, \dots, Y_n \in A_n)$ for all $A_i \in \mathcal{I}(\mathbb{E}_i)$ and all $A_i \in \mathcal{D}(\mathbb{E}_i)$, $i = 1, \dots, n$.
3. $P(X_1 \in A_1, \dots, X_n \in A_n) \leq P(Y_1 \in A_1, \dots, Y_n \in A_n)$ for all closed sets $A_i \in \mathcal{I}(\mathbb{E}_i)$ and all closed sets $A_i \in \mathcal{D}(\mathbb{E}_i)$, $i = 1, \dots, n$.
4. $P(X_1 \in A_1, \dots, X_n \in A_n) \leq P(Y_1 \in A_1, \dots, Y_n \in A_n)$ for all compact generated sets $A_i \in \mathcal{I}(\mathbb{E}_i)$ (for a definition, see Lindqvist (1988)) and all compact generated sets $A_i \in \mathcal{D}(\mathbb{E}_i)$, $i = 1, \dots, n$.

If, in addition, the \mathbb{E}_i are normally ordered (see Lindqvist (1988)), then $X^{(n)} \prec_{\text{cc}}^n Y^{(n)}$ if and only if (2.2) holds for f_i which are continuous (in addition to the other properties required of the f_i).

The following example shows how product structures can be related to lattice nonproduct structures for concordance orderings.

Example 2.2. Suppose that the state space \mathbb{E} is the set of all subsets of a finite set $\{e_1, \dots, e_n\}$. Then \mathbb{E} is a lattice under ordinary set inclusion. Let P_1 and P_2 be probability measures on \mathbb{E} such that $P_1 \prec_{\text{cc-wk}} P_2$. On \mathbb{E} it is natural to define the variable $\delta(A) = (1_A(e_1), \dots, 1_A(e_n))$, $A \in \mathbb{E}$. This is a random vector with n zero/one-valued coordinates indicating elements belonging to A (and 1_A denotes the indicator function). Denote by $P_1^\delta = P_1 \delta^{-1}$ and $P_2^\delta = P_2 \delta^{-1}$ the distributions of δ on $\{0, 1\}^n$ with respect to P_1 and P_2 , respectively. On $\{0, 1\}^n$ we consider the usual coordinatewise ordering ' \leq '. It then follows, by definition, that $P_1 \prec_{\text{cc-wk}} P_2$ if and only if $P_1^\delta \leq_{\text{cc}}^n P_2^\delta$. The ordering $P_1 \prec_{\text{cc-wk}} P_2$ for two probability measures on \mathbb{E} can be defined by two orderings defined in Li and Xu (2000, Definition 5.1), that is, by majorization of P_1 by P_2 from the roots and by majorization of P_1 by P_2 from the leaves.

We shall now proceed to state, without proof (the usual arguments can be used), the standard properties of dependence orderings for ' \prec_{cc}^n ' (for the case of real-valued random vectors, see Müller and Stoyan (2002, Theorem 3.8.7)). We omit the ' n ' in the notation ' \prec_{cc}^n ' in this statement.

Proposition 2.1. *Let $X^{(n)}$ and $Y^{(n)}$ be random elements of $\mathbb{E}^{(n)}$. Then the following properties hold.*

- (P1) (Bivariate concordance.) $X^{(n)} \prec_{\text{cc}} Y^{(n)}$ implies that, for any pair of indices (i, j) , $1 \leq i < j \leq n$, the bivariate marginal vectors satisfy $(X_i, X_j) \prec_{\text{cc}}^2 (Y_i, Y_j)$ and, therefore, that $\text{cov}(f(X_i), g(X_j)) \leq \text{cov}(f(Y_i), g(Y_j))$ for each $f \in \mathcal{I}_+^*(\mathbb{E}_i)$ and $g \in \mathcal{I}_+^*(\mathbb{E}_j)$.
- (P2) (Transitivity.) $X^{(n)} \prec_{\text{cc}} Y^{(n)}$ and $Y^{(n)} \prec_{\text{cc}} Z^{(n)}$ imply that $X^{(n)} \prec_{\text{cc}} Z^{(n)}$.

(P3) (Reflexivity.) $X^{(n)} \prec_{cc} X^{(n)}$.

(P4) (Antisymmetry.) $X^{(n)} \prec_{cc} Y^{(n)}$ and $Y^{(n)} \prec_{cc} X^{(n)}$ imply that the distributions of $X^{(n)}$ and $Y^{(n)}$ are equal.

(P5) (Weak convergence.) Let $(X_k^{(n)}, Y_k^{(n)}, k \in \mathbb{N})$ be a sequence of random elements of $\mathbb{E}^{(n)}$. If $\mathbb{E}^{(n)}$ is normally ordered then $X_k^{(n)} \xrightarrow{w}_{k \rightarrow \infty} X^{(n)}$ and $Y_k^{(n)} \xrightarrow{w}_{k \rightarrow \infty} Y^{(n)}$ imply that $X^{(n)} \prec_{cc} Y^{(n)}$ if $X_k^{(n)} \prec_{cc} Y_k^{(n)}$ for all $k \in \mathbb{N}$, where ' \xrightarrow{w} ' denotes weak convergence.

(P6) (Invariance with respect to permutation of indices.) $X^{(n)} \prec_{cc} Y^{(n)}$ implies that

$$X^{\pi(n)} = (X_{\pi(1)}, \dots, X_{\pi(n)}) \prec_{cc} (Y_{\pi(1)}, \dots, Y_{\pi(n)}) = Y^{\pi(n)}$$

for all permutations π of $\{1, 2, \dots, n\}$.

(P7) (Invariance with respect to monotone transforms.) $X^{(n)} \prec_{cc} Y^{(n)}$ implies that

$$(f_1(X_1), \dots, f_n(X_n)) \prec_{cc} (f_1(Y_1), \dots, f_n(Y_n))$$

for all increasing and for all decreasing functions $f_i: \mathbb{E}_i \rightarrow \mathbb{E}_i, i = 1, \dots, n$.

(P8) (Closure under marginalization.) If $X^{(n)} \prec_{cc} Y^{(n)}$ then, for any marginal vectors $X_M = (X_i: i \in M)$ and $Y_M = (Y_i: i \in M)$ with $M \subseteq \{1, \dots, n\}$, it follows that $X_M \prec_{cc}^{|M|} Y_M$.

(P9) (Equality of one-dimensional marginals.) $X^{(n)} \prec_{cc} Y^{(n)}$ implies that, for all $i = 1, \dots, n$, the distributions of the respective coordinates X_i and Y_i are equal.

(P10) (Closure under identical concatenation.) $X^{(n)} \prec_{cc} Y^{(n)}$ implies that, for all $K, L \subseteq \{1, \dots, n\}, (X_K, X_L) \prec_{cc}^{|K|+|L|} (Y_K, Y_L)$. Here $K \cap L \neq \emptyset$ is allowed.

(P11) (Closure under independent concatenation.) Let $X_j^{(n)} \prec_{cc} Y_j^{(n)}$ for $j = 1, 2$ and assume both that $X_1^{(n)}$ is independent of $X_2^{(n)}$ and that $Y_1^{(n)}$ is independent of $Y_2^{(n)}$. Then $(X_1^{(n)}, X_2^{(n)}) \prec_{cc} (Y_1^{(n)}, Y_2^{(n)})$.

3. Main results

We shall use the following definition for stochastic processes.

Definition 3.1. Let $T \subseteq \mathbb{R}$ be an index set for stochastic processes $X = (X_t: t \in T)$ and $Y = (Y_t: t \in T)$ with $X_t, Y_t: (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{E}, \mathcal{E}, <), t \in T$. We say that X and Y are concordant stochastically ordered (and write $X \prec_{cc} Y$) if, for all $n \geq 2$ and all $t_1 < t_2 < \dots < t_n$, we have

$$(X_{t_1}, \dots, X_{t_n}) \prec_{cc}^n (Y_{t_1}, \dots, Y_{t_n})$$

on \mathbb{E}^n . If, in addition, E is a lattice, we say that X and Y are supermodular stochastically ordered (and write $X \prec_{sm} Y$) if, for all $n \geq 2$ and all $t_1 < t_2 < \dots < t_n$, we have

$$(X_{t_1}, \dots, X_{t_n}) \prec_{sm}^n (Y_{t_1}, \dots, Y_{t_n})$$

on \mathbb{E}^n .

3.1. Discrete-time Markov processes

We say that a stochastic kernel $K : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1]$ is stochastically monotone if

$$\int f(x)K(s, dx)$$

is increasing in s for each $f \in \mathcal{I}^*(\mathbb{E})$. It is known (see, e.g. Müller and Stoyan (2002, Section 5.2)) that a stochastic kernel K is stochastically monotone if and only if $K(x, \cdot) \prec_{st} K(y, \cdot)$ for all $x < y$. An equivalent condition is that $\mu K \prec_{st} \nu K$ for all μ and ν such that $\mu \prec_{st} \nu$, where μK denotes the measure defined by $\mu K(A) = \int K(s, A)\mu(ds)$, $A \in \mathcal{E}$. It is worth mentioning that, for $\mathbb{E} = \mathbb{N}$, using the traditional notation $P^X = [p^X(i, j)]_{i, j \in \mathbb{N}}$ for the transition matrix of X (that is, $p^X(i, j) := K^X(i, \{j\})$), stochastic monotonicity can be expressed in a very simple form: we say that P^X is stochastically monotone if (see Keilson and Kester (1977))

$$T^{-1}P^X T(i, j) \geq 0, \quad i, j \in \mathbb{N}, \tag{3.1}$$

where T is the lower-triangular matrix with 0s above the main diagonal and 1s elsewhere.

Let $X = (X_t : t \in \mathbb{Z})$ and $Y = (Y_t : t \in \mathbb{Z})$, $X_t, Y_t : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{E}, \mathcal{E}, <)$, be discrete-time, stationary, homogeneous Markov processes. Assume that π is an invariant (stationary) one-dimensional marginal distribution common to both X and Y , and denote the one-step transition kernels for X and Y by $K^X : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1]$ and $K^Y : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1]$, respectively. Denote the respective transition kernels for the time reversed processes, \tilde{X} and \tilde{Y} by \tilde{K}^X and \tilde{K}^Y . Note that \tilde{K}^X and \tilde{K}^Y can be seen as adjoint operators of K^X and K^Y on the space of square integrable functions $L^2(\mathbb{E}, \pi)$, respectively. Indeed, for $f, g \in L^2(\mathbb{E}, \pi)$, $K^X f$ is defined by

$$K^X f(x) = \int f(y)K^X(x, dy)$$

and $(f, g)_\pi = \int f(x)g(x)\pi(dx)$. Then $(\tilde{K}^X f, g)_\pi = (f, K^X g)_\pi$ for all $f, g \in L^2(\mathbb{E}, \pi)$ and, thus, $\tilde{K}^X = (K^X)^*$, where A^* denotes the adjoint of the operator A . Similarly, $\tilde{K}^Y = (K^Y)^*$.

It will be convenient to state the following stochastic monotonicity assumption.

Assumption 3.1. (STM1.) *Either K^Y and \tilde{K}^X are stochastically monotone or K^X and \tilde{K}^Y are stochastically monotone.*

Theorem 3.1. *For the above-defined stationary Markov processes X and Y with common invariant distribution π , under STM1 the following relations are equivalent.*

- (i) $X \prec_{cc} Y$.
- (ii) $(X_0, X_1) \prec_{cc}^2 (Y_0, Y_1)$.
- (iii) $(f, K^X g)_\pi \leq (f, K^Y g)_\pi$ for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$.
- (iv) $(f, \tilde{K}^X g)_\pi \leq (f, \tilde{K}^Y g)_\pi$ for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$.

Proof. We consider the case in which K^X and \tilde{K}^Y are monotone. To prove that (ii) implies (i) we proceed by induction on n . For $n = 1$, (ii) is the required relation. Suppose that, for $n - 1 > 0$, we have $(X_0, \dots, X_{n-1}) \prec_{cc} (Y_0, \dots, Y_{n-1})$. Then, for $f_i \in \mathcal{I}_+^*(\mathbb{E})$ and $f_i \in \mathcal{D}_+^*(\mathbb{E})$, $i = 0, \dots, n$, denoting the joint distribution of (X_0, \dots, X_n) by P^{X_0, \dots, X_n} (and

using similar notation for conditional distributions), we have

$$\begin{aligned}
 & \mathbb{E}_\pi \left[\prod_{i=0}^n f_i(X_i) \right] \\
 &= \int_{\mathbb{E}^n} \mathbf{P}^{X_0, \dots, X_{n-1}}(d(x_0, \dots, x_{n-1})) \prod_{i=0}^{n-1} f_i(x_i) \int_{\mathbb{E}} \mathbf{P}^{X_n | X_{n-1}=x_{n-1}}(dx_n) f_n(x_n) \\
 &= \int_{\mathbb{E}^n} \mathbf{P}^{X_0, \dots, X_{n-1}}(d(x_0, \dots, x_{n-1})) \prod_{i=0}^{n-1} f_i(x_i) \left[\int_{\mathbb{E}} K^X(x_{n-1}, dx_n) f_n(x_n) \right] \\
 &\leq \int_{\mathbb{E}^n} \mathbf{P}^{Y_0, \dots, Y_{n-1}}(d(x_0, \dots, x_{n-1})) \prod_{i=0}^{n-1} f_i(x_i) \int_{\mathbb{E}} K^X(x_{n-1}, dx_n) f_n(x_n) \\
 &= \int_{\mathbb{E}} \mathbf{P}^{Y_{n-1}}(dx_{n-1}) \int_{\mathbb{E}^{n-1}} \mathbf{P}^{Y_0, \dots, Y_{n-2} | Y_{n-1}=x_{n-1}}(d(x_0, \dots, x_{n-2})) \\
 &\quad \times \prod_{i=0}^{n-1} f_i(x_i) \int_{\mathbb{E}} K^X(x_{n-1}, dx_n) f_n(x_n) \\
 &= \underbrace{\int_{\mathbb{E}} \pi(dx_{n-1}) \int_{\mathbb{E}} K^X(x_{n-1}, dx_n)}_{\int_{\mathbb{E}^2} \mathbf{P}^{X_{n-1}, X_n}(d(x_{n-1}, x_n))} \prod_{i=n-1}^n f_i(x_i) \\
 &\quad \times \left[\int_{\mathbb{E}^{n-1}} \mathbf{P}^{Y_0, \dots, Y_{n-2} | Y_{n-1}=x_{n-1}}(d(x_0, \dots, x_{n-2})) \prod_{i=0}^{n-2} f_i(x_i) \right] \\
 &\leq \int_{\mathbb{E}^2} \mathbf{P}^{Y_{n-1}, Y_n}(d(x_{n-1}, x_n)) \prod_{i=n-1}^n f_i(x_i) \\
 &\quad \times \int_{\mathbb{E}^{n-1}} \mathbf{P}^{Y_0, \dots, Y_{n-2} | Y_{n-1}=x_{n-1}}(d(x_0, \dots, x_{n-2})) \prod_{i=0}^{n-2} f_i(x_i) \\
 &= \mathbb{E}_\pi \left[\prod_{i=0}^n f_i(Y_i) \right].
 \end{aligned}$$

Here the first inequality follows from the induction hypotheses and the second inequality follows from stationarity and (ii). Note that, because K^X is monotone, the term in square brackets on the third line is increasing in x_{n-1} for $f_i \in \mathcal{I}_+^*(\mathbb{E})$ and decreasing in x_{n-1} for $f_i \in \mathcal{D}_+^*(\mathbb{E})$. Because K^Y is monotone, the same is true of the term in square brackets on the eighth line.

Theorem 3.2. *Assume that the state space \mathbb{E} is linearly ordered. For the above-defined stationary Markov processes X and Y with common invariant distribution, under STM1 we have $X \prec_{\text{sm}} Y$ if and only if $(X_0, X_1) \prec_{\text{sm}}^2 (Y_0, Y_1)$.*

Proof. The proof follows by analogy to the case $\mathbb{E} = \mathbb{R}$; see Theorem 3.2 of Hu and Pan (2000).

Now consider two homogeneous Markov chains, X and Y , on a countable state space \mathbb{E} , with common invariant distribution π . Denote by $P^X = [p^X(i, j)]$ and $P^Y = [p^Y(i, j)]$ the

respective transition matrices of X and Y . By standard approximation arguments, we obtain the following corollary to Theorem 3.1.

Corollary 3.1. *For the above-defined stationary Markov processes X and Y with common invariant distribution π and countable state space \mathbb{E} , under STM1 the conditions of Theorem 3.1 are equivalent to*

$$\sum_{i \in F, j \in G} \pi(i)p^X(i, j) \leq \sum_{i \in F, j \in G} \pi(i)p^Y(i, j),$$

for all $F, G \in \mathcal{I}(\mathbb{E})$ and all $F, G \in \mathcal{D}(\mathbb{E})$.

3.2. Continuous-time Markov processes

Let $X = (X_t : t \in \mathbb{R})$ and $Y = (Y_t : t \in \mathbb{R})$, $X_t, Y_t : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{E}, \mathcal{E}, <)$, be stationary, homogeneous Markov processes. Denote their respective families of transition kernels by

$$\mathbb{K}^X = (K_t^X : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1], t \geq 0) \quad \text{and} \quad \mathbb{K}^Y = (K_t^Y : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1], t \geq 0),$$

and the respective families of transition kernels for the stationary time-reversed processes \tilde{X} and \tilde{Y} by

$$\tilde{\mathbb{K}}^X = (\tilde{K}_t^X : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1], t \geq 0) \quad \text{and} \quad \tilde{\mathbb{K}}^Y = (\tilde{K}_t^Y : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1], t \geq 0).$$

Assume that π is an invariant distribution common to both \mathbb{K}^X and \mathbb{K}^Y ; that is,

$$\int K_t^X(x, dy)\pi(dx) = \int K_t^Y(x, dy)\pi(dx) = \pi(dy) \quad \text{for all } t > 0.$$

We say that \mathbb{K}^X and \mathbb{K}^Y are stochastically monotone if, for each $t > 0$, K_t^X and, respectively, K_t^Y are stochastically monotone as defined in the previous section. If \mathbb{E} is countable and $Q^X = [Q^X(x, y)]$ and $Q^Y = [Q^Y(x, y)]$ denote the intensity matrices (infinitesimal generators) of the corresponding chains X and Y , then the following condition, due to Massey (1987), is useful: if Q^X is bounded and conservative, then \mathbb{K}^X is stochastically monotone if and only if

$$\sum_{y \in F} Q^X(x_1, y) \leq \sum_{y \in F} Q^X(x_2, y)$$

for all $F \in \mathcal{I}(\mathbb{E})$ and all x_1 and x_2 such that $x_1 < x_2$ with $x_1 \in F$ or $x_2 \notin F$. An analogous condition for arbitrary continuous-time Markov jump processes (also for unbounded generators) was given by Chen (2004, Theorem 5.47). For diffusions on $\mathbb{E} = \mathbb{R}^n$, necessary and sufficient conditions for stochastic monotonicity were given by Chen and Wang (1993). It is worth mentioning that, for $\mathbb{E} = \mathbb{N}$, we say that $Q^X = [Q^X(i, j)]_{i, j \in \mathbb{N}}$ is stochastically monotone if $T^{-1}Q^XT(i, j) \geq 0$ for all $i \neq j$ (cf. (3.1)).

It is convenient to state the following assumption.

Assumption 3.2. (STM2.) *Either \mathbb{K}^Y and $\tilde{\mathbb{K}}^X$ are stochastically monotone or \mathbb{K}^X and $\tilde{\mathbb{K}}^Y$ are stochastically monotone.*

Utilizing arguments similar to those in the proof of Theorem 3.1, we obtain the following corollary.

Corollary 3.2. *For the above-defined stationary Markov processes X and Y with common invariant distribution, under STM2 we have $X <_{cc} Y$ if and only if $(X_0, X_t) <_{cc}^2 (Y_0, Y_t)$ for all $t > 0$.*

In order to relate the above conditions to generators, we need additional regularity assumptions. Assume that X and Y are Feller processes on a normally ordered space $(\mathbb{E}, \mathcal{E}, <)$ and have corresponding semigroups

$$T^X = (T_t^X : \mathcal{C}_b(\mathbb{E}) \rightarrow \mathcal{C}_b(\mathbb{E}), t \geq 0) \quad \text{and} \quad T^Y = (T_t^Y : \mathcal{C}_b(\mathbb{E}) \rightarrow \mathcal{C}_b(\mathbb{E}), t \geq 0)$$

on $\mathcal{C}_b(\mathbb{E})$ (the space of real-valued, bounded, continuous functions) possessing strong generators A^X , with domain $D_{A^X} \subseteq \mathcal{C}_b(\mathbb{E})$, and, respectively, A^Y , with domain $D_{A^Y} \subseteq \mathcal{C}_b(\mathbb{E})$. We assume that $T_0^X = T_0^Y = I$ (the identity operator). The corresponding families of transition kernels of X and Y are respectively

$$\mathbb{K}^X = (K_t^X : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1], t \geq 0) \quad \text{and} \quad \mathbb{K}^Y = (K_t^Y : \mathbb{E} \times \mathcal{E} \rightarrow [0, 1], t \geq 0)$$

(such that $(T_t^X f)(x) = \int_{\mathbb{E}} K_t^X(x, dy) f(y)$).

Lemma 3.1. *Suppose that STM2 holds and that $D_{A^X} \cap D_{A^Y}$ is dense in $\mathcal{C}_b(\mathbb{E})$. Then $X \prec_{cc} Y$ implies that*

$$\int_{\mathbb{E}} f(x) A^X g(x) \pi(dx) \leq \int_{\mathbb{E}} f(x) A^Y g(x) \pi(dx) \tag{3.2}$$

for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$ such that $f \in \mathcal{C}_b(\mathbb{E})$ and $g \in D_{A^X} \cap D_{A^Y}$.

Proof. The relation $X \prec_{cc} Y$ implies that $(X_0, X_t) \prec_{cc}^2 (Y_0, Y_t)$ for all $t > 0$. Therefore, for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$ such that $f \in \mathcal{C}_b(\mathbb{E})$ and $g \in D_{A^X} \cap D_{A^Y}$, we have

$$\int_{\mathbb{E}} \pi(dx) f(x) \underbrace{\int_{\mathbb{E}} K_t^X(x, dy) g(y)}_{(T_t^X g)(x)} \leq \int_{\mathbb{E}} \pi(dx) f(x) \underbrace{\int_{\mathbb{E}} K_t^Y(x, dy) g(y)}_{(T_t^Y g)(x)}.$$

Hence, for all $h > 0$,

$$\int_{\mathbb{E}} \pi(dx) f(x) \frac{1}{h} [T_h^X - I]g(x) \leq \int_{\mathbb{E}} \pi(dx) f(x) \frac{1}{h} [T_h^Y - I]g(x). \tag{3.3}$$

For any $\varepsilon > 0$, there exists some sufficiently small $h(\varepsilon)$ such that, for all $h < h(\varepsilon)$,

$$\left\| \frac{1}{h} [T_h^X - I]g - A^X g \right\| \leq \varepsilon$$

(with the uniform norm). It follows that

$$\begin{aligned} & \left| \int_{\mathbb{E}} \pi(dx) f(x) \frac{1}{h} [T_h^X - I]g(x) - \int_{\mathbb{E}} \pi(dx) f(x) (A^X g)(x) \right| \\ & \leq \int_{\mathbb{E}} \pi(dx) \|f\| \left\| \frac{1}{h} [T_h^X - I]g - A^X g \right\| \\ & \leq \|f\| \varepsilon. \end{aligned}$$

Therefore, we can let $h \rightarrow 0$ in (3.3) and remove it from each integral, to obtain (3.2).

We will now show that (3.2) is essentially necessary and sufficient for concordance ordering. For simplicity, we restrict our considerations to countable state spaces and bounded generators

$$Q^X = [Q^X(x, y)] \quad \text{and} \quad Q^Y = [Q^Y(x, y)].$$

The argument can be applied with obvious modifications to arbitrary state spaces and bounded generators. For the time-reversed processes we use the corresponding notation \overleftarrow{Q}^X and \overleftarrow{Q}^Y . We assume that $\mathcal{E} = 2^{\mathbb{E}}$.

Theorem 3.3. *Suppose that $(\mathbb{E}, \mathcal{E}, <)$ is countable and that the above-defined stationary chains X and Y have bounded intensity matrices Q^X and Q^Y , respectively. Then, under STM2, the following properties are equivalent.*

- (i) $X <_{cc} Y$.
- (ii) $(X_0, X_t) <_{cc}^2 (Y_0, Y_t)$ for all $t > 0$.
- (iii) $(f, T_t^X g)_\pi \leq (f, T_t^Y g)_\pi$ for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$, for all $t > 0$.
- (iv) $(f, Q^X g)_\pi \leq (f, Q^Y g)_\pi$ for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$.
- (v) $(f, \overleftarrow{Q}^X g)_\pi \leq (f, \overleftarrow{Q}^Y g)_\pi$ for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and all $f, g \in \mathcal{D}_+^*(\mathbb{E})$.
- (vi) For all $F, G \in \mathcal{I}(\mathbb{E})$ and all $F, G \in \mathcal{D}(\mathbb{E})$,

$$\sum_{x \in F, y \in G} \pi(x) Q^X(x, y) \leq \sum_{i \in F, j \in G} \pi(x) Q^Y(x, y).$$

Proof. That (i) and (ii) are equivalent follows from the statement of Corollary 3.2. That (iii) implies (iv) follows from the statement of Lemma 3.1. The equivalence of (v) and (iv) follows from the fact that we can interchange the roles of the processes and their time reversals. The equivalence of (vi) and (iv) follows by standard approximation arguments.

In order to finish the proof we need to show that (iv) implies (iii). We consider the case in which $f, g \in \mathcal{I}_+^*(\mathbb{E})$ and \mathbb{K}^X and \mathbb{K}^Y are stochastically monotone, and shall utilize the following representation (which follows from assuming bounded intensity matrices):

$$K_t^X(x, \{y\}) = \lim_{n \rightarrow \infty} \left(I + \frac{t}{n} Q^X \right)^n(x, y), \quad x, y \in \mathbb{E}.$$

We first show, by induction, that for sufficiently small $\eta > 0$ we have

$$\begin{aligned} & \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} (I + \eta Q^X)^n(x, y) g(y) \\ & \leq \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} (I + \eta Q^Y)^n(x, y) g(y). \end{aligned} \tag{3.4}$$

For $n = 1$, this is immediate from (iv), so assume that we have proved (3.4) for all $m \leq n$. Then, into (3.4), substitute

$$g(x) := \sum_{y \in \mathbb{E}} (I + \eta Q^X)(x, y) g(y) = (I + \eta Q^X)g(x),$$

which is an increasing function of x because $I + \eta Q^X$ is a monotone operator as defined in Massey (1987) (by induction $(I + \eta Q^X)^n(g)$ is then also an increasing function). We obtain

$$\begin{aligned}
 & \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} (I + \eta Q^X)^{n+1}(x, y) g(y) \\
 &= \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} (I + \eta Q^X)^n(x, y) \left[\sum_{z \in \mathbb{E}} (I + \eta Q^X)(y, z) g(z) \right] \\
 &\leq \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} (1 + \eta Q^Y)^n(x, y) \left[\sum_{z \in \mathbb{E}} (I + \eta Q^X)(y, z) g(z) \right] \\
 &= \sum_{y \in \mathbb{E}} \pi(y) \left[\sum_{z \in \mathbb{E}} (I + \eta Q^X)(y, z) g(z) \right] \sum_{x \in \mathbb{E}} \frac{\pi(x)}{\pi(y)} f(x) (I + \eta Q^Y)^n(x, y) \\
 &= \sum_{y \in \mathbb{E}} \pi(y) \sum_{x \in \mathbb{E}} (I + \eta \tilde{Q}^Y)^n(y, x) f(x) \left[\sum_{z \in \mathbb{E}} (I + \eta Q^X)(y, z) g(z) \right] \\
 &\leq \sum_{y \in \mathbb{E}} \pi(y) \sum_{x \in \mathbb{E}} (I + \eta \tilde{Q}^Y)^n(y, x) f(x) \left[\sum_{z \in \mathbb{E}} (I + \eta Q^Y)(y, z) g(z) \right] \\
 &= \sum_{y \in \mathbb{E}} \pi(y) \sum_{x \in \mathbb{E}} \frac{\pi(x)}{\pi(y)} (I + \eta Q^Y)^n(x, y) f(x) \left[\sum_{z \in \mathbb{E}} (I + \eta Q^Y)(y, z) g(z) \right] \\
 &= \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{z \in \mathbb{E}} \sum_{y \in \mathbb{E}} (I + \eta Q^Y)^n(x, y) (I + \eta Q^Y)(y, z) g(z) \\
 &= \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} (I + \eta Q^Y)^{n+1}(x, y) g(y).
 \end{aligned}$$

Here the first inequality follows from the induction hypothesis (3.4) and the monotonicity of $(I + \eta Q^X)g(x)$. The second inequality is obtained by applying (iv) with the substitution $f(x) \rightarrow \sum_{y \in \mathbb{E}} (I + \eta \tilde{Q}^Y)^n(x, y) f(y)$, which function is increasing and nonnegative for sufficiently small η . The third equality follows, by direct computation, from

$$\begin{aligned}
 \pi(x) (I + \eta Q^Y)^n(x, y) &= \sum_{k=0}^n \binom{n}{k} \eta^k \pi(x) (Q^Y)^k(x, y), \\
 \pi(x) (Q^Y)^n(x, y) &= \sum_{z \in \mathbb{E}} \tilde{Q}^Y(z, x) \pi(z) (Q^Y)^{n-1}(z, y).
 \end{aligned}$$

(Note that any operator raised to the power 0 is equal to the identity operator.)

Now let $\eta := t/n$ in (3.4); then

$$\begin{aligned}
 & \lim_{n \rightarrow \infty} \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} \left(I + \frac{t}{n} Q^X \right)^n(x, y) g(y) \\
 &\leq \lim_{n \rightarrow \infty} \sum_{x \in \mathbb{E}} \pi(x) f(x) \sum_{y \in \mathbb{E}} \left(I + \frac{t}{n} Q^Y \right)^n(x, y) g(y).
 \end{aligned}$$

Since $\pi(x)f(x)$, $x \in \mathbb{E}$, is a counting density of a finite measure on \mathbb{E} and the sequences of functions

$$\sum_{y \in \mathbb{E}} \left(I + \frac{t}{n} Q^X \right)^n(x, y)g(y), \quad n \in \mathbb{N}, \quad \sum_{y \in \mathbb{E}} \left(I + \frac{t}{n} Q^Y \right)^n(x, y)g(y), \quad n \in \mathbb{N},$$

are bounded and respectively uniformly convergent to the bounded functions

$$\sum_{y \in \mathbb{E}} K_t^X(x, \{y\})g(y), \quad n \in \mathbb{N}, \quad \sum_{y \in \mathbb{E}} K_t^Y(x, \{y\})g(y), \quad n \in \mathbb{N},$$

by interchanging limit and integration in (iv) we obtain (iii) in the form

$$\int_{\mathbb{E}^2} P^{X_0, X_1}(d(u, v))f(u)g(v) \leq \int_{\mathbb{E}^2} P^{Y_0, Y_1}(d(u, v))f(u)g(v),$$

for all $f, g \in \mathcal{I}_+^*(\mathbb{E})$. For decreasing functions we repeat the above argument.

3.3. Monotonicity assumptions

A natural question is whether the monotonicity assumptions on the kernels arising in the above theorems are necessary. The following example indicates that they are not necessary in order to obtain concordance ordering for finite sequences.

Example 3.1. Consider two homogeneous Markov chains, X and Y , on a finite state space \mathbb{E} , with common invariant distribution π . Denote by $P^X = [p^X(i, j)]_{i, j \in \mathbb{E}}$ and $P^Y = [p^Y(i, j)]_{i, j \in \mathbb{E}}$ the transition matrices of X and Y , respectively. Assume that, for $a, b, c, d \in \mathbb{E}$, we have $a < c$ and $b < d$ (that is, $(a, d) \in \mathbb{E}^2$ and $(c, b) \in \mathbb{E}^2$ are not comparable) and that, for $P^{X_0, X_1}((i, j)) := \pi(i)p^X(i, j)$ and $P^{Y_0, Y_1}((i, j)) := \pi(i)p^Y(i, j)$, defined for all $(i, j) \in \mathbb{E}^2$, we have $P^{X_0, X_1}((a, d)) \geq \alpha$ and $P^{X_0, X_1}((c, b)) \geq \alpha$, for some $\alpha > 0$.

Moreover, assume that P^{Y_0, Y_1} and P^{X_0, X_1} are related by

$$\begin{aligned} P^{Y_0, Y_1}((a, b)) &= P^{X_0, X_1}((a, b)) + \alpha, & P^{Y_0, Y_1}((c, d)) &= P^{X_0, X_1}((c, d)) + \alpha, \\ P^{Y_0, Y_1}((a, d)) &= P^{X_0, X_1}((a, d)) - \alpha, & P^{Y_0, Y_1}((c, b)) &= P^{X_0, X_1}((c, b)) - \alpha, \\ P^{Y_0, Y_1}((u, v)) &= P^{X_0, X_1}((u, v)) & \text{for all other } (u, v) &\in \mathbb{E}^2 \end{aligned}$$

(see Corollary 2.1). Then the one-dimensional marginal distribution of both (X_0, X_1) and (Y_0, Y_1) is π , and the transition matrix $P^Y = [p^Y(i, j)]_{i, j \in \mathbb{E}}$ is related to $P^X = [p^X(i, j)]_{i, j \in \mathbb{E}}$ as follows:

$$\begin{aligned} p^Y(a, d) &= p^X(a, d) - \frac{\alpha}{\pi(a)}, & p^Y(c, b) &= p^X(c, b) - \frac{\alpha}{\pi(c)}, \\ p^Y(a, b) &= p^X(a, b) + \frac{\alpha}{\pi(a)}, & p^Y(c, d) &= p^X(c, d) + \frac{\alpha}{\pi(c)}, \\ p^Y(u, v) &= p^X(u, v) & \text{for all other } (u, v) &\in \mathbb{E}^2. \end{aligned}$$

If $\mathbb{E} = \{1, 2, 3, 4\}$, ‘ $<$ ’ is ‘ \leq ’ (natural ordering), $(a, d) = (2, 4)$, $(c, b) = (4, 2)$, and $\alpha = \frac{1}{12}$, then the rows $P^X(i, \cdot)$, $i = 1, 2, 3, 4$, of the transition matrix

$$P^X = \frac{1}{36} \begin{pmatrix} 18 & 0 & 18 & 0 \\ 6 & 12 & 6 & 12 \\ 7 & 11 & 7 & 11 \\ 5 & 13 & 5 & 13 \end{pmatrix},$$

treated as distributions on \mathbb{E} , satisfy

$$P^X(1, \cdot) \leq_{st} P^X(3, \cdot) \leq_{st} P^X(2, \cdot) \leq_{st} P^X(4, \cdot);$$

in particular, $P^X(2, \cdot)$ is not equal to $P^X(3, \cdot)$, so P^X is not stochastically monotone.

Because P^X is doubly stochastic, its unique invariant vector is $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and the time reversal, \tilde{X} , of X has transition matrix $\tilde{P}^X = [\tilde{p}^X(i, j)]_{i,j \in \mathbb{E}}$ of the form

$$\tilde{P}^X = (P^X)^\top = \frac{1}{36} \begin{pmatrix} 18 & 6 & 7 & 5 \\ 0 & 12 & 11 & 13 \\ 18 & 6 & 7 & 5 \\ 0 & 12 & 11 & 13 \end{pmatrix},$$

where ${}^\top$ denotes transpose. The rows $\tilde{P}^X(i, \cdot)$, $i = 1, 2, 3, 4$, of this matrix, treated as distributions on \mathbb{E} , satisfy

$$\tilde{P}^X(1, \cdot) = \tilde{P}^X(3, \cdot) \leq_{st} \tilde{P}^X(2, \cdot) = \tilde{P}^X(4, \cdot);$$

in particular, $\tilde{P}^X(2, \cdot)$ is not equal to $\tilde{P}^X(3, \cdot)$, so \tilde{P}^X is not stochastically monotone. The stationary two-dimensional distribution of (X_0, X_1) is given by

$$[P^{X_0, X_1}((i, j))]_{i,j=1,\dots,4} = \frac{1}{144} \begin{pmatrix} 18 & 0 & 18 & 0 \\ 6 & 12 & 6 & 12 \\ 7 & 11 & 7 & 11 \\ 5 & 13 & 5 & 13 \end{pmatrix}.$$

The distribution of (Y_0, Y_1) is given by

$$[P^{Y_0, Y_1}((i, j))]_{i,j=1,\dots,4} = \frac{1}{144} \begin{pmatrix} 18 & 0 & 18 & 0 \\ 6 & 24 & 6 & 0 \\ 7 & 11 & 7 & 11 \\ 5 & 1 & 5 & 25 \end{pmatrix},$$

with transition matrix

$$P^Y = \frac{1}{36} \begin{pmatrix} 18 & 0 & 18 & 0 \\ 6 & 24 & 6 & 0 \\ 7 & 11 & 7 & 11 \\ 5 & 1 & 5 & 25 \end{pmatrix},$$

which is not stochastically monotone, but doubly stochastic. It can be checked that the time reversal, \tilde{P}^Y , of P^Y is also not monotone. By direct computation we see that, for any triple of functions f_i , $i = 0, 1, 2$,

$$\begin{aligned} R &:= E_\pi \left[\prod_{i=0}^2 f_i(Y_i) \right] - E_\pi \left[\prod_{i=0}^2 f_i(X_i) \right] \\ &= \alpha(f_2(d) - f_2(b)) \left(f_1(c) \left[\sum_{x \in \mathbb{E}} \tilde{p}^X(c, x) f_0(x) \right] - f_1(a) \left[\sum_{x \in \mathbb{E}} \tilde{p}^X(a, x) f_0(x) \right] \right) \\ &\quad + \alpha(f_0(c) - f_0(a)) \left(f_1(d) \left[\sum_{x \in \mathbb{E}} p^X(d, x) f_2(x) \right] - f_1(b) \left[\sum_{x \in \mathbb{E}} p^X(b, x) f_2(x) \right] \right). \end{aligned}$$

For $(a, d) = (2, 4)$ and $(c, b) = (4, 2)$, we obtain

$$R = \alpha(f_2(4) - f_2(2)) \left(f_1(4) \left[\sum_{x \in \mathbb{E}} \tilde{p}^X(4, x) f_0(x) \right] - f_1(2) \left[\sum_{x \in \mathbb{E}} \tilde{p}^X(2, x) f_0(x) \right] \right) + \alpha(f_0(4) - f_0(2)) \left(f_1(4) \left[\sum_{x \in \mathbb{E}} p^X(4, x) f_2(x) \right] - f_1(2) \left[\sum_{x \in \mathbb{E}} p^X(2, x) f_2(x) \right] \right).$$

Recalling that

$$\tilde{P}^X(2, \cdot) = \tilde{P}^X(4, \cdot), \quad P^X(2, \cdot) \leq_{st} P^X(4, \cdot),$$

we conclude that, for functions $f_i, i = 0, 1, 2$, which are monotone in the same direction, we have $R \geq 0$ and, therefore, $(X_0, X_1, X_2) \leq_{cc}^3 (Y_0, Y_1, Y_2)$.

Proceeding in a similar way and applying Corollary 2.1 to the pair of states $((a, d) = (2, 3), (c, b) = (3, 2))$, we find that in this case $R \geq 0$ does not hold for general functions $f_i, i = 0, 1, 2$, which are monotone in the same direction. Thus, $(X_0, X_1, X_2) \leq_{cc}^3 (Y_0, Y_1, Y_2)$ does not hold, although for two-dimensional marginal distributions we have $(X_0, X_1) \leq_{cc} (Y_0, Y_1)$ and $(X_1, X_2) \leq_{cc} (Y_1, Y_2)$. A closer look reveals that the essential properties are

$$\tilde{P}^X(2, \cdot) \leq_{st} \tilde{P}^X(4, \cdot), \quad P^X(2, \cdot) \leq_{st} P^X(4, \cdot)$$

in the first case and

$$\tilde{P}^X(3, \cdot) \leq_{st} \tilde{P}^X(2, \cdot), \quad P^X(3, \cdot) \leq_{st} P^X(2, \cdot)$$

in the second case, and while in the first case we shift the probability mass α in a direction which is comonotone with the inherent stochastic monotonicity of P^X and \tilde{P}^X , in the second case we shift the probability mass α in a direction which is converse to the inherent stochastic monotonicity of P^X and \tilde{P}^X .

The next example indicates that monotonicity assumptions for all states are not necessary in order to obtain supermodular ordering for finite Markov sequences.

Example 3.2. Let \mathbb{E} be a discrete-lattice ordered space with partial order ' $<$ ' and assume that $b < a$ for $a, b \in \mathbb{E}, a \neq b$. Let $X = (X_n : n = 0, 1, 2)$ be a stationary, homogeneous Markov chain with state space \mathbb{E} , transition matrix P^X , and one-dimensional stationary distribution π . Let \tilde{P}^X be the transition matrix of the time reversal, \tilde{X} , of X . With the rows of P^X and \tilde{P}^X treated as distributions on \mathbb{E} , we assume that

$$P^X(b, \cdot) <_{st} P^X(a, \cdot), \quad \tilde{P}^X(b, \cdot) <_{st} \tilde{P}^X(a, \cdot),$$

and that

$$P^{X_0, X_1}(a, b) \geq \alpha, \quad P^{X_0, X_1}(b, a) \geq \alpha,$$

for some $\alpha > 0$. We compare X with the stationary, homogeneous Markov chain $Y = (Y_n : n = 0, 1, 2)$ with transition matrix P^Y obtained from

$$\begin{aligned} p^Y(a, b) &= p^X(a, b) - \frac{\alpha}{\pi(a)}, & p^Y(b, a) &= p^X(b, a) - \frac{\alpha}{\pi(b)}, \\ p^Y(a, a) &= p^X(a, a) + \frac{\alpha}{\pi(a)}, & p^Y(b, b) &= p^X(b, b) + \frac{\alpha}{\pi(b)}, \\ p^Y(u, v) &= p^X(u, v) \quad \text{for all other } (u, v) \in \mathbb{E}^2. \end{aligned}$$

Note that Y also has invariant distribution π . Let $f : (\mathbb{E}^3, \prec^3) \rightarrow \mathbb{R}$ be a bounded supermodular function. Direct computation yields

$$E_\pi[f(Y_0, Y_1, Y_2)] - E_\pi[f(X_0, X_1, X_2)] = \sum_{x_0 \in E} \bar{p}^X(a, x_0)[f(x_0, a, a) - f(x_0, a, b)] - \sum_{x_0 \in E} \bar{p}^X(b, x_0)[f(x_0, b, a) - f(x_0, b, b)] \tag{3.5}$$

$$+ \sum_{x_2 \in E} p^X(a, x_2)[f(a, a, x_2) - f(b, a, x_2)] - \sum_{x_2 \in E} p^X(b, x_2)[f(a, b, x_2) - f(b, b, x_2)] \tag{3.6}$$

$$+ \frac{\alpha}{\pi(a)}[f(a, a, a) - f(a, a, b) - f(b, a, a) + f(b, a, b)] \tag{3.7}$$

$$+ \frac{\alpha}{\pi(b)}[f(b, b, b) - f(b, b, a) - f(a, b, a) + f(a, b, a)]. \tag{3.8}$$

If we assume now that $f \in \mathcal{L}_{sm}(\mathbb{E})$, then, from supermodularity, (3.7) and (3.8) are nonnegative. Utilizing the increasing differences property (which follows from supermodularity), we find that $f(x_0, a, a) - f(x_0, a, b)$ and $f(x_0, b, a) - f(x_0, b, b)$ are increasing in x_0 . Furthermore, for a fixed x_0 , $f(x_0, a, a) - f(x_0, a, b) \geq f(x_0, b, a) - f(x_0, b, b)$. Thus, $\bar{P}^X(b, \cdot) \prec_{st} \bar{P}^X(a, \cdot)$ implies that (3.5) is nonnegative. By similar arguments and the fact that $P^X(b, \cdot) \prec_{st} P^X(a, \cdot)$, we find that (3.6) is also nonnegative. Because this holds for any $f \in \mathcal{L}_{sm}(\mathbb{E})$, we conclude that $(X_0, X_1, X_2) \prec_{sm} (Y_0, Y_1, Y_2)$.

4. Applications

4.1. Asymptotic variance

A nice idea for calculating expected values for random variables having distribution π , that is, expressions of the form $\pi(f) = \int f(x)\pi(dx)$, where f is a real function on a state space $(\mathbb{E}, \mathcal{E})$, is to construct a Harris recurrent Markov chain $X = (X_t : t \in \mathbb{Z}_+)$ such that π is its stationary (invariant) distribution, and then use $\hat{\mu}_n(f) = (1/n) \sum_{i=1}^n f(X_i)$ as the Monte Carlo approximation of $\pi(f)$. The class of possible chains with π as their stationary distribution is usually large, and a common performance criterion for the estimators obtained from using the different transition kernels is to measure the asymptotic variance

$$v(f, K^X) := \lim_{n \rightarrow \infty} \frac{1}{n} \text{var}_\pi \left(\sum_{i=1}^n f(X_i) \right)$$

for the version of the process stationary under π . This choice is well motivated since, under some regularity conditions, the central limit theorem for X holds, i.e.

$$\sqrt{n}(\hat{\mu}_n(f) - \pi(f)) \xrightarrow{D} N(0, v(f, K^X)) \quad \text{as } n \rightarrow \infty,$$

where ‘ \xrightarrow{D} ’ denotes convergence in distribution and $N(0, v(f, K^X))$ denotes the zero-mean, normal random variable with variance $v(f, K^X)$. In order to compare asymptotic variances under π it is worth recalling that it is possible to express v as

$$v(f, K^X) = (f, f)_\pi - \pi(f)^2 + 2 \sum_{k=1}^\infty (f, (K^X)^k f)_\pi, \tag{4.1}$$

where $(K^X)^k$ denotes the k -fold composition of K^X as an operator acting on $f \in L^2(\mathbb{E}, \pi)$. Moreover, for a time-reversible K^X ,

$$v(f, K^X) = \int_{-1}^1 \frac{1 + \lambda}{1 - \lambda} E_{f, K^X}(d\lambda), \tag{4.2}$$

where $E_{f, K^X}(d\lambda) = (f, E_{K^X}(d\lambda)f)_\pi$, E_{K^X} being the resolution of the identity in the spectral decomposition theorem for self-adjoint operators K^X acting on $f \in L^2_0(\mathbb{E}, \pi) = \{g \in L^2(\mathbb{E}, \pi) : \pi(g) = 0\}$. Peskun (1973) introduced a partial ordering on Markov kernels with the same invariant distribution π as follows: $K^X \leq K^Y$ if $K^X(x, F \setminus \{x\}) \leq K^Y(x, F \setminus \{x\})$ for all $F \in \mathcal{E}$ and π -almost surely all $x \in \mathbb{E}$. Using (4.2), Tierney (1998) generalized a result of Peskun (1973) proving that if $K^X \leq K^Y$ then $(f, K^Y f)_\pi \leq (f, K^X f)_\pi$ for all $f \in L^2(\mathbb{E}, \pi)$, which means that one-step autocorrelations are comparable and, moreover, $v(f, K^Y) \leq v(f, K^X)$ for all $f \in L^2_0(\mathbb{E}, \pi)$, under the additional assumption that K^X and K^Y are time reversible (i.e. self-adjoint as operators acting on $L^2_0(\mathbb{E}, \pi)$). Mira and Geyer (1999) introduced another ordering, ‘ \leq_1 ’, in which $K^X \leq_1 K^Y$ if $(f, K^Y f)_\pi \leq (f, K^X f)_\pi$ for all $f \in L^2_0(\mathbb{E}, \pi)$, and proved that if K^X and K^Y are time reversible and irreducible, then $K^X \leq_1 K^Y$ if and only if $v(f, K^Y) \leq v(f, K^X)$ for all $f \in L^2_0(\mathbb{E}, \pi)$. Mira (2001) considered a finite \mathbb{E} and a monotone f (introducing some ordering on \mathbb{E} which makes f monotone), and introduced the south-west ordering, ‘ \leq_{sw} ’, for arbitrary matrices, in which $M_1 \leq_{sw} M_2$ if $T^\top \text{diag}(\pi) M_1 T^\top(i, j) \leq T^\top \text{diag}(\pi) M_2 T^\top(i, j)$ for all $i, j \in \mathbb{E}$ not indexing the first row or the last column. Here T^\top denotes the transpose of the matrix T defined below (3.1), and $\text{diag}(\pi)$ denotes the diagonal matrix with π on the diagonal and 0s elsewhere. Mira (2001) observed that if $(I - P^X + A)^{-1} \leq_{sw} (I - P^Y + A)^{-1}$ then $v(f, K^Y) \leq v(f, K^X)$, where $P^X = [K^X(x, \{y\})]_{x, y \in \mathbb{E}}$ and A is the transition matrix for the corresponding independent, identically distributed chain, that is, the matrix whose rows all equal π . Stationarity-preserving transfers (which imply the south-west ordering for transition matrices) considered by Mira coincide with the transformations described in Lemma 2.1 and Corollary 2.1, and the south-west ordering for transition matrices is clearly related to the weak concordance ordering (see Definition 2.2). We obtain a related result immediately from Theorem 3.1 and (4.1).

Corollary 4.1. *For stationary Markov processes $X = (X_t : t \in \mathbb{Z}_+)$ and $Y = (Y_t : t \in \mathbb{Z}_+)$ on $(\mathbb{E}, \mathcal{E}, <)$ with common invariant distribution π , under STM1 we have*

$$X <_{cc} Y \implies v(f, K^X) \leq v(f, K^Y),$$

for all $f \in \mathcal{I}^*_+(\mathbb{E})$ and all $f \in \mathcal{D}^*_+(\mathbb{E})$.

The rate of convergence to stationarity for a transient Markov process with transition kernel K^X is related to the spectral radius of K^X treated as an operator on $L^2(\mathbb{E}, \pi)$, which for a finite state space is the second-largest eigenvalue in absolute value. Fast convergence happens for eigenvalues small in absolute value, while, as follows from (4.2), small asymptotic variance occurs for small positive eigenvalues and large negative eigenvalues. For positive operators (positive-definite matrices), these two goals are not in conflict; however, in many other cases they are.

4.2. Cheeger-type constants

In this section, we assume for simplicity that \mathbb{E} is finite. As before, consider two stationary, ergodic Markov chains, X and Y , on a state space \mathbb{E} , with common invariant distribution π .

Define Cheeger constants h^X and h'^X for X by

$$h^X = \min_S \frac{P^{X_0, X_1}(S \times S^c)}{\pi(S)\pi(S^c)}, \quad h'^X = \min_S \frac{P^{X_0, X_1}(S \times S^c)}{\pi(S) \wedge \pi(S^c)},$$

where the minima are taken over all subsets S of \mathbb{E} with $\pi(S) \in (0, 1)$, and S^c denotes the complement of S .

It is well known that if X is reversible then $h^X \leq 8\lambda_1^X$ and $(h'^X)^2 \leq 2\lambda_1^X$, where $\lambda_1^X < 1$ is the second-largest eigenvalue of P^X . (These are Cheeger inequalities; see, e.g. Chen and Wang (2000) or Chen (2005, pp. 67–88).)

The original Cheeger inequality, found by Cheeger, bounds the eigenvalues of the Laplacian on a Riemannian manifold. A discrete version is due to Alon and Millman (1985) (see also Diaconis and Stroock (1991)). From the following example it will be clear that, in some special situations, $P^{X_0, X_1} \prec_{cc}^2 P^{Y_0, Y_1}$ implies that $h^X \geq h^Y$ and $h'^X \geq h'^Y$. It would be of interest to find some general conditions which, together with ' \prec_{cc}^2 ', would imply inequalities for Cheeger constants.

Example 4.1. Assume that, for $a, b \in \mathbb{E}$, $a \neq b$, we have $a < b$. Note that $(a, b) \in \mathbb{E}^2$ and $(b, a) \in \mathbb{E}^2$ are not comparable coordinatewise. Assume, moreover, that, for $P^{X_0, X_1}(i, j) := \pi(i)p^X(i, j)$ and $P^{Y_0, Y_1}(i, j) := \pi(i)p^Y(i, j)$, defined for all $(i, j) \in \mathbb{E}^2$, we have $P^{X_0, X_1}(a, b) \geq \alpha$ and $P^{X_0, X_1}(b, a) \geq \alpha$, for some $\alpha > 0$, and

$$\begin{aligned} P^{Y_0, Y_1}(a, a) &= P^{X_0, X_1}(a, a) + \alpha, & P^{Y_0, Y_1}(b, b) &= P^{X_0, X_1}(b, b) + \alpha, \\ P^{Y_0, Y_1}(a, b) &= P^{X_0, X_1}(a, b) + \alpha, & P^{Y_0, Y_1}(b, a) &= P^{X_0, X_1}(b, a) - \alpha, \\ P^{Y_0, Y_1}(u, v) &= P^{X_0, X_1}(u, v) \quad \text{for all other } (u, v) \in \mathbb{E}^2. \end{aligned}$$

Then, from Corollary 2.1, $P^{X_0, X_1} \prec_{cc}^2 P^{Y_0, Y_1}$. The above equalities define a mass transformation which guarantees that $P^{X_0, X_1}(S \times S^c) \geq P^{Y_0, Y_1}(S \times S^c)$, since it moves mass into the main diagonal. Because π is the same for both processes, for each S such that $\pi(S) \in (0, 1)$ we have

$$\frac{P^{X_0, X_1}(S \times S^c)}{\pi(S)\pi(S^c)} \geq \frac{P^{Y_0, Y_1}(S \times S^c)}{\pi(S)\pi(S^c)}$$

and, therefore, $h^X \geq h^Y$ and, similarly, $h'^X \geq h'^Y$.

From Theorem 3.1, if we assume STM1 then in this case we have $X \prec_{cc} Y$, and both the corresponding inequalities for the Cheeger constants and the asymptotic variance reduction from Section 4.1 hold.

Recall that Theorem 3.3 for stochastic, monotone, reversible processes can be rephrased using bilinear forms: $X \prec_{cc} Y$ if and only if $D^X(f, g) \geq D^Y(f, g)$ for all nonnegative, comonotone functions f and g , where $D^X(f, g) = \langle f, -Q^X g \rangle$, the scalar product being defined in $L^2(\mathbb{E}, \pi)$. It would be natural to define Cheeger-type constants using increasing sets S , and study their role in mixing time and speed-of-convergence problems. This will be a topic of a further study.

4.3. Positive dependence

A natural way to define positive dependence for a random vector (or, alternatively, for a distribution on a state space) $X = (X_1, \dots, X_n)$ is to use a dependency ordering in order to compare it with its independent, identically distributed version, i.e. with $X^\perp = (X_1^\perp, \dots, X_n^\perp)$,

where $X_i \stackrel{D}{=} X_i^\perp$, and $X_1^\perp, \dots, X_n^\perp$ are independent. For example, if $\mathbb{E} = \mathbb{R}$ then $X^\perp \leq_{cc} X$ is equivalent to the fact that X is positive orthant dependent (for definitions of this and other related concepts, see, e.g. Szekli (1995)). Positive orthant dependence is weaker than the association of X defined by the condition that $\text{cov}(f(X), g(X)) \geq 0$ for all $f, g \in \mathcal{I}^*(\mathbb{R}^n)$, which is not possible to express as X being greater than X^\perp for some ordering. However, Christofides and Veggelatos (2004) showed that association implies that $X^\perp \leq_{sm} X$ (positive supermodular dependence). In fact, they showed that the weak association (defined by

$$\text{cov}(f(X_i, i \in A), g(X_i, i \in A^c)) \geq 0$$

for all real, increasing functions f and g of appropriate dimension and all $A \subset \{1, \dots, n\}$) implies positive supermodular dependence. Rüschemdorf (2004) defined a positive dependence weaker than weak association, which he called weak association in sequence, by

$$\text{cov}(1_{\{X_i > t\}}, g(X_{i+1}, \dots, X_n)) \geq 0$$

for all increasing functions g , all $t \in \mathbb{R}$, and all $i = 1, \dots, n - 1$. He showed that weak association in sequence implies positive supermodular dependence. Hu *et al.* (2004) gave examples showing that the mentioned positive dependence concepts are, however, different.

For general state spaces \mathbb{E} , the structure of the dependence can be much more complex, because, roughly speaking, there appear ‘space’, ‘time’, and joint ‘space–time’ dependency regions. By analogy, we can define X to be positive ‘orthant’ dependent if

$$P(X_1 \in A_1, \dots, X_n \in A_n) \geq P(X_1 \in A_1) \cdots P(X_n \in A_n) \tag{4.3}$$

for all $A_i \in \mathcal{I}_p(\mathbb{E}_i)$, $i = 1, \dots, n$, and all $A_i \in \mathcal{D}_p(\mathbb{E}_i)$, $i = 1, \dots, n$. This condition is just $X^\perp \prec_{cc-wk} X$.

If (4.3) holds for all comonotone sets, we similarly have positive monotone set dependence for X , that is, $X^\perp \prec_{cc} X$. From Theorem 3.1 we immediately obtain a corollary.

Corollary 4.2. *For a stationary Markov process $X = (X_t : t \in \mathbb{Z}_+)$ on $(\mathbb{E}, \mathcal{E}, \prec)$ with invariant distribution π , under the condition that both K^X and \bar{K}^X are stochastically monotone, if $(X_0^\perp, X_1^\perp) \prec_{cc}^2 (X_0, X_1)$ then $X^\perp \prec_{cc} X$.*

It is immediate that if K^X is stochastically monotone then $(X_0^\perp, X_1^\perp) \prec_{cc}^2 (X_0, X_1)$ holds if π is associated. Note that $X^\perp \prec_{cc} X$ expresses positive dependence in the time evolution of X , which is explicit if we restate this condition as the positive upper-orthant dependence of $(f_1(X_1), \dots, f_n(X_n))$ for all $f_i \in \mathcal{I}_+^*(\mathbb{E})$, $i = 1, \dots, n$, and the positive lower-orthant dependence of $(f_1(X_1), \dots, f_n(X_n))$ for all $f_i \in \mathcal{D}_+^*(\mathbb{E})$, $i = 1, \dots, n$. This corollary is in the same spirit as Theorem C, Section 3.7, of Szekli (1995), which states that, for a stationary process $X = (X_t : t \in \mathbb{R})$ defined as in (our) Section 3.2, if \mathbb{K}^X and $\bar{\mathbb{K}}^X$ are stochastically monotone and π is associated on \mathbb{E} , then

$$\text{cov}(f(X_{t_i}, i = 1, \dots, k), g(X_{t_i}, i = k + 1, \dots, n)) \geq 0 \tag{4.4}$$

for all $f \in \mathcal{I}^*(\mathbb{E}^k)$, $g \in \mathcal{I}^*(\mathbb{E}^{n-k})$, $k = 1, \dots, n - 1$, and t_i with $t_1 < \dots < t_n$. Note that property (4.4) is a rather strong positive dependence in the time evolution of X ; e.g. it implies that $(f_1(X_{t_1}), \dots, f_n(X_{t_n}))$ is weakly associated in sequence for all $f_i \in \mathcal{I}_+^*(\mathbb{E})$ and, therefore, is also positive supermodular dependent, which allows us to compare maxima, minima, and other supermodular functionals of the time evolution of X , $(f_1(X_{t_1}), \dots, f_n(X_{t_n}))$, with the corresponding independent versions.

It is worth mentioning that, in order to obtain a joint space and time positive dependence for X , we require additional assumptions. For example, it is known (see, e.g. Liggett (1985, pp. 80–82) and Szekli (1995, Theorem A, Section 3.7)) that if \mathbb{K}^X is stochastically monotone, π is associated on \mathbb{E} , and $A^X(fg) \geq fA^Xg + gA^Xf$ for all increasing functions f and g , then X is space–time associated (i.e. $\text{cov}(\phi(X_{t_i}, i = 1, \dots, n), \psi(X_{t_i}, i = 1, \dots, n)) \geq 0$ for all increasing functions ϕ and ψ).

Acknowledgements

We thank the referee for his/her remarks, which helped to improve the presentation of this work. We thank Cornelia Sauer for helpful discussions on the subject of this paper. Thanks are also due to the Mittag-Leffler Institute, where part of the present research was done, and particularly to Professor I. Kaj, who invited R. Szekli to visit the institute.

References

- ALON, N. AND MILLMAN, V. D. (1985). λ_1 , isoperimetric inequalities for graphs, and superconcentrators. *J. Combinatorial Theory B* **38**, 73–88.
- CHEN, M.-F. (2004). *From Markov Chains to Non-Equilibrium Particle Systems*. World Scientific, Singapore.
- CHEN, M.-F. (2005). *Eigenvalues, Inequalities and Ergodic Theory*. Springer, London.
- CHEN, M.-F. AND WANG, F.-Y. (1993). On order preservation and positive correlations for multidimensional diffusion processes. *Prob. Theory Relat. Fields* **95**, 421–428.
- CHEN, M.-F. AND WANG, F.-Y. (2000). Cheeger's inequalities for general symmetric forms and existence criteria for spectral gap. *Ann. Prob.* **28**, 235–257.
- CHRISTOFIDES, T. C. AND VAGGELATOU, E. (2004). A connection between supermodular ordering and positive/negative association. *J. Multivariate Anal.* **88**, 138–151.
- DADUNA, H. AND SZEKLI, R. (1995). Dependencies in Markovian networks. *Adv. Appl. Prob.* **27**, 226–254.
- DALEY, D. J. (1968). The correlation structure of the output process of some single server queuing systems. *Ann. Math. Statist.* **39**, 1007–1019.
- DIACONIS, P. AND STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1**, 36–61.
- GRIFFEATH, D. (1979). *Additive and Cancellative Interacting Particle Systems* (Lecture Notes Math. **724**). Springer, Berlin.
- HARDY, G. H., LITTLEWOOD, J. E. AND POLYA, G. (1952). *Inequalities*, 2nd edn. Cambridge University Press.
- HOEFFDING, W. (1940). Maßstabinvariante Korrelationstheorie. *Schr. Math. Inst. Inst. Angew. Math. Univ. Berlin* **5**, 179–233.
- HU, T. AND PAN, X. (2000). Comparisons of dependence for stationary Markov processes. *Prob. Eng. Inf. Sci.* **14**, 299–315.
- HU, T., MÜLLER, A. AND SCARSINI, M. (2004). Some counterexamples in positive dependence. *J. Statist. Planning Inference* **124**, 153–158.
- JOE, H. (1990). Multivariate concordance. *J. Multivariate Anal.* **35**, 12–30.
- JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- KEILSON, J. AND KESTER, A. (1977). Monotone matrices and monotone Markov processes. *Stoch. Process. Appl.* **5**, 231–241.
- KULIK, R. AND SZEKLI, R. (2004). Dependence orderings for some functionals of multivariate point processes. *J. Multivariate Anal.* **92**, 145–173.
- LI, H. AND XU, S. H. (2000). Stochastic bounds and dependence properties of survival times in a multicomponent shock model. *J. Appl. Prob.* **37**, 1020–1043.
- LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.
- LINDQVIST, B. H. (1988). Association of probability measures. *J. Multivariate Anal.* **26**, 111–132.
- LORENTZ, G. G. (1953). An inequality for rearrangements. *Amer. Math. Monthly* **60**, 176–179.
- MASSEY, W. A. (1987). Stochastic ordering for Markov processes on partially ordered spaces. *Math. Operat. Res.* **12**, 350–367.
- MIRA, A. (2001). Efficiency increasing and stationarity preserving probability mass transfers for MCMC. *Statist. Prob. Lett.* **54**, 405–411.
- MIRA, A. AND GEYER, C. J. (1999). Ordering Monte Carlo Markov chains. Submitted.
- MÜLLER, A. AND STOYAN, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley, Chichester.

- PESKUN, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.
- RÜSCHENDORF, L. (1980). Inequalities for the expectation of δ monotone functions. *Z. Wahrscheinlichkeitsth.* **54**, 341–349.
- RÜSCHENDORF, L. (2004). Comparison of multivariate risks and positive dependence. *Adv. Appl. Prob.* **41**, 391–406.
- SHAKED, M. AND SHANTHIKUMAR, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Boston, MA.
- SZEKLI, R. (1995). *Stochastic Ordering and Dependence in Applied Probability* (Lecture Notes Statist. **97**). Springer, New York.
- SZEKLI, R., DISNEY, R. L. AND HUR, S. (1994). MR/GI/1 queues with positively correlated arrival streams. *J. Appl. Prob.* **31**, 497–514.
- TCHEN, A. (1980). Inequalities for distributions with given marginals. *Ann. Prob.* **8**, 811–827.
- TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Prob.* **8**, 1–9.
- VAN DOORN, E. (1981). *Stochastic Monotonicity of Birth–Death Processes* (Lecture Notes Statist. **4**). Springer, Berlin.
- WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4**, 1280–1289.