Helsinki University of Technology Institute of Mathematics Research Reports
Espoo 2008 A554

STOCHASTIC RELATIONS OF RANDOM VARIABLES AND PROCESSES

Lasse Leskelä

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Institute of Mathematics Research Reports
Espoo 2008 A554

STOCHASTIC RELATIONS OF RANDOM VARIABLES AND PROCESSES

Lasse Leskelä

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Mathematics and Systems Analysis

Lasse Leskelä: *Stochastic relations of random variables and processes*; Helsinki University of Technology Institute of Mathematics Research Reports A554 (2008).

Abstract: *This paper generalizes the notion of stochastic order to a relation between probability measures over arbitrary measurable spaces. This generalization is motivated by the observation that for the stochastic ordering of two stationary Markov processes, it suffices that the generators of the processes preserve some, not necessarily reflexive or transitive, subrelation of the order relation. The main contributions of the paper are: a functional characterization of stochastic relations, necessary and sufficient conditions for the preservation of stochastic relations, and an algorithm for finding subrelations preserved by probability kernels. The theory is illustrated with applications to hidden Markov processes, population processes, and queueing systems.*

AMS subject classifications: 60B99, 60E15, 60G99, 60J99

Keywords: stochastic order, stochastic relation, subrelation, coupling, probability kernel, random dynamical system, Markov process

Correspondence

Lasse Leskelä
Helsinki University of Technology
Department of Mathematics and Systems Analysis
P.O. Box 1100
FI-02015 TKK
Finland

http://www.iki.fi/lsl/
lasse.leskela@iki.fi

ISBN 978-951-22-9591-3 (print)
ISBN 978-951-22-9592-0 (PDF)
ISSN 0784-3143 (print)
ISSN 1797-5867 (PDF)

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Mathematics and Systems Analysis
P.O. Box 1100, FI-02015 TKK, Finland
email: math@tkk.fi http://math.tkk.fi/

1 Introduction

Comparison techniques based on stochastic orders [21, 23, 26] are key to obtaining upper and lower bounds for complicated random variables and processes in terms of simpler random elements. Consider for example two ergodic discrete-time Markov processes X and Y with stationary distributions μ_X and μ_Y, taking values in a common ordered state space, and denote by $\le_{\text{st}}$ the corresponding stochastic order. Then the upper bound

$$\mu_X \le_{\text{st}} \mu_Y \tag{1.1}$$

can be established [12] without explicit knowledge of μ_X by verifying that the corresponding transition probability kernels P_X and P_Y satisfy

$$x \le y \implies P_X(x, \cdot) \le_{\text{st}} P_Y(y, \cdot). \tag{1.2}$$

Analogous conditions for continuous-time Markov processes on countable spaces have been derived by Whitt [29] and Massey [20], and later extended to more general jump processes by Brandt and Last [3].

The starting point of this paper is to generalize the notion of stochastic order by denoting $X \sim_{\text{st}} Y$, if there exists a coupling $(\hat{X}, \hat{Y})$ of X and Y such that $\hat{X} \sim \hat{Y}$ almost surely, where $\sim$ denotes some relation between the state spaces of X and Y. The main motivation for this definition is that (1.2) is by no means necessary for (1.1); a less stringent sufficient condition is that

$$x \sim y \implies P_X(x, \cdot) \sim_{\text{st}} P_Y(y, \cdot) \tag{1.3}$$

for some, not necessarily symmetric or transitive, nontrivial subrelation of the underlying order relation. Another advantage of the generalized definition is that X and Y are no longer required to take values in the same state space, leading to greater flexibility in the search for bounding random elements Y. For example, to study whether $f(X) \le_{\text{st}} g(Y)$ for some given real functions f and g defined on the state spaces of X and Y, we may define a relation $x \sim y$ by the condition $f(x) \le g(y)$ [5].

The main contributions of the paper are: a functional characterization of stochastic relations, necessary and sufficient conditions for the preservation of stochastic relations in the sense of (1.3), and an algorithm for finding subrelations preserved by probability kernels. The functional characterization (Section 2) is given in terms of relational conjugates that were implicitly defined by Strassen [25, Theorem 11], and the proof goes along similar lines, the new feature being the use of compact sets and upper semicontinuous functions instead of completions of measures. López and Sanz have characterized the preservation of stochastic relations for Markov processes on countable spaces in terms of a subtle order construction [18]. Section 3 describes an equivalent, considerably simpler characterization based on relational conjugates, together with an iterative algorithm for finding the maximal subrelation of a given relation preserved by a pair of probability kernels. The main results are extended to the context of general random processes and Markov processes in

Section 4. Applications to hidden Markov processes, population processes, and queueing systems are discussed in Section 5. Section 6 concludes the paper.

2 Stochastic relations

2.1 Definitions

Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be measurable spaces, and denote by $\mathcal{P}(S_i)$ the family of probability measures on $(S_i, \mathcal{S}_i)$. Unless otherwise mentioned, all spaces shall implicitly be assumed Polish (complete separable metrizable) and equipped with the Borel sigma-algebra. A *coupling* of probability measures $\mu_1 \in \mathcal{P}(S_1)$ and $\mu_2 \in \mathcal{P}(S_2)$ is a probability measure $\mu \in \mathcal{P}(S_1 \times S_2)$ with marginals μ_1 and μ_2, that is, $\mu \circ \pi_i^{-1} = \mu_i$ for $i = 1, 2$, where π_i denotes the projection map from $S_1 \times S_2$ onto S_i. If μ is a coupling of μ_1 and μ_2, we also say that μ couples μ_1 and μ_2 [16, 27].

A *measurable relation* between S_1 and S_2 is measurable subset of $S_1 \times S_2$. All relations in this paper are assumed to be closed (in the product topology of $S_1 \times S_2$), if not otherwise mentioned. Given a nontrivial ($R \neq \emptyset$) measurable relation R between S_1 and S_2, we write $x_1 \sim x_2$, if $(x_1, x_2) \in R$. For probability measures $\mu_1 \in \mathcal{P}(S_1)$ and $\mu_2 \in \mathcal{P}(S_2)$ we denote

$$\mu_1 \sim_{\text{st}} \mu_2,$$

and say that μ_1 is stochastically related to μ_2, if there exists a coupling μ of μ_1 and μ_2 such that $\mu(R) = 1$. The relation $R_{\text{st}} = \{(\mu_1, \mu_2) : \mu_1 \sim_{\text{st}} \mu_2\}$ is called the *stochastic relation* generated by R. Observe that two Dirac measures satisfy $\delta_{x_1} \sim_{\text{st}} \delta_{x_2}$ if and only if $x_1 \sim x_2$. In this way the stochastic relation R_{st} may be regarded as a natural randomization of the underlying relation R.

A random variable X_1 is *stochastically related* to a random variable X_2, denoted by $X_1 \sim_{\text{st}} X_2$, if the distribution of X_1 is stochastically related to the distribution of X_2. Observe that X_1 and X_2 do not need to be defined on the same probability space. Recall that a coupling of random variables X_1 and X_2 is a bivariate random variable whose distribution couples the distributions of X_1 and X_2. Hence $X_1 \sim_{\text{st}} X_2$ if and only if there exists a coupling $(\hat{X}_1, \hat{X}_2)$ of X_1 and X_2 such that $\hat{X}_1 \sim \hat{X}_2$ almost surely.

Example 2.1 (Stochastic equality)**.** The stochastic relation generated by the equality relation $\{(x, y) : x = y\}$ on S is the equality on $\mathcal{P}(S)$. Hence $X =_{\text{st}} Y$ if and only if X and Y have the same distribution.

Example 2.2 (Stochastic ϵ-distance)**.** Define a relation on the real line by denoting $x \approx y$, if $|x - y| \leq \epsilon$. If $X_1 \approx_{\text{st}} X_2$, then the cumulative distribution functions of X_1 and X_2 satisfy

$$F_2(x - \epsilon) \leq F_1(x) \leq F_2(x + \epsilon) \quad \text{for all } x. \tag{2.1}$$

Conversely, if (2.1) holds, it is not hard to verify that the quantile functions $G_i(r) = \inf\{x : F_i(x) \ge r\}$ satisfy $|G_1(r) - G_2(r)| \le \epsilon$ for all $r \in (0,1)$. Hence the bivariate random variable $\hat{X} = (G_1(\xi), G_2(\xi))$, with ξ uniformly distributed on $(0,1)$, couples X_1 and X_2 and satisfies $\hat{X}_1 \approx \hat{X}_2$ with probability one. Thus (2.1) is necessary and sufficient for $X_1 \approx_{\text{st}} X_2$.

Example 2.3 (Stochastic majorization)**.** Let S be closed subset of $\mathbb{R}^n$, and denote by $x_{[1]} \ge \cdots \ge x_{[n]}$ the components of $x \in S$ in decreasing order. The *weak majorization order* on S is defined by denoting $x \preceq^{\text{wm}} y$, if $\sum_{i=1}^k x_{[i]} \le \sum_{i=1}^k y_{[i]}$ for all $k = 1, \dots, n$; and the *majorization order* by denoting $x \preceq^{\text{m}} y$, if $x \preceq^{\text{wm}} y$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. The $\preceq^{\text{m}}$-increasing real functions are called Schur-convex, and a function is $\preceq^{\text{wm}}$-increasing if and only if it is coordinatewise increasing and Schur-convex [19, Theorem 3.A.8]. The standard characterization for stochastic orders (Remark 2.6 in Section 2.3) hence shows that $X \preceq^{\text{m}}_{\text{st}} Y$ (resp. $X \preceq^{\text{wm}}_{\text{st}} Y$) if and only if $\mathrm{E}\, f(X) \le \mathrm{E}\, f(Y)$ for all positive measurable Schur-convex (resp. coordinatewise increasing Schur-convex) functions f.

2.2 Relational conjugates

To develop a convenient way to check whether two probability measures are stochastically related or not, we shall define the *right conjugate* of $B_1 \subset S_1$ and the *left conjugate* of $B_2 \subset S_2$ with respect to a relation R by

$$\begin{aligned} B_1^{\rightarrow} &= \cup_{x \in B_1} \{y \in S_2 : x \sim y\}, \\ B_2^{\leftarrow} &= \cup_{y \in B_2} \{x \in S_1 : x \sim y\}. \end{aligned}$$

The conjugates of positive functions f_i on S_i are defined analogously by

$$\begin{aligned} f_1^{\rightarrow}(y) &= \sup_{x \in S_1 : x \sim y} f_1(x), \quad y \in S_2, \\ f_2^{\leftarrow}(x) &= \sup_{y \in S_2 : x \sim y} f_2(y), \quad x \in S_1, \end{aligned}$$

where we adopt the convention that the supremum of the empty set is zero. Relational conjugates of sets and functions are interlinked via

$$(1_{B_1})^{\rightarrow} = 1_{B_1^{\rightarrow}}, \tag{2.2}$$

where 1_{B_1} denotes the indicator function of B_1, and

$$\{x : f_1(x) > r\}^{\rightarrow} = \{y : f_1^{\rightarrow}(y) > r\}, \tag{2.3}$$

which is valid for all $r \ge 0$.

The following result summarizes the basic topological properties of right conjugates. By symmetry, analogous results are valid for left conjugates.

Lemma 2.4. *Let R be a closed relation between two Polish spaces. Then:*

(i) $B^{\rightarrow}$ is closed for compact B. Especially, $\{x\}^{\rightarrow}$ is closed for all x.

(ii) $f^{\to}$ *is upper semicontinuous (u.s.c.) for all positive u.s.c.* f *on* S_1 *with compact support.*

Proof. Assume B is compact, and consider a sequence $y_n \to y$ such that $y_n \in B^{\to}$ for all n. Then for all n there exists $x_n \in B$ such that $x_n \sim y_n$. Because B is compact, there exists $x \in B$ such that $x_n \to x$ as $n \to \infty$ along some subsequence of the natural numbers. Hence $(x_n, y_n) \to (x, y)$ as $n \to \infty$ along the same subsequence, which implies that $x \sim y$, and thus $y \in B^{\to}$.

Assume next that f is positive u.s.c. with compact support on S_1. We shall first show that

$$\{x : f(x) \geq r\}^{\to} = \{y : f^{\to}(y) \geq r\} \quad \text{for all } r > 0. \tag{2.4}$$

Observe first that if y belongs to the left side of (2.4), then $f(x) \geq r$ for some $x \sim y$, so that $f^{\to}(y) \geq r$. To prove the converse statement, assume next that $f^{\to}(y) \geq r$. Then the sets $K_n = \{f \geq r - 1/n\} \cap \{y\}^{\leftarrow}$ are nonempty and compact for all $n > 1/r$, because $\{y\}^{\leftarrow}$ is closed by property (i). Hence Cantor's intersection theorem implies that

$$\{x : f(x) \geq r\} \cap \{y\}^{\leftarrow} = \cap_{n>1/r} K_n,$$

is nonempty, so that $f(x) \geq r$ for some $x \sim y$. We may now use (2.4) together with property (i) to conclude that $\{y : f^{\to}(y) \geq r\}$ is closed for all $r > 0$. Obviously, $\{y : f^{\to}(y) \geq 0\} = S_2$ is closed as well. □

2.3 Functional characterization

The following result characterizes stochastic relations using relational conjugates of sets and functions. The key part of the characterization is essentially Strassen's Theorem 11 [25], written in a new notation. The new contributions are (ii) and (iv), providing classes of test sets and functions with Borel-measurable conjugates (Lemma 2.4) that are large enough to characterize stochastic relations without resorting to completions of measures.

Theorem 2.5. *Let* R *be a closed relation between Polish spaces* S_1 *and* S_2*. Then* $\mu \sim_{\mathrm{st}} \nu$ *is equivalent to each of the following:*

(i) $\mu(B) \leq \nu(B^{\to})$ *for all measurable* B *such that* $B^{\to}$ *is measurable.*

(ii) $\mu(B) \leq \nu(B^{\to})$ *for all compact* B*.*

(iii) $\int_{S_1} f \, d\mu \leq \int_{S_2} f^{\to} \, d\nu$ *for all positive measurable* f *such that* $f^{\to}$ *is measurable.*

(iv) $\int_{S_1} f \, d\mu \leq \int_{S_2} f^{\to} \, d\nu$ *for all positive u.s.c.* f *with compact support.*

Remark 2.6. If R is an order (reflexive and transitive) relation on S, then using the properties $B \subset B^{\to} = (B^{\to})^{\to}$ and $f \leq f^{\to} = (f^{\to})^{\to}$ we see that (i) and (iii) in Theorem 2.5 become equivalent to well-known characterizations of stochastic orders [12, 25]:

(i') $\mu(B) \le \nu(B)$ for all measurable upper sets B.

(iii') $\int_S f\,d\mu \le \int_S f\,d\nu$ for all measurable positive increasing functions f.

Remark 2.7. When S_1 and S_2 are countable, the measurability requirements of Theorem 2.5 become void, and the word "compact" becomes replaced by "finite".

Proof of Theorem 2.5. $\mu \sim_{\rm st} \nu \implies$ (i). Let λ be a coupling of μ and ν such that $\lambda(R) = 1$. Then because

$$(B \times S_2) \cap R = (B \times B^{\rightarrow}) \cap R,$$

we see that

$$\mu(B) = \lambda(B \times S_2) = \lambda(B \times B^{\rightarrow}) \le \lambda(S_1 \times B^{\rightarrow}) = \nu(B^{\rightarrow})$$

for all measurable $B \subset S_1$ such that $B^{\rightarrow}$ is measurable.

(i) $\implies$ (ii). Clear by Lemma 2.4.

(ii) $\implies$ (iv). Let f be a positive compactly supported u.s.c. function on S_1. Then equality (2.4) shows that

$$\mu(\{x : f(x) \ge r\}) \le \nu(\{y : f^{\rightarrow}(y) \ge r\})$$

for all $r > 0$. The validity of (iv) hence follows by integrating both sides of the above inequality with respect to r over $(0, \infty)$.

(iv) $\implies \mu \sim_{\rm st} \nu$. By virtue of [25, Theorem 7], it suffices to show that

$$\int_{S_1} f\,d\mu + \int_{S_2} g\,d\nu \le \sup_{(x,y)\in R} (f(x) + g(y)) \tag{2.5}$$

for all bounded continuous f and g on S_1 and S_2, respectively, and without loss of generality we may assume f and g are positive and bounded by one. Given any such functions f and g, and a number $\epsilon > 0$, choose a compact set $K \subset S_1$ such that $\mu(K^c) \le \epsilon$ [1, Theorem 1.3], and define $f_0 = f 1_K$. Because f_0 is u.s.c., we see using (iv) that $\int_{S_1} f_0\,d\mu \le \int_{S_2} f_0^{\rightarrow}\,d\nu$, so that

$$\int_{S_1} f\,d\mu + \int_{S_2} g\,d\nu \le \int_{S_2} (f_0^{\rightarrow} + g)\,d\nu + \epsilon. \tag{2.6}$$

In light of (2.2), assumption (iv) further implies that $\mu(K) \le \nu(K^{\rightarrow})$, because 1_K is u.s.c.. Thus $\nu((K^{\rightarrow})^c) \le \epsilon$, so by splitting the ν-integral into $K^{\rightarrow}$ and its complement we see that

$$\int (f_0^{\rightarrow} + g)\,d\nu \le \sup_{y\in K^{\rightarrow}} (f_0^{\rightarrow}(y) + g(y)) + 2\epsilon.$$

Because $f_0^{\rightarrow} \le f^{\rightarrow}$ and $K^{\rightarrow} \subset S_1^{\rightarrow}$, the above inequality combined with (2.6) shows that

$$\int_{S_1} f\,d\mu + \int_{S_2} g\,d\nu \le \sup_{y\in S_1^{\rightarrow}} (f^{\rightarrow}(y) + g(y)) + 3\epsilon.$$

After letting $\epsilon \to 0$ and observing that

$$\sup_{y \in S_1^{\to}} (f^{\to}(y) + g(y)) = \sup_{(x,y) \in R} (f(x) + g(y)),$$

we may conclude that (2.5) holds.

Finally, observe that the proof of (i) $\implies$ (iii) is completely analogous to the proof of (ii) $\implies$ (iv), and the implication (iii) $\implies$ (iv) follows immediately by Lemma 2.4. $\square$

3 Preservation of stochastic relations

3.1 Coupling of probability kernels

Monotone functions are key objects in the study of order relations. When passing from orders to general relations, the role of monotone functions is taken over by function pairs (f_1, f_2) such that $x_1 \sim x_2 \implies f_1(x_1) \sim f_2(x_2)$. To study stochastic relations, we need a randomized version of the above property. Recall that a *probability kernel* from a measurable space S to a measurable space S' is a mapping $P : S \times \mathcal{S}' \to \mathbb{R}$ such that $P(x, \cdot)$ is a probability measure for all x, and $x \mapsto P(x, B)$ is measurable for all $B \in \mathcal{S}'$. Probability kernels may alternatively be viewed as mappings $\mathcal{P}(S) \ni \mu \mapsto \mu P \in \mathcal{P}(S')$ by defining $\mu P(B) = \int_S P(x, B)\, \mu(dx)$.

Given a closed relation R between Polish spaces S_1 and S_2, and probability kernels P_1 on S_1 and P_2 on S_2, we say that the pair (P_1, P_2) *stochastically preserves* R, if any of the equivalent conditions in Theorem 3.1 holds.

Theorem 3.1. *The following are equivalent:*

(i) $x_1 \sim x_2 \implies P_1(x_1, \cdot) \sim_{\text{st}} P_2(x_2, \cdot)$.

(ii) $\mu_1 \sim_{\text{st}} \mu_2 \implies \mu_1 P_1 \sim_{\text{st}} \mu_2 P_2$.

(iii) $P_1(x_1, B) \le P_2(x_2, B^{\to})$ *for all* $x_1 \sim x_2$ *and compact* $B \subset S_1$.

Proof. (i) $\implies$ (ii). Given $\mu_1 \sim_{\text{st}} \mu_2$, choose a coupling of μ_1 and μ_2 such that $\mu(R) = 1$. Theorem 2.5 then shows that

$$\mu_1 P_1(B) = \int_R P_1(x_1, B)\, \mu(dx) \le \int_R P_2(x_2, B^{\to})\, \mu(dx) = \mu_2 P_2(B^{\to})$$

for all compact $B \subset S_1$, so that $\mu_1 P_1 \sim_{\text{st}} \mu_2 P_2$.

The implication (ii) $\implies$ (i) follows immediately by choosing $\mu_i = \delta_{x_i}$, while the equivalence (i) $\iff$ (iii) is clear by Theorem 2.5. $\square$

Remark 3.2. A probability kernel P is said to stochastically preserve a relation R on S, if $x_1 \sim x_2 \implies P(x_1, \cdot) \sim_{\text{st}} P(x_2, \cdot)$. Order-preserving probability kernels are usually called monotone [21].

The main result of this section is the following coupling characterization of relation-preserving pairs of probability kernels. For technical reasons related to local uniformization of Markov jump processes in Section 4.3, we shall consider probability kernels P_i from S_i to S_i', where S_i' is a measurable space not necessarily equal to S_i. A probability kernel P from $S_1 \times S_2$ to $S_1' \times S_2'$ is called a *coupling* of probability kernels P_1 and P_2, if the probability measure $P(x, \cdot)$ couples the probability measures $P_1(x_1, \cdot)$ and $P_2(x_2, \cdot)$ for all $x = (x_1, x_2) \in S_1 \times S_2$.

Theorem 3.3. *Given closed relations R between S_1 and S_2, and R' between S_1' and S_2', assume that*

$$x_1 \overset{R}{\sim} x_2 \implies P_1(x_1, \cdot) \overset{R'}{\sim}_{\mathrm{st}} P_2(x_2, \cdot).$$

Then there exists a coupling P of P_1 and P_2 such that $P(x, R') = 1$ for all $x \in R$.

The proof of Theorem 3.3 requires some preliminaries on topology and measure theory that are discussed next. Denote by π_i the projection from $S_1 \times S_2$ to S_i, and define the projection maps $\hat{\pi}_i : \mathcal{P}(S_1 \times S_2) \to \mathcal{P}(S_i)$ by $\hat{\pi}_i \mu = \mu \circ \pi_i^{-1}$, so that $\hat{\pi}_i \mu$ equals the i-th marginal of μ. From now on, all sets of probability measures shall be considered as topological spaces equipped with the weak topology.

Lemma 3.4. *The projection $\hat{\pi}_i : \mathcal{P}(S_1 \times S_2) \to \mathcal{P}(S_i)$ is continuous and open with respect to the weak topology, $i = 1, 2$.*

Proof. Assume that $\mu^n \xrightarrow{w} \mu$ in $\mathcal{P}(S_1 \times S_2)$, and let $f \in C_b(S_i)$. Then because $f \circ \pi_i \in C_b(S_1 \times S_2)$, it follows that $\hat{\pi}_i \mu^n(f) = \mu^n(f \circ \pi_i) \to \mu(f \circ \pi_i) = \hat{\pi}_i \mu(f)$. Hence $\hat{\pi}_i$ is continuous. The openness of $\hat{\pi}_i$ follows from Eifler [8, Theorem 2.5], because the map π_i is continuous, open, and onto. □

Lemma 3.5. *For any $\mu_1 \in \mathcal{P}(S_1)$ and $\mu_2 \in \mathcal{P}(S_2)$, the set $K(\mu_1, \mu_2)$ of all couplings of μ_1 and μ_2 is compact in the weak topology of $\mathcal{P}(S_1 \times S_2)$.*

Proof. Given $\epsilon > 0$, choose compacts sets $C_i \subset S_i$ such that $\mu_i(C_i^c) \le \epsilon/2$, $i = 1, 2$. Define $C = C_1 \times C_2$. Then because $C^c = (C_1^c \times S_2) \cup (S_1 \times C_2^c)$, it follows that

$$\mu(C^c) \le \mu_1(C_1^c) + \mu_2(C_2^c) \le \epsilon$$

for all $\mu \in K(\mu_1, \mu_2)$. Hence $K(\mu_1, \mu_2)$ is relatively compact by Prohorov's theorem. The equality $K(\mu_1, \mu_2) = \hat{\pi}_1^{-1}(\mu_1) \cap \hat{\pi}_2^{-1}(\mu_2)$ further shows that $K(\mu_1, \mu_2)$ is closed, because $\hat{\pi}_i$ are continuous by Lemma 3.4. □

Lemma 3.6. *Let P be a probability kernel from S to S'. Then the map $x \mapsto P(x, \cdot)$ is $\mathcal{B}(S)/\mathcal{B}(\mathcal{P}(S'))$-measurable, where $\mathcal{B}(\mathcal{P}(S'))$ denotes the Borel σ-algebra generated by the weak topology on $\mathcal{P}(S')$.*

Proof. For any $f \in C_b(S')$, one may check by approximating f with simple functions that the map $x \mapsto P(x, f)$ is measurable. Hence it follows that the set $\{x : P(x, \cdot) \in A\}$ is $\mathcal{B}(S)$-measurable for any $A = \cap_{k=1}^n \{\mu : \mu(f_k) \in B_k\}$, where $f_k \in C_b(S')$ and B_k are open subsets of the real line. Because the sets of the above type form a basis for the weak topology of $\mathcal{P}(S')$, and because the space $\mathcal{P}(S')$ equipped with the weak topology is Polish [1] and hence Lindelöf, it follows that any open set in $\mathcal{P}(S')$ can be represented as a countable union of the basis sets. Hence $\{x : P(x, \cdot) \in A\}$ is measurable for all open subsets A in $\mathcal{P}(S')$, and the claim follows. □

A *set-valued mapping* from a set S to a set S' is a function that assigns to each element in S a subset of S'. A set-valued mapping F from a measurable space S to a topological space S' is *measurable* [28], if the inverse image

$$F^-(A) = \{x \in S : F(x) \cap A \neq \emptyset\}$$

is measurable for all closed $A \subset S'$.

Lemma 3.7. *Let P_i be probability kernels from S_i to S_i', $i = 1, 2$. Then the set-valued mapping $F : x \mapsto K(P(x_1, \cdot), P_2(x_2, \cdot))$ is measurable.*

Proof. Because the set $F(x) \subset \mathcal{P}(S_1' \times S_2')$ is compact for all x by Lemma 3.5, it is sufficient to verify that $F^-(A)$ is measurable for all open sets A (Himmelberg [10, Theorem 3.1]). Let us hence assume that $A \subset \mathcal{P}(S_1' \times S_2')$ is open. Observe that $F^-(A) = \pi_1^{-1}(B_1) \cap \pi_2^{-1}(B_2)$, where

$$B_i = \{x_i \in S_i : P_i(x_i, \cdot) \in \hat{\pi}_i(A)\}.$$

Now Lemma 3.4 implies that $\hat{\pi}_i(A)$ is open, and Lemma 3.6 further shows that B_i is measurable. Thus $F^-(A)$ is measurable. □

Proof of Theorem 3.3. Assume without loss of generality that $R' \neq \emptyset$, and let $\mathcal{P}(R) = \{\mu \in \mathcal{P}(S_1' \times S_2') : \mu(R') = 1\}$. Because R' is closed, it follows from Portmanteau's theorem that $\mathcal{P}(R')$ is closed. Define the set-valued mappings F and G from $S_1 \times S_2$ to $\mathcal{P}(S_1' \times S_2')$ by $F(x) = K(P_1(x_1, \cdot), P_2(x_2, \cdot))$, and

$$G(x) = \begin{cases} F(x) \cap \mathcal{P}(R'), & x \in R, \\ F(x), & \text{else.} \end{cases}$$

Then for any closed $A' \subset \mathcal{P}(S_1' \times S_2')$,

$$G^-(A') = \big(R \cap F^-(\mathcal{P}(R') \cap A')\big) \cup \big(R^c \cap F^-(A')\big).$$

Because $\mathcal{P}(R')$ is closed, it follows from Lemma 3.7 that $G^-(A')$ is measurable. Moreover, because F is compact-valued by Lemma 3.5, we may conclude that G is a measurable set-valued mapping such that $G(x)$ is compact and nonempty for all x. A measurable selection theorem of Kuratowski and Ryll–Nardzewski [15] (see alternatively Srivastava [24, Theorem 5.2.1]) now shows that there exists a measurable function $g : S_1 \times S_2 \to \mathcal{P}(S_1' \times S_2')$ such that $g(x) \in G(x)$ for all x. By defining $P(x, B) = [g(x)](B)$ for $x \in S_1 \times S_2$ and measurable $B \subset S_1' \times S_2'$, we see that P is a probability kernel from $S_1 \times S_2$ to $S_1' \times S_2'$ with the desired properties. □

3.2 Subrelation algorithm

This section presents an algorithm for finding the maximal subrelation of a closed relation that is stochastically preserved by a pair (P_1, P_2) of continuous[1] probability kernels. Given a closed relation R and continuous probability kernels P_i on S_i, $i = 1, 2$, define recursively the relations $R^{(n)}$ by $R^{(0)} = R$,

$$R^{(n+1)} = \left\{ x \in R^{(n)} : \; (P_1(x_1, \cdot), P_2(x_2, \cdot)) \in R^{(n)}_{\rm st} \right\},$$

and denote

$$R^* = \bigcap_{n=0}^{\infty} R^{(n)}. \tag{3.1}$$

Lemma 3.10 below shows that the relations $R^{(n)}$ are closed and hence measurable, so the stochastic relations $R^{(n)}_{\rm st}$ are well-defined. The following theorem underlines the key role of R^* in characterizing the existence of subrelations stochastically preserved by a pair of probability kernels.

Theorem 3.8. *Assume P_1 and P_2 are continuous. Then R^* is the maximal closed subrelation of R that is stochastically preserved by (P_1, P_2). Especially, there exists a nontrivial closed subrelation stochastically preserved by (P_1, P_2) if and only if $R^* \neq \emptyset$.*

The following three lemmas summarize the topological preliminaries required for the proof of Theorem 3.8.

Lemma 3.9. *Let R be a closed relation between Polish spaces S_1 and S_2. Then the relation $R_{\rm st}$ is closed in the weak topology of $\mathcal{P}(S_1) \times \mathcal{P}(S_2)$.*

Proof. Assume $\mu_n \xrightarrow{w} \mu$ and $\nu_n \xrightarrow{w} \nu$ such that $\mu_n \sim_{\rm st} \nu_n$ for all n. Then for all n there exists a coupling λ_n of μ_n and ν_n such that $\lambda_n(R) = 1$. Because the sequences μ_n and ν_n are tight, and because $\lambda_n((C_1 \times C_2)^c) \leq \mu_n(C_1^c) + \nu_n(C_2^c)$ for all compact C_1 and C_2, it follows that the sequence λ_n is tight, so there exists $\lambda \in \mathcal{P}(S_1 \times S_2)$ such that $\lambda_n \xrightarrow{w} \lambda$ as $n \to \infty$ along some subsequence [11, Theorem 16.3]. The continuity of $\hat{\pi}_i$ (Lemma 3.4) implies that λ is a coupling of μ and ν, and Portmanteau's theorem shows that $\lambda(R) = 1$. Hence $\mu \sim_{\rm st} \nu$. □

Lemma 3.10. *Given continuous probability kernels P_1 and P_2, define*

$$M(R) = \{x \in R : \; (P_1(x_1, \cdot), P_2(x_2, \cdot)) \in R_{\rm st}\} \tag{3.2}$$

for measurable relations R. Then:

(i) $M(R) \subset M(R')$ for $R \subset R'$.

(ii) M maps closed relations into closed relations.

[1] A probability kernel P from S to S' is called *continuous* if $P(x_n, \cdot) \to P(x, \cdot)$ in distribution whenever $x_n \to x$. In other words, P is continuous if and only if the map $x \mapsto P(x, \cdot)$ from S to $\mathcal{P}(S')$ is continuous, when $\mathcal{P}(S')$ is equipped with the weak topology.

Proof. For (i) it suffices to observe that $R \subset R'$ implies $R_{\text{st}} \subset R'_{\text{st}}$. For (ii), observe that $M(R) = R \cap f^{-1}(R_{\text{st}})$, where the function $f : S_1 \times S_2 \to \mathcal{P}(S_1) \times \mathcal{P}(S_2)$ is defined by $f(x) = (P_1(x_1, \cdot), P_2(x_2, \cdot))$. Because R_{st} is closed (Lemma 3.9) and f is continuous, it follows that $M(R)$ is closed. □

Lemma 3.11. *Let $R^{(0)} \supset R^{(1)} \supset \cdots$ be closed relations between Polish spaces S_1 and S_2, and let $R^* = \cap_{n=0}^{\infty} R^{(n)}$. Assume that $(\mu_1, \mu_2) \in R^{(n)}_{\text{st}}$ for all n. Then $(\mu_1, \mu_2) \in R^*_{\text{st}}$.*

Proof. By definition, for all n there exists a coupling λ_n of μ and ν such that $\lambda_n(R^{(n)}) = 1$. Because the set of couplings of μ and ν is compact by Lemma 3.5, there exists a coupling λ of μ and ν such that $\lambda_n \xrightarrow{w} \lambda$ as $n \to \infty$ along a subsequence of $\mathbb{Z}_+$. Further, observe that $\lambda_n(R^{(m)}) \geq \lambda_n(R^{(n)}) = 1$ for all $m \leq n$, which implies that $\lim_{n\to\infty} \lambda_n(R^{(m)}) = 1$ for all m. Portmanteau's theorem now shows that $\lambda(R^{(m)}) = 1$ for all m, so it follows that $\lambda(R^*) = 1$. □

Proof of Theorem 3.8. Let M be the map defined in (3.2). If $x \in R^*$, then $x \in R^{(n+1)} = M(R^{(n)})$ shows that $(P_1(x_1, \cdot), P_2(x_2, \cdot)) \in R^{(n)}_{\text{st}}$ for all n. Because the relations $R^{(n)}$ are closed by Lemma 3.10, we see using Lemma 3.11 that $(P_1(x_1, \cdot), P_2(x_2, \cdot)) \in R^*_{\text{st}}$. Hence (P_1, P_2) stochastically preserves R^*. On the other hand, if R' is a closed subrelation of R that is stochastically preserved by (P_1, P_2), then $R' = M(R') \subset M(R) = R^{(1)}$ by Lemma 3.10. Induction shows that $R' \subset R^{(n)}$ for all n, and thus $R' \subset R^*$. □

4 Random processes

4.1 Random sequences

Given a relation R between S_1 and S_2, the coordinatewise relation between the product spaces S_1^n and S_2^n for $n \leq \infty$ is defined by

$$R^n = \{(x, y) \in S_1^n \times S_2^n : x_i \sim y_i \text{ for all } i\}.$$

The stochastic relation generated by R^n is called the *stochastic coordinatewise relation*. The following example shows that the stochastic coordinatewise relation of two random sequences cannot be verified just by looking at the one-dimensional marginal distributions.

Example 4.1. Let ξ_1 and ξ_2 be independent random variables uniformly distributed on the unit interval, and define $X = (\xi_1, \xi_1)$ and $Y = (\xi_1, \xi_2)$. Then $X_i =_{\text{st}} Y_i$ for all $i = 1, 2$, but X and Y are not related with respect to the stochastic coordinatewise equality.

The proof of the following result is a straightforward modification of its continuous-time analogue Theorem 4.6, and shall hence be omitted.

Theorem 4.2. *Two random sequences X and Y satisfy $X \overset{R^n}{\sim}_{\mathrm{st}} Y$ if and only if $(X_{t_1}, \ldots, X_{t_k}) \overset{R^k}{\sim}_{\mathrm{st}} (Y_{t_1}, \ldots, Y_{t_k})$ for all finite parameter combinations $(t_1, \ldots, t_k)$.*

The following result, which is completely analogous to [12, Proposition 1], gives a sufficient condition for $X \sim_{\mathrm{st}} Y$ in terms of conditional probabilities. Let P_i be a probability kernel from S_1^{i-1} to S_1 representing the regular conditional distribution of X_i given $(X_1, \ldots, X_{i-1})$, and define the kernels Q_i in a similar way for Y [11, Theorem 6.3].

Theorem 4.3. *Assume that $X_1 \sim_{\mathrm{st}} Y_1$, and that for all i,*

$$P_i(x_1, \ldots, x_{i-1}, dx_i) \sim_{\mathrm{st}} Q_i(y_1, \ldots, y_{i-1}, dy_i)$$

whenever $(x_1, \ldots, x_{i-1}) \sim (y_1, \ldots, y_{i-1})$. Then $X \sim_{\mathrm{st}} Y$.

Proof. Let λ_1 be a coupling of the distributions of X_1 and Y_1 such that $\lambda_1(R) = 1$. For convenience, we shall use x^n as a shorthand for $(x_1, \ldots, x_n)$. By Theorem 3.3 there exists for each i a coupling Λ_i of probability kernels P_i and Q_i such that $\Lambda_i((x^{i-1}, y^{i-1}), R) = 1$ whenever $x^{i-1} \sim y^{i-1}$. Then it is easy to verify by induction that the probability measure

$$\lambda_n(B) = \int \cdots \int 1(z^n \in B)\, \Lambda_n(z^{n-1}, dz_n) \cdots \Lambda_2(z^1, dz_2)\, \lambda_1(dz_1)$$

couples the distributions of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, and $\lambda_n(R^n) = 1$ for any finite n. In the case where n is infinite, the proof is completed by applying Theorem 4.2. □

The next example shows that the condition in Theorem 4.3 is not necessary in general.

Example 4.4. Let $X = (\xi, 1-\xi)$ and $Y = (2\xi, 1-\xi)$, where ξ has uniform distribution on the unit interval. Then $X \le_{\mathrm{st}} Y$ by construction, but $P_2(x_1, \cdot) \le_{\mathrm{st}} Q_2(y_2, \cdot)$ only for $x_1 \ge y_1/2$.

4.2 Continuous-time random processes

Denote by $D_i = D_i(\mathbb{R}_+, S_i)$ the space of functions from $\mathbb{R}_+$ into S_i that are right-continuous and have left limits, and equip D_i with the Skorohod topology, which makes it Polish [9, Section 3.5]. The coordinatewise relation between D_1 and D_2 is defined by

$$R^D = \{(x, y) \in D_1 \times D_2 : \; x(t) \sim y(t) \text{ for all } t \in \mathbb{R}_+\},$$

and we denote by R^D_{st} the corresponding stochastic relation between random processes with paths in D_i (identified as D_i-valued random elements). When there is no risk of confusion, the same notation $\sim_{\mathrm{st}}$ shall be used for a random process (corresponding to R^D_{st} and its finite-dimensional distributions (corresponding to R^n_{st}).

Lemma 4.5. *R^D is a closed relation between D_1 and D_2, whenever R is closed.*

Proof. Assume that x_i and x_i^n are functions in D_i such that $x_i^n \to x_i$ as $n \to \infty$, and $(x_1^n, x_2^n) \in R^D$ for all n. Denote $\Delta = \Delta_1 \cup \Delta_2$, where Δ_i is the set of points $t \in \mathbb{R}_+$ where x_i is discontinuous. It is well-known [9, Section 3.5] that Δ_i is countable, and that $x_i^n(t) \to x_i(t)$ in S_i for all $t \notin \Delta_i^c$. Hence $x_1(t) \sim x_2(t)$ for all $t \notin \Delta$, because R is closed.

Observe next that if $t \in \Delta$, then there exists a sequence $t_k \in (t, \infty) \cap \Delta^c$ such that $t_k \to t$. Then $x_i(t_k) \to x_i(t)$ by the right-continuity of x_i, and again the fact that R is closed implies $x_1(t) \sim x_2(t)$. □

Theorem 4.6. *Two random processes X and Y with paths in $D_1(\mathbb{R}_+, S_1)$ and $D_2(\mathbb{R}_+, S_2)$, respectively, satisfy $X \sim_{\text{st}} Y$ if and only if $(X_{t_1}, \ldots, X_{t_n}) \sim_{\text{st}} (Y_{t_1}, \ldots, Y_{t_n})$ for all finite parameter combinations $(t_1, \ldots, t_n)$.*

Proof. The necessity is obvious. To prove the converse, define for each positive integer m the discretization map $\pi_i^m : D_i \to S_i^{m^2+1}$ by

$$\pi_i^m(x) = (x(k/m))_{k=0}^{m^2},$$

and the corresponding interpolation map $\eta_i^m : S_i^{m^2+1} \to D_i$ by

$$\eta_i^m(\alpha)(t) = \begin{cases} \alpha_k, & t \in [k/m, (k+1)/m), \quad k = 0, 1, \ldots, m^2 - 1, \\ \alpha_{m^2}, & t \in (t, \infty). \end{cases}$$

The functions π_i^m and η_i^m are measurable with respect to the Borel σ-algebras on D_i and $S_i^{m^2+1}$ [9, Proposition 3.7.1]. Let $(\hat{X}_1^m, \hat{X}_2^m)$ be a coupling of $\pi_1^m(X_1)$ and $\pi_2^m(X_2)$ such that $(\hat{X}_1^m, \hat{X}_2^m) \in R^{m^2+1}$ almost surely. Then $(\eta_1^m(\hat{X}_1^m), \eta_2^m(\hat{X}_2^m))$ couples $\eta_1^m \circ \pi_1^m(X_1)$ and $\eta_2^m \circ \pi_2^m(X_2)$, and moreover, $(\eta_1^m(\hat{X}_1^m), \eta_2^m(\hat{X}_2^m)) \in R^D$ almost surely. Hence

$$\eta_1^m \circ \pi_1^m(X_1) \sim_{\text{st}} \eta_2^m \circ \pi_2^m(X_2).$$

Because $\eta_i^m \circ \pi_i^m(x_i)$ converges to x_i in D_i as $m \to \infty$ for all $x_i \in D_i$ ([9, Problem 3.12], [1, Lemma 3]), it follows that $\eta_i^m \circ \pi_i^m(X_i) \xrightarrow{w} X_i$. Lemmas 3.9 and 4.5 show that R_{st}^D is closed in the weak topology, so that $X_1 \sim_{\text{st}} X_2$. □

4.3 Markov processes

In the sequel, the notation $X(\mu, t)$ refers to the state of a Markov process X at time t with initial distribution μ, and we shall use $X(x, t)$ as shorthand for $X(\delta_x, t)$. Markov processes X_1 and X_2 are said to *stochastically preserve* a relation R, if for all t,

$$x_1 \sim x_2 \implies X_1(x_1, t) \sim_{\text{st}} X_2(x_2, t),$$

or equivalently (see Theorem 3.1),

$$\mu_1 \sim_{\text{st}} \mu_2 \implies X_1(\mu_1, t) \sim_{\text{st}} X_2(\mu_2, t).$$

The following theorem presents a simple but powerful result, which together with the subrelation algorithm (see Theorem 4.9 below) provides a method for stochastically relating (potentially unknown) stationary distributions of Markov processes based on their generators.

Theorem 4.7. *Let X_1 and X_2 be Markov processes with stationary distributions μ_1 and μ_2 such that $X_i(x_i, t) \xrightarrow{w} \mu_i$ as $t \to \infty$ for all initial states x_i. Given any measurable relation R, a sufficient condition for $\mu_1 \sim_{\mathrm{st}} \mu_2$ is that X_1 and X_2 stochastically preserve some nontrivial closed subrelation of R.*

Proof. Choose a pair of initial states $(x_1, x_2) \in R'$, where R' is a closed subrelation of R stochastically preserved by X_1 and X_2. Then $X_1(x_1, t)$ and $X_2(x_2, t)$ are stochastically related with respect to R' for all t, so by Lemma 3.9 we see that $(\mu_1, \mu_2) \in R'_{\mathrm{st}}$. Because $R'_{\mathrm{st}} \subset R_{\mathrm{st}}$, it follows that $(\mu_1, \mu_2) \in R_{\mathrm{st}}$. □

Let X_1 and X_2 be discrete-time Markov processes with transition probability kernels P_1 and P_2, respectively. The following result characterizes precisely when X_1 and X_2 stochastically preserve a relation R. A Markov process $\hat{X}$ taking values in $S_1 \times S_2$ is called a *Markovian coupling* of X_1 and X_2, if $\hat{X}(x, t)$ couples $X_1(x_1, t)$ and $X_2(x_2, t)$ for all t and all $x = (x_1, x_2)$. A measurable set B is called *invariant* for a Markov process X, if $x \in B$ implies $X(x, t) \in B$ for all t almost surely.

Theorem 4.8. *The following are equivalent:*

(i) X_1 and X_2 stochastically preserve the relation R.

(ii) $P_1(x_1, B) \le P_2(x_2, B^{\rightarrow})$ for all $x_1 \sim x_2$ and compact $B \subset S_1$.

(iii) P_1 and P_2 stochastically preserve the relation R.

(iv) There is a Markovian coupling of X_1 and X_2 for which R is invariant.

Proof. The implications (iv) $\Longrightarrow$ (i) $\Longrightarrow$ (iii) are direct consequences of the definitions, while (ii) $\Longleftrightarrow$ (iii) follows by Theorem 3.1. For (iii) $\Longrightarrow$ (iv), observe that Theorem 3.3 implies the existence of a coupling P of the probability kernels P_1 and P_2 such that $P(x, R) = 1$ for all $x \in R$. Let $\hat{X}$ be a discrete-time Markov process with transition probability kernel P. Induction then shows that $\hat{X}$ is a Markovian coupling of X_1 and X_2 for which R is invariant. □

Theorems 3.8 and 4.8 yield the following characterization for subrelations of a closed relation R that are stochastically preserved by discrete-time Markov processes X_1 and X_2 with continuous transition probability kernels P_1 and P_2. Denote by R^* the output (3.1) of the subrelation algorithm in Section 3.2.

Theorem 4.9. *R^* is the maximal closed subrelation of R that is stochastically preserved by X_1 and X_2. Especially, X_1 and X_2 stochastically preserve a nontrivial closed subrelation of R if and only if $R^* \neq \emptyset$.*

Markov jump processes shall be consider next. Recall that a map $Q : S \times \mathcal{B}(S) \to \mathbb{R}_+$ is called a *rate kernel* on S, if $Q(x, dy) = q(x)P(x, dy)$ for some probability kernel P and a positive measurable function q. A rate kernel Q is called *nonexplosive*, if the standard construction using a discrete-time Markov process with transition probability kernel P generates a Markov jump process with paths in $D(\mathbb{R}_+, S)$ [11, Theorem 12.18].

Theorem 4.10 below characterizes precisely when a pair of Markov jump processes X_1 and X_2 with nonexplosive rate kernels Q_1 and Q_2 stochastically preserves a closed relation R. The construction of the Markovian coupling is based on the local uniformization of the rate kernels Q_i using the probability kernels $\hat{P}_i$ from $S_1 \times S_2$ to S_i, defined by

$$\hat{P}_i(x, B_i) = \frac{q_i(x_i)}{q(x)} P_i(x_i, B_i) + \left(1 - \frac{q_i(x_i)}{q(x)}\right) \delta(x_i, B_i), \tag{4.1}$$

where $q(x) = 1 + q_1(x_1) + q_2(x_2)$ [17, Section 3].

Theorem 4.10. *The following are equivalent:*

(i) X_1 *and* X_2 *stochastically preserve the relation* R.

(ii) *For all* $x_1 \sim x_2$ *and compact* $B \subset S_1$ *such that* $\delta(x_1, B) = \delta(x_2, B^{\to})$,

$$Q_1(x_1, B) - q_1(x_1)\delta(x_1, B) \le Q_2(x_2, B^{\to}) - q_2(x_2)\delta(x_2, B^{\to}).$$

(iii) *The probability kernels in (4.1) satisfy* $\hat{P}_1(x, \cdot) \sim_{\mathrm{st}} \hat{P}_2(x, \cdot)$ *for all* $x \in R$.

(iv) *There is a Markovian coupling of* X_1 *and* X_2 *for which* R *is invariant.*

Countable spaces admit the following slightly more convenient characterization, due to the fact that all sets are measurable.

Theorem 4.11. *If the spaces* S_1 *and* S_2 *are countable, then the properties of Theorem 4.10 are equivalent to requiring that for all* $x_1 \sim x_2$*:*

$$Q_1(x_1, B_1) \le Q_2(x_2, B_1^{\to}) \tag{4.2}$$

for all $B_1 \subset S_1$ *such that* $x_1 \notin B_1$ *and* $x_2 \notin B_1^{\to}$, *and*

$$Q_1(x_1, B_2^{\leftarrow}) \ge Q_2(x_2, B_2) \tag{4.3}$$

for all $B_2 \subset S_2$ *such that* $x_1 \notin B_2^{\leftarrow}$ *and* $x_2 \notin B_2$.

Remark 4.12. For order relations on countable spaces it suffices to verify (4.2) for all upper sets B_1 and (4.3) for all lower sets B_2 (see Remark 2.6), so Theorem 4.11 becomes equivalent to Massey's characterization [20, Theorem 3.4].

To prove Theorem 4.10 we need the following special form of Theorem 3.3 to account for the fact that $\hat{P}_i$ are probability kernels from $S_1 \times S_2$ to S_i, and not from S_i to S_i.

Lemma 4.13. *Let P_i be probability kernels from $S_1 \times S_2$ to S_i such that $P_1(x,\cdot) \sim_{\rm st} P_2(x,\cdot)$ for all $x \in R$. Then there exists a probability kernel P on $S_1 \times S_2$ such that $P(x,\cdot)$ couples $P_1(x,\cdot)$ and $P_2(x,\cdot)$ for all $x \in S_1 \times S_2$, and $P(x,R) = 1$ for all $x \in R$.*

Proof. Denote $\hat S_i = S_1 \times S_2$ and $\hat S_i' = S_i$, $i = 1,2$, and define $\hat R = \{(x,x) : x \in R\}$ and $\hat R' = R$. Then by Theorem 3.3 there exists a probability kernel $\hat P$ from $\hat S_1 \times \hat S_2$ to $\hat S_1' \times \hat S_2'$ such that $\hat P((x,y),\cdot)$ couples $P_1(x,\cdot)$ and $P_2(y,\cdot)$ for all $x \in S_1 \times S_2$ and $y \in S_1 \times S_2$, and $\hat P((x,x),R) = 1$ for all $x \in R$. Define $P(x,B) = \hat P((x,x),B)$ for all $x \in S_1 \times S_2$ and measurable $B \subset S_1 \times S_2$. □

Proof of Theorem 4.10. (i) $\implies$ (ii). Choose $x_1 \sim x_2$ and a compact $B \subset S_1$ such that $\delta(x_1,B) = \delta(x_2,B^{\to})$. Then Theorem 2.5 shows that

$$\mathrm{P}(X_1(x_1,t) \in B) - \delta(x_1,B) \le \mathrm{P}(X_2(x_2,t) \in B^{\to}) - \delta(x_2,B^{\to})$$

for all t. Dividing both sides above by t, and taking $t \downarrow 0$, we thus see using Kolmogorov's backward equation [11, Theorem 12.25] the validity of (ii).

(ii) $\implies$ (iii). Observe that for any compact $B \subset S_1$,

$$\begin{aligned}\hat P_2(x,B^{\to}) - \hat P_1(x,B) = q(x)^{-1}\{&(Q_2(x_2,B^{\to}) - q(x_2)\delta(x_2,B^{\to}))\\ &- (Q_1(x_1,B) - q(x_1)\delta(x_1,B))\}\\ &+ \delta(x_2,B^{\to}) - \delta(x_1,B).\end{aligned}$$

Because $\delta(x_2,B^{\to}) \ge \delta(x_1,B)$ for all $x_1 \sim x_2$, we see that $\hat P_1(x,B) \le \hat P_2(x,B^{\to})$ for all $x \in R$. Hence using Theorem 2.5 we may conclude that $\hat P_1(x,\cdot) \sim_{\rm st} \hat P_2(x,\cdot)$ for all $x \in R$.

(iii) $\implies$ (iv). Let $\hat P_i$ be the probability kernels defined in (4.1). Lemma 4.13 then shows the existence of a probability kernel P on $S_1 \times S_2$ such that $P(x,\cdot)$ couples $\hat P_1(x,\cdot)$ and $\hat P_2(x,\cdot)$ for all $x \in S_1 \times S_2$, and $P(x,R) = 1$ for all $x \in R$. Define a rate kernel Q on $S_1 \times S_2$ by $Q(x,B) = q(x)P(x,B)$. Then

$$\int_{S_1 \times S_2} (f_i(y_i) - f_i(x_i))\, Q(x,dy) = \int_{S_i} (f_i(y_i) - f_i(x_i))\, Q_i(x_i,dy_i)$$

for all $x \in S_1 \times S_2$ and all bounded measurable f on S_i, $i = 1,2$, which implies that Q is nonexplosive (Chen [4, Theorem 37]). Let $\hat X$ be a Markov jump process generated by Q using the standard construction [11, Theorem 12.18]. Then $\hat X(x,t)$ couples $X_1(x_1,t)$ and $X_2(x_2,t)$ for all $x \in S_1 \times S_2$ and for all t (Chen [4, Theorem 13]). Moreover, R is invariant for $\hat X$, because $P(x,R) = 1$ for all $x \in R$.

(iv) $\implies$ (i). Clear by definition. □

Proof of Theorem 4.11. Assume first that (4.2) and (4.3) hold, and choose $x_1 \sim x_2$ and $B_1 \subset S_1$ so that $\delta(x_1,B_1) = \delta(x_2,B_1^{\to})$. If $x_1 \notin B_1$, and $x_2 \notin B_1^{\to}$, then the validity of Theorem 4.10:(ii) is obvious from (4.2). In the other case where $x_1 \in B_1$, and $x_2 \in B_1^{\to}$, denote $B_2 = (B_1^{\to})^c$. Then it follows that $B_2^{\leftarrow} \subset B_1^c$, and thus $x_1 \notin B_2^{\leftarrow}$. This further implies that $x_2 \notin B_2$, because $x_1 \sim x_2$. Hence (4.3) shows that

$$Q_2(x_2, B_2) \le Q_1(x_1, B_2^{\leftarrow}) \le Q_1(x_1, B_1^c),$$

from which we again see that Theorem 4.10:(ii) is valid.

Assume next that Theorem 4.10:(iv) holds, and let $x_1 \sim x_2$ be such that $x_1 \notin B_1$ and $x_2 \notin B_1^{\rightarrow}$. Then (4.2) follows by Theorem 2.5, because $Q_1(x_1, B_1) = \hat{P}_1(x, B_1)$ and $Q_2(x_2, B_1^{\rightarrow}) = \hat{P}_2(x_2, B_1^{\rightarrow})$. Inequality (4.3) can be verified by a symmetrical argument. □

This section is concluded by an analogue of Theorem 4.9. A rate kernel $Q(x, dy) = q(x)P(x, dy)$ such that q is a continuous function and P is a continuous probability kernel shall be called *continuous*. Given Markov jump processes X_1 and X_2 with continuous nonexplosive rate kernels Q_1 and Q_2, define the relations $R^{(n)}$ by $R^{(0)} = R$, and

$$R^{(n+1)} = \left\{ x \in R^{(n)} : \; (\hat{P}_1(x, \cdot), \hat{P}_2(x, \cdot)) \in R^{(n)}_{\mathrm{st}} \right\}, \tag{4.4}$$

where $\hat{P}_1$ and $\hat{P}_2$ are given by (4.1). Moreover, denote $R^* = \cap_{n=0}^{\infty} R^{(n)}$, as in Section 3.2.

Theorem 4.14. *R^* is the maximal closed subrelation of R that is stochastically preserved by X_1 and X_2. Especially, X_1 and X_2 stochastically preserve a nontrivial closed subrelation of R if and only if $R^* \neq \emptyset$.*

Proof. The continuity of Q_1 and Q_2 guarantees the continuity of $\hat{P}_1$ and $\hat{P}_2$. The proof is hence completed by repeating the steps in the proof of Theorem 3.8, with notational modifications to take into account that here $\hat{P}_i$ are probability kernels from $S_1 \times S_2$ to S_i, and not from S_i to S_i. □

5 Applications

5.1 Hidden Markov processes

The goal of this section is to stochastically compare two hidden Markov processes Y_1 and Y_2 of the form $Y_i = f_i \circ X_i$, where X_i is a Markov process taking values in S_i', and f_i is a continuous function from S_i' to S_i. Although the results in this section have natural counterparts for Markov jump processes, we shall only treat the case where X_i are discrete-time Markov processes with transition probability kernels P_i.

Theorem 5.1. *Let R be a closed relation between S_1 and S_2, and assume that*

$$P_1(x_1, f_1^{-1}(B)) \le P_2(x_2, f_2^{-1}(B^{\rightarrow})) \tag{5.1}$$

for all x_1 and x_2 such that $f_1(x_1) \sim f_2(x_2)$ and all compact $B \subset S_1$. Then $Y_1(t) \sim_{\mathrm{st}} Y_2(t)$ whenever $Y_1(0) \sim_{\mathrm{st}} Y_2(0)$.

The proof of Theorem 5.1 relies on the notion of induced relation, defined as follows. Given a closed relation R between S_1 and S_2, and two continuous functions $\phi_i : S_i' \to S_i$, $i = 1, 2$, we denote

$$R' = \{(x_1', x_2') \in S_1' \times S_2' : (\phi_1(x_1'), \phi_2(x_2')) \in R\}.$$

The closed relation R' is said to be *induced from* R by the pair (ϕ_1, ϕ_2), and we denote by R'_{st} the stochastic relation generated by R'.

Lemma 5.2. *Probability measures μ_1 on S_1' and μ_2 on S_2' satisfy $(\mu_1, \mu_2) \in R'_{\mathrm{st}}$ if and only if $(\mu_1 \circ \phi_1^{-1}, \mu_2 \circ \phi_2^{-1}) \in R_{\mathrm{st}}$.*

Proof. Observe that $R' = \phi^{-1}(R)$, where the function ϕ is defined by $\phi(x_1, x_2) = (\phi_1(x_1), \phi_2(x_2))$. If $(\mu_1, \mu_2) \in R'_{\mathrm{st}}$, then let λ be a coupling of μ_1 and μ_2 such that $\lambda(R') = 1$. Then $\lambda \circ \phi^{-1}$ couples the probability measures $\mu_1 \circ \phi_1^{-1}$ and $\mu_2 \circ \phi_2^{-1}$, and moreover, that $\lambda \circ \phi^{-1}(R) = 1$.

Assume next that $(\nu_1, \nu_2) \in R_{\mathrm{st}}$, where $\nu_i = \mu_i \circ \phi_i^{-1}$. Choose a compact $B' \subset S_1'$, and observe that the right conjugate of B' with respect to R' equals

$$[B']^{\rightarrow}_{R'} = \phi_2^{-1}([\phi_1(B')]^{\rightarrow}_{R}),$$

where $[\phi_1(B')]^{\rightarrow}_{R}$ denotes the right conjugate of $\phi_1(B')$ with respect to R. Moreover, because $\phi_1(B')$ is compact, Theorem 2.5 shows that

$$\mu_1(B') = \nu_1(\phi_1(B')) \le \nu_2([\phi_1(B')]^{\rightarrow}_{R}) = \mu_2([B']^{\rightarrow}_{R'}),$$

and hence $(\mu_1, \mu_2) \in R'_{\mathrm{st}}$. ☐

Proof of Theorem 5.1. Let R' be the relation between S_1' and S_2' induced from R by the pair (f_1, f_2). Lemma 5.2 then shows that for all t,

$$Y_1(t) \overset{R}{\sim}_{\mathrm{st}} Y_2(t) \quad \text{if and only if} \quad X_1(t) \overset{R'}{\sim}_{\mathrm{st}} X_2(t),$$

and moreover, (5.1) is equivalent to requiring that the pair (P_1, P_2) stochastically preserves R'. ☐

Example 5.3 (Non-Markov processes)**.** Let Y_1 and Y_2 be non-Markov processes with values in S_1 and S_2, respectively. Following Whitt [29], let us assume that Y_i can be made Markov by keeping track of additional information, say a random processes Z_i with values in S_i'. Under this assumption, $Y_i = \pi_i \circ X_i$, where $X_i = (Y_i, Z_i)$ is a Markov process taking values $S_i \times S_i'$, and π_i denotes the projection from $S_i \times S_i'$ onto S_i. Denoting the transition probability kernel of X_i by P_i, condition (5.1) in Theorem 5.1 becomes equivalent to

$$\sup_{z_1 \in S_1'} P_1((y_1, z_1), B \times S_1') \le \inf_{z_2 \in S_2'} P_2((y_2, z_2), B^{\rightarrow} \times S_2') \qquad (5.2)$$

for all $y_1 \sim y_2$ and all compact $B \subset S_1$. Inequality (5.2) together with $Y_1(0) \sim_{\mathrm{st}} Y_2(0)$ is thus sufficient for $Y_1(t) \sim_{\mathrm{st}} Y_2(t)$ for all t. This formulation is conceptually similar, though not equivalent, to [29, Theorem 1].

Example 5.4 (Lumpability)**.** A Markov process X with values in S_1 is called *lumpable* with respect to $f : S_1 \to S_2$, if $f \circ X$ is Markov for any initial distribution of X [13]. It is well known [6] that X is lumpable if and only if its transition probability kernel satisfies

$$P(x, f^{-1}(B)) = P(y, f^{-1}(B))$$

for all x and y such that $f(x) = f(y)$ and all measurable $B \subset S_2$. This is equivalent to saying that the pair (P, P) stochastically preserves the relation $\{(x, y) \in S_1 \times S_1 : f(x) = f(y)\}$ induced by (f, f) from the equality on S_2. Moreover, if X is lumpable with respect to f, then the pair (P, P') stochastically preserves the relation $\{(x, y) \in S_1 \times S_2 : f(x) = y\}$ induced from the equality on S_2 by the pair (f, χ), where P' denotes the transition probability kernel of $f \circ X$, and χ is the identity map on S_2. The notion of lumpability may be generalized by calling X lumpable with respect to a relation R, if there exists a Markov process Y such that X and Y stochastically preserve R [18].

Example 5.5 (Stochastic induced order)**.** For order relations, condition (5.1) in Theorem 5.1 can be rephrased as

$$P_1(x_1, f_1^{-1}(B)) \le P_2(x_2, f_2^{-1}(B))$$

for all upper sets $B \subset S$ and all x_1 and x_2 such that $f_1(x_1) \le f_2(x_2)$ (see Remark 2.6). This is a discrete-time analogue to [5, Theorem 6].

5.2 Population processes

Let us denote by $e_1, \ldots, e_m$ the unit vectors of $\mathbb{Z}^m$, and define $e_0 = 0$ and $e_{i,j} = -e_i + e_j$ for notational convenience. A *Markov population process* [14] is a nonexplosive Markov jump process taking values[2] in $S \subset \mathbb{Z}^m$, generated by the transitions

$$x \mapsto x + e_{i,j} \quad \text{at rate } \alpha_{i,j}(x), \quad i, j \in \{0, 1, \ldots, m\}, \quad i \neq j.$$

where $\alpha_{i,j}$ are positive functions on S such that $\alpha_{i,j}(x) = 0$ for $x + e_{i,j} \notin S$. The functions $\alpha_{0,i}$ and $\alpha_{i,0}$ may be regarded as the arrival and departure rates of individuals for colony i, and $\alpha_{i,j}$ represents the transfer rate of individuals from colony i to colony j. Population processes with values in $S \subset \mathbb{Z}_+^m$ may be viewed as Markovian queueing networks [22] or interacting particle systems [16].

The following result characterizes precisely when two population processes stochastically preserve a relation R between $S \subset \mathbb{Z}^m$ and $S' \subset \mathbb{Z}^{m'}$. To state the result, define for $x \in S$ and $y \in S'$, and for sets of index pairs $U \subset \{0, \ldots, m\}^2$ and $V \subset \{0, \ldots, m'\}^2$,

$$U_{\to}(x, y) = \{(k, l) : x + e_{i,j} \sim y + e_{k,l} \text{ for some } (i, j) \in U\}$$

[2] Often it is natural to assume that the elements of S have positive coordinates, but in this section there is no need to make this restriction.

and

$$V_{\leftarrow}(x,y) = \{(i,j) : x + e_{i,j} \sim y + e_{k,l} \text{ for some } (k,l) \in V\}.$$

Theorem 5.6. *Let X and X' be population processes taking values in $S \subset \mathbb{Z}^m$ and $S' \subset \mathbb{Z}^{m'}$ generated by transition rate functions $\alpha_{i,j}$ and $\alpha'_{k,l}$, respectively. Then X and X' stochastically preserve a relation R if and only if for all $x \sim y$:*

$$\sum_{(i,j)\in U} \alpha_{i,j}(x) \le \sum_{(k,l)\in U_{\rightarrow}(x,y)} \alpha'_{k,l}(y) \tag{5.3}$$

for all $U \subset \{(i,j) : \ x + e_{i,j} \not\sim y\}$, and

$$\sum_{(i,j)\in V_{\leftarrow}(x,y)} \alpha_{i,j}(x) \ge \sum_{(k,l)\in V} \alpha'_{k,l}(y) \tag{5.4}$$

for all $V \subset \{(k,l) : \ x \not\sim y + e_{k,l}\}$.

Proof. Observe that for any $x \sim y$ and $B \subset S$ such that $x \notin B$ and $y \notin B^{\rightarrow}$, the rate kernels of X and X' satisfy

$$Q(x,B) = \sum_{(i,j)\in U} \alpha_{i,j}(x) \quad \text{and} \quad Q'(y,B^{\rightarrow}) = \sum_{(k,l)\in U_{\rightarrow}(x,y)} \alpha'_{k,l}(y),$$

where $U = \{(i,j) : x + e_{i,j} \in B\}$. Moreover, $x + e_{i,j} \not\sim y$ for all $(i,j) \in U$, because $y \notin B^{\rightarrow}$. Hence inequality (5.3) is equivalent to (4.2) in Theorem 4.11. By symmetry, we see that (5.4) is equivalent (4.3). □

Example 5.7 (Partial coordinatewise order)**.** Define a relation between S and S' by denoting $x \le_M y$, if $x_i \le y_i$ for all $i \in M$, where M is a fixed subset of $\{1,\dots,m\} \cap \{1,\dots,m'\}$. We shall next show that X and X' stochastically preserve the relation $\le_M$ if and only if for all $x \le_M y$ and all k in the set $M_0(x,y) = \{i \in M : x_i = y_i\}$:

$$\sum_{i\in I} \alpha_{i,k}(x) \le \sum_{i\in I\cup([0,m']\setminus M_0(x,y))} \alpha'_{i,k}(y) \tag{5.5}$$

for all $I \subset [0,m] \setminus \{k\}$, and

$$\sum_{j\in J\cup([0,m]\setminus M_0(x,y))} \alpha_{k,j}(x) \ge \sum_{j\in J} \alpha'_{k,j}(y) \tag{5.6}$$

for all $J \subset [0,m'] \setminus \{k\}$.

To justify the above claim, observe that for any $x \le_M y$ and U is as in Theorem 5.6, we can write $U = \cup_{k\in K}(I_k \times \{k\})$ for some $I_k \subset [0,M] \setminus \{k\}$ and $K \subset M_0(x,y)$. Hence $U_{\rightarrow}(x,y) = \cup_{k\in K}(J_k \times \{k\})$, where $U_{\rightarrow}(x,y) = \cup_{k\in K}(I_k \cup ([0,m'] \setminus M_0(x,y))) \times \{k\})$. By summing both sides of (5.5) over $k \in K$, with I_k in place of I, we see that (5.3) holds. On the other hand, it is easy to see that (5.3) implies (5.5), and a symmetric argument shows the equivalence of (5.4) and (5.6).

The above characterization extends and sharpens earlier comparison results for Markovian queueing networks [7, 16, 17]. For population processes where transfers between colonies do not occur, (5.5) and (5.6) simplify to

$$\alpha_{0,k}(x) \le \alpha'_{0,k}(y) \quad \text{and} \quad \alpha_{k,0}(x) \ge \alpha'_{k,0}(y)$$

for all $x \le_M y$ and all $k \in M$ such that $x_k = y_k$ [2, Lemma 1].

5.3 Parallel queueing system

Consider a system of two queues in parallel, where customers arrive to queue k at rate $\lambda_k \in (0,1)$ and have unit service rate. Assuming all interarrival and service times are exponential, the queue length process $X = (X_1, X_2)$ is a Markov population process on $\mathbb{Z}_+^2$ with transition rates $\alpha_{0,k}(x) = \lambda_k$ and $\alpha_{k,0}(x) = 1(x_k > 0)$, $k = 1, 2$. We shall also consider a modification of the system, where load is balanced by routing incoming traffic to the shortest queue, modeled as a Markov population process $X^{\mathrm{LB}} = (X_1^{\mathrm{LB}}, X_2^{\mathrm{LB}})$ with transition rates

$$\begin{aligned}
\alpha_{0,1}^{\mathrm{LB}}(x) &= (\lambda_1 + \lambda_2) 1(x_1 < x_2) + \lambda_1 1(x_1 = x_2), \\
\alpha_{0,2}^{\mathrm{LB}}(x) &= (\lambda_1 + \lambda_2) 1(x_1 > x_2) + \lambda_2 1(x_1 = x_2),
\end{aligned}$$

and $\alpha_{k,0}^{\mathrm{LB}}(x) = \alpha_{k,0}(x)$ for $k = 1, 2$. Common sense suggests that load balancing decreases the total number of customers in the system, so that

$$X_1^{\mathrm{LB}}(t) + X_2^{\mathrm{LB}}(t) \le_{\mathrm{st}} X_1(t) + X_2(t). \tag{5.7}$$

However, the justification of (5.7) appears difficult, because using Theorem 5.6 can check that the processes X^{LB} and X do not stochastically preserve the coordinatewise order on $\mathbb{Z}_+^2$, nor the order

$$R^{\mathrm{sum}} = \{(x, y) : |x| \le |y|\},$$

where $|x| = x_1 + x_2$. On the other hand, it is known [30] that (5.7) holds for all t, whenever $X^{\mathrm{LB}}(0) = X(0)$, which suggests that X^{LB} and X might stochastically preserve some strict subrelation of R^{sum}. The following theorem summarizes the output of the subrelation algorithm applied to the rate kernels of X^{LB} and X.

Theorem 5.8. *Starting from $R^{(0)} = R^{\mathrm{sum}}$, the subrelation iteration (4.4) produces the sequence of relations*

$$R^{(n)} = \left\{ (x, y) : |x| \le |y| \text{ and } x_1 \vee x_2 \le y_1 \vee y_2 + (y_1 \wedge y_2 - n)^+ \right\}, \tag{5.8}$$

which converges to

$$R^* = \{(x, y) : |x| \le |y| \text{ and } x_1 \vee x_2 \le y_1 \vee y_2\}.$$

The limiting relation R^* may be identified as the weak majorization order $\preceq^{\mathrm{wm}}$ on $\mathbb{Z}_+^2$ (Example 2.3). As a consequence,

$$X^{\mathrm{LB}}(0) \preceq^{\mathrm{wm}} X(0) \implies X^{\mathrm{LB}}(t) \preceq^{\mathrm{wm}}_{\mathrm{st}} X(t) \quad \text{for all } t.$$

Especially, $X^{\mathrm{LB}}(0) \preceq^{\mathrm{wm}} X(0)$ implies (5.7), and moreover,

$$X_1^{\mathrm{LB}}(t) \vee X_2^{\mathrm{LB}}(t) \leq_{\mathrm{st}} X_1(t) \vee X_2(t),$$

which indicates that the queue lengths corresponding to X^{LB} are more balanced than those corresponding to X. The proof of Theorem 5.8 is based on the following lemma, the proof of which is omitted.

Lemma 5.9. *The function $\alpha_n(x) = x_1 \vee x_2 \vee (|x| - n)$ on $\mathbb{Z}^2$ satisfies:*

(i) $\alpha_n(x) = |x| - x_1 \wedge x_2 \wedge n$.

(ii) $\alpha_n(x - e_k) \leq \alpha_{n+1}(x)$ *for all* k.

(iii) $\alpha_n(x + e_1) > \alpha_n(x + e_2)$ *if and only if* $x_1 > x_2 \vee (|x| - n)$.

(iv) $\alpha_n(x - e_1) < \alpha_n(x - e_2)$ *if and only if* $x_1 > x_2 \vee (|x| - n - 1)$.

(v) $\alpha_n(x + e_k) > \alpha_n(x)$ *for* k *such that* $x_k = x_1 \vee x_2$.

(vi) $\alpha_n(x - e_k) = \alpha_{n+1}(x)$ *for* k *such that* $x_k = x_1 \wedge x_2$.

Proof of Theorem 5.8. Define the function $\alpha_n(x)$ as in Lemma 5.9. Then the relations $R^{(n)}$ defined in (5.8) can be written as

$$R^{(n)} = \{(x, y) : |x| \leq |y|,\ \alpha_n(x) \leq \alpha_n(y)\},$$

which shows that $R^{(n)}$ is an order for all n. To show that $R^{(n)}$ is the sequence of relations produced by iteration (4.4), it is sufficient (Remark 4.12) to show that the properties

- $Q^{\mathrm{LB}}(x, U) \leq Q(y, U)$ for all $R^{(n)}$-upper sets U such that $x, y \notin U$,
- $Q^{\mathrm{LB}}(x, U) \geq Q(y, U)$ for all $R^{(n)}$-lower sets U such that $x, y \notin U$,

are valid for all $(x, y) \in R^{(n+1)}$, and that at least one of the above properties fails for $(x, y) \in R^{(n)} \setminus R^{(n+1)}$.

Assume $(x, y) \in R^{(n+1)}$ (actually, $R^{(n)}$ is here enough), and let U be an $R^{(n)}$-upper set such that $x, y \notin U$. Then $x - e_k, y - e_k \notin U$ for all k.

(i) If $y + e_1, y + e_2 \in U$, then $Q^{\mathrm{LB}}(x, U) \leq \lambda_1 + \lambda_2 = Q(y, U)$.

(ii) Assume $y + e_1, y + e_2 \notin U$, and choose l so that $y_l = y_1 \vee y_2$. Then Lemma 5.9:(v) together with $\alpha_n(x) \leq \alpha_n(y)$ imply that $\alpha_n(x + e_k) \leq \alpha_n(x) + 1 \leq \alpha_n(y + e_l)$ for all k. Hence $x + e_k \notin U$ for all k, because U is $R^{(n)}$-upper, so that $Q(x, U) = 0 = Q^{\mathrm{LB}}(y, U)$.

(iii) Assume $y + e_1 \in U, y + e_2 \notin U$. Then $\alpha_n(y + e_1) > \alpha_n(y + e_2)$, so Lemma 5.9:(iii) shows that $y_1 > y_2$ and $y_1 > |y| - n$. Now if $x + e_k \in U$, then $\alpha_n(x + e_k) > \alpha_n(y + e_2) = \alpha_n(y)$. Further $\alpha_n(x) \leq \alpha_n(y)$ shows that $\alpha_n(x) = \alpha_n(y)$, so that $\alpha_n(x) = y_1 > |x| - n$. Because $\alpha_n(x + e_k) > \alpha_n(x) > |x| - n$, it follows that $x_k = x_1 \vee x_2 = y_1$. As a consequence, $|x| \leq |y|$ implies that $x_1 \wedge x_2 \leq y_2 < y_1$, so that $x_k > x_1 \wedge x_2$. Hence $\alpha_k^{\mathrm{LB}}(x) = 0$, which implies that $Q^{\mathrm{LB}}(x, U) = 0 \leq Q(y, U)$. The case where $y + e_1 \notin U, y + e_1 \in U$ is similar.

Assume next that $(x, y) \in R^{(n+1)}$, and let U be an $R^{(n)}$-lower set such that $x, y \notin U$. Then $x + e_k, y + e_k \notin U$ for all k.

(i) Assume $y - e_1 \in U$, $y - e_2 \notin U$. Then $y_2 = 0$ or $\alpha_n(y - e_1) < \alpha_n(y - e_2)$, so with the help of Lemma 5.9:(iv), we see that $y_1 > y_2 \vee (|y| - n - 1)$. Choose k so that $x_k = x_1 \vee x_2$, and observe that

$(x, y - e_1) \notin R^{(n)}$ implies that either $|x| = |y|$ or $\alpha_n(x) \geq \alpha_n(y)$, so that $x_k > 0$. If $x_1 \neq x_2$, then $\alpha_n(x - e_k) = \alpha_n(x) - 1 \leq \alpha_n(y) - 1 = \alpha_n(y - e_1)$. If $x_1 = x_2$, then $|x| \leq |y|$ together with $y_1 > y_2$ implies that $x_1 < y_1$. Hence $\alpha_n(x - e_k) = x_1 \vee (|x| - n - 1) \leq (y_1 - 1) \vee (|y| - n - 1) = \alpha_n(y - e_1)$. Thus, we may conclude that $x - e_k \in U$, which shows that $Q^{\mathrm{LB}}(x, U) \geq 1 = Q(y, U)$. By symmetry, the same conclusion holds under the assumption $y - e_1 \notin U$, $y - e_2 \in U$.

(ii) Assume $y - e_1, y - e_2 \in U$, and choose l so that $y_l = y_1 \wedge y_2$. Then using Lemma 5.9:(ii) and Lemma 5.9:(vi) we find that for all k,

$$\alpha_n(x - e_k) \leq \alpha_{n+1}(x) \leq \alpha_{n+1}(y) = \alpha_n(y - e_l).$$

Further, if $|x| = |y|$, then Lemma 5.9:(i) together with $x_1 \vee x_2 \leq \alpha_{n+1}(x) \leq \alpha_{n+1}(y)$ shows that $x_1 \wedge x_2 = |x| - x_1 \vee x_2 \geq y_1 \wedge y_2 \wedge (n+1)$, so that $x_1 \wedge x_2 \geq 1$. On the other hand, if $|x| < |y|$, then $x \notin U$ implies that $\alpha_n(x) > \alpha_n(y - e_l) \geq \alpha_{n+1}(x)$, so it follows that $x_1 \vee x_2 < |x| - n$, which again shows that $x_1 \wedge x_2 \geq 1$. Hence $x - e_k \in U$ for all k, and we may conclude that $Q^{\mathrm{LB}}(x, U) = 2 = Q(y, U)$.

(iii) Assume $y - e_l \notin U$ for all l. Then $Q^{\mathrm{LB}}(x, U) \geq 0 = Q(y, U)$.

Finally, assume that $(x, y) \in R^{(n)} \setminus R^{(n+1)}$. Then

$$\alpha_{n+1}(y) < \alpha_{n+1}(x) \leq \alpha_n(x) \leq \alpha_n(y),$$

which implies that $y_1 \vee y_2 < |y| - n$ and $\alpha_n(x) = \alpha_n(y) = |y| - n$. Consider the $R^{(n)}$-lower set $U = \{z : |z| < |y|,\ \alpha_n(z) < \alpha_n(y)\}$. Then $x, y \notin U$, $x + e_k, y + e_k \notin U$ for all k, and $y - e_l \in U$ for all l. On the other hand, $\alpha_n(x) = |y| - n$ together with $|x| \leq |y|$ shows that $\alpha_n(x) = x_1 \vee x_2$, so that $\alpha_n(x - e_k) = \alpha_n(x)$ for some k. Especially, $x - e_k \notin U$, so that $Q^{\mathrm{LB}}(x, U) \leq 1 < 2 = Q(y, U)$. □

6 Conclusion

This paper presented a systematic study of stochastic relations, which naturally extend the notion of stochastic orders to relations between random variables and processes that may take values in different state spaces. The key points of the paper may be summarized by Theorem 3.8, which characterizes the existence of subrelations stochastically preserved by a pair of probability kernels, and Theorem 4.7, which underlines the relevance of subrelation techniques in stochastically comparing stationary distributions of Markov processes. Finite-state Markov processes and diffusions are two important classes of processes that were not discussed in the paper. In finite state spaces the subrelation algorithm converges in finite time, which calls for numerical analysis of the runtime. The analysis of stochastic relations for diffusion processes requires the identification of suitable test functions that behave well with respect to taking relational conjugates. These issues may be considered interesting topics for future research.

Acknowledgements

The main part of the work presented has been carried out at Centrum voor Wiskunde en Informatica (CWI) and Eindhoven University of Technology, the Netherlands. The research has been supported by the Dutch BSIK/BRICKS PDC2.1 project, Helsingin Sanomat Foundation, and the Academy of Finland.

References

[1] Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, second edition.

[2] Borst, S., Jonckheere, M., and Leskelä, L. (2007). Stability of parallel queueing systems with coupled service rates. Discrete Event Dyn. Syst., To appear.

[3] Brandt, A. and Last, G. (1994). On the pathwise comparison of jump processes driven by stochastic intensities. *Math. Nachr.*, 167:21–42.

[4] Chen, M.-F. (1986). Coupling for jump processes. *Acta Math. Sin.*, 2(2):123–136.

[5] Doisy, M. (2000). A coupling technique for stochastic comparison of functions of Markov processes. *J. Appl. Math. Decis. Sci.*, 4(1):39–64.

[6] Dynkin, E. B. (1965). *Markov processes Vol. I*. Springer.

[7] Economou, A. (2003). Necessary and sufficient conditions for the stochastic comparison of Jackson networks. *Probab. Eng. Inform. Sci.*, 17:143–151.

[8] Eifler, L. Q. (1976). Open mapping theorems for probability measures on metric spaces. *Pac. J. Math.*, 66:89–97.

[9] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley.

[10] Himmelberg, C. J. (1975). Measurable relations. *Fund. Math.*, 87:53–72.

[11] Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer, second edition.

[12] Kamae, T., Krengel, U., and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.*, 5(6):899–912.

[13] Kemeny, J. and Snell, J. (1960). *Finite Markov chains*. Princeton Universty Press.

[14] Kingman, J. F. C. (1969). Markov population processes. *J. Appl. Probab.*, 6:1–18.

[15] Kuratowski, K. and Ryll-Nardzewski, C. (1965). A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 13:397–403.

[16] Lindvall, T. (1992). *Lectures on the Coupling Method*. Wiley.

[17] López, F. J., Martínez, S., and Sanz, G. (2000). Stochastic domination and Markovian couplings. *Adv. Appl. Probab.*, 32:1064–1076.

[18] López, F. J. and Sanz, G. (2002). Markovian couplings staying in arbitrary subsets of the state space. *J. Appl. Probab.*, 39:197–212.

[19] Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press.

[20] Massey, W. A. (1987). Stochastic orderings for Markov processes on partially ordered spaces. *Math. Oper. Res.*, 12(2):350–367.

[21] Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley.

[22] Serfozo, R. (1999). *Introduction to Stochastic Networks*. Springer.

[23] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer Series in Statistics. Springer, New York.

[24] Srivastava, S. M. (1998). *A course on Borel sets*. Springer.

[25] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36(2):423–439.

[26] Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability.* Springer.

[27] Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration.* Springer.

[28] Wagner, D. H. (1977). Survey of measurable selection theorems. *SIAM J. Control Optimization*, 15(5):859–903.

[29] Whitt, W. (1986). Stochastic comparisons for non-Markov processes. *Math. Oper. Res.*, 11(4):608–618.

[30] Winston, W. (1977). Optimality of the shortest line discipline. *J. Appl. Probab.*, 14(1):181–189.

(continued from the back cover)

A548 Kalle Mikkola
Weakly coprime factorization, continuous-time systems, and strong-H^p and Nevanlinna fractions
August 2008

A547 Wolfgang Desch, Stig-Olof Londen
A generalization of an inequality by N. V. Krylov
June 2008

A546 Olavi Nevanlinna
Resolvent and polynomial numerical hull
May 2008

A545 Ruth Kaila
The integrated volatility implied by option prices, a Bayesian approach
April 2008

A544 Stig-Olof Londen, Hana Petzeltová
Convergence of solutions of a non-local phase-field system
March 2008

A543 Outi Elina Maasalo
Self-improving phenomena in the calculus of variations on metric spaces
February 2008

A542 Vladimir M. Miklyukov, Antti Rasila, Matti Vuorinen
Stagnation zones for A-harmonic functions on canonical domains
February 2008

A541 Teemu Lukkari
Nonlinear potential theory of elliptic equations with nonstandard growth
February 2008

A540 Riikka Korte
Geometric properties of metric measure spaces and Sobolev-type inequalities
January 2008

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The reports are available at *http://math.tkk.fi/reports/* .
The list of reports is continued inside the back cover.

A553 Rolf Stenberg
A nonstandard mixed finite element family
September 2008

A552 Janos Karatson, Sergey Korotov
A discrete maximum principle in Hilbert space with applications to nonlinear cooperative elliptic systems
August 2008

A551 István Faragó, Janos Karatson, Sergey Korotov
Discrete maximum principles for the FEM solution of some nonlinear parabolic problems
August 2008

A550 István Faragó, Róbert Horváth, Sergey Korotov
Discrete maximum principles for FE solutions of nonstationary diffusion-reaction problems with mixed boundary conditions
August 2008

A549 Antti Hannukainen, Sergey Korotov, Tomás Vejchodský
On weakening conditions for discrete maximum principles for linear finite element schemes
August 2008

ISBN 978-951-22-9591-3 (print)
ISBN 978-951-22-9592-0 (PDF)
ISSN 0784-3143 (print)
ISSN 1797-5867 (PDF)