DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

METHODS FOR COMPARISON OF MARKOV PROCESSES BY STOCHASTIC DOMINANCE

Edward J. Green
California Institute of Technology

Richard D. McKelvey
California Institute of Technology

Edward W. Packel
Lake Forest College

SOCIAL SCIENCE WORKING PAPER 395

July 1981

ABSTRACT

A technique is developed for proving existence and obtaining bounds for the concentration of a stationary distribution for a given Markov process on the basis of comparisons, via stochastic dominance, with a different Markov process, having a known stationary distribution.

METHODS FOR COMPARISON OF MARKOV PROCESSES BY STOCHASTIC DOMINANCE

Edward J. Green, Richard D. McKelvey, and Edward W. Packel

1. Introduction

This paper provides a technique for proving existence and obtaining bounds for the concentration of a stationary distribution for a given Markov process on the basis of comparisons, via stochastic dominance, with a different Markov process, having a known stationary distribution. Thus, given a particular Markov process $\bar{p}$ on the measurable sets of X, we look for a nonnegative real-valued function L on X, and another Markov process $\bar{q}$ on X, such that for every $x \in X$, the transition measure $\bar{p}_x$ generated by $\bar{p}$ always concentrates at least as much of the distribution in each lower contour set of L as does the transition measure $\bar{q}_x$. This by itself is not enough to guarantee a limiting distribution for $\bar{p}$. However if $\bar{p}$ is sufficiently smooth, and if $\bar{q}$ satisfies an additional condition which we call ''stochastically increasing'' with respect to L, then if $\bar{q}$ has a stationary distribution, so does $\bar{p}$, and further, the limiting distribution for $\bar{p}$ is more concentrated (with respect to L) than is that of $\bar{q}$.

2. The Main Results

Let X be a measurable topological space and $\underline{X}$ be a σ - Algebra on X. Probability measures on X are denoted p, q, etc., i.e. $p : \underline{X} \to \mathbb{R}$, while Markov processes are denoted $\bar{p}$, $\bar{q}$, etc., i.e. $\bar{p} : \underline{X} \times X \to \mathbb{R}$. Let $L : X \to \mathbb{R}^+$ be a measurable function on X. We use the following notation.

For any $c \in \mathbb{R}$,

$$S_c = \{x \in X \mid L(x) \leq c\}. \qquad (1.1)$$

We assume S_c is compact for all c. Also, if $\bar{p} : \underline{X} \times X \to \mathbb{R}$ is a Markov process on X, we use the following notation.

For any $y \in X$, $\bar{p}_y$ is the measure $\bar{p}(\cdot|y) : \underline{X} \to \mathbb{R}$. For any $A \in \underline{X}$, $\bar{p}_A$ is the function $\bar{p}(A|\cdot) : X \to \mathbb{R}$

Now, given any two probability measures, p and q on X, we say <u>q dominates p with respect to L</u>, written $q >> p$, iff, for all $c \in \mathbb{R}$,

$$p(S_c) \geq q(S_c) \qquad (1.2)$$

Second, given two Markov processes on X, $\bar{p}$ and $\bar{q}$, we say that <u>$\bar{q}$ dominates $\bar{p}$ with respect to L</u>, written $\bar{q} >> \bar{p}$, iff, for all $y \in X$,

$$\bar{q}_y >> \bar{p}_y \qquad (1.3)$$

Thirdly, we say that the Markov process $\bar{q}$ is <u>stochastically increasing with respect to L</u>, iff, for all $x,y \in X$,

$$L(y) \geq L(x) \Rightarrow \bar{q}_y >> \bar{q}_x \qquad (1.4)$$

<u>Lemma 1.</u> Let p,q be probability measures on X. If $q >> p$ with

respect to L, then

$$\int L dp \leq \int L dq$$

<u>Proof:</u> For any integer, n, define $L_n : X \to \mathbb{R}^+$ by

$$L_n = \sum_{i=1}^{2^{n^2}} \frac{1}{2^n} \chi_{E_i} = \sum_{i=1}^{N} c\, \chi_{E_i}$$

where $N = 2^{n^2}$, $c = \frac{1}{2^n}$, and $E_i = X - S_{\frac{i}{2^n}} = \{x \in X \mid L(x) > \frac{i}{2^n}\}$. Then $\{L_n\}_{n=1}^{\infty}$ is a sequence of simple functions converging monotonically to L. Further, since $q >> p$, $p(E_i) \leq q(E_i)$ for all i, hence

$$\int L_n dp = \int \sum_{i=1}^{N} c\, \chi_{E_i} dp = \sum_{i=0}^{N} cp(E_i)$$

$$\leq \sum_{i=0}^{N} cq(E_i) = \int L_n dq.$$

Taking limits, as $n \to \infty$, the result follows.

Q.E.D.

<u>Lemma 2.</u> Let $\bar{p}$ and $\bar{q}$ be two Markov processes on X, and let p and q be two probability measures on X such that

(a) $\bar{q}$ is stochastically increasing

(b) $\bar{q} >> \bar{p}$

(c) $q >> p$

Then, defining $p^1 : \underline{X} \to \mathbb{R}$ and $q^1 : \underline{X} \to \mathbb{R}$ by, $\forall\, A \in \underline{X}$

$$p^1(A) = \int \bar{p}_A dp$$
$$q^1(A) = \int \bar{q}_A dp$$

it follows that $q^1 >> p^1$.

<u>Proof:</u> We must show, for all $c \in \mathbb{R}$, that $q^1(S_c) \leq p^1(S_c)$ (where $S_c = \{x \in X \mid L(x) \leq c\}$). But, since $\bar{q} >> \bar{p}$, we have $\bar{p}_{S_c}(y) \geq \bar{q}_{S_c}(y)$ for all y. Hence

$$p^1(S_c) = \int \bar{p}_{S_c} dp \geq \int \bar{q}_{S_c} dp$$

Further, by assumption (a), $\bar{q}_{S_c}$ is monotone decreasing with L. So $p >> q$ with respect to $\bar{q}_{S_c}$. Hence, applying Lemma 1,

$$\int \bar{q}_{S_c} dp \geq \int \bar{q}_{S_c} dq = q^1(S_c).$$

Q.E.D.

<u>Theorem 1.</u> Let $\bar{p}$ and $\bar{q}$ be two Markov processes on X such that

(a) $\bar{q}$ is stochastically increasing

(b) $\bar{q} >> \bar{p}$

(c) For all $A \in \underline{X}$, $\bar{p}_A$ is a continuous function on X.

Then if $\bar{q}$ has a stationary distribution, q^*, then $\bar{p}$ has a stationary distribution, p^*, and $q^* >> p^*$.

Proof: A set Π, of probability measures is defined to be tight if, for every $\varepsilon > 0$, there exists a compact set, F_ε, such that $p(F_\varepsilon) \geq 1 - \varepsilon$ for all $p \in \Pi$. Here we define

$$\Pi = \{p : \underline{X} \to \mathbb{R} \mid p \text{ is a probability measure and } q^* >> p\}$$

It is easily verified that Π is tight since $\{q^*\}$ is tight.

We now define $T : \Pi \to \Pi$, for all $A \in \underline{M}$ and $p \in \Pi$, by

$$Tp(A) = \int p_A dp$$

It follows, from Lemma 2, that T maps Π into Π.

Let C(X) denote the set of bounded, continuous real valued functions on X and let Ω denote the set of regular, countably additive set functions on the measurable subsets $\underline{X}$ of X. For each $f \in C(X)$ define the linear functional $\tilde{f} : \Omega \to \mathbb{R}$ by, for all $\lambda \in \Omega$,

$$\tilde{f}(\lambda) = \int f d\lambda,$$

Endow Ω with the weak topology generated by $\{\tilde{f} \mid f \in C(X)\}$ and let Π inherit the relative weak topology from Ω. Then under this topology,

(i) Π is a convex subset of the locally convex topological linear space Ω.

(ii) Π is (weakly) closed in Ω since any $\lambda \notin \Pi$ can be ''separated'' from Π by an appropriately chosen $\tilde{f}$ (using the fact that $p \in \Pi \to q^* >> p$).

(iii) Π is (weakly) compact by Lemma 2.8 of Futia [1980, p. 27] since the closure of any tight set of probability measures is compact in the weak topology we are using.

The Tychonoff fixed point theorem (see Smart [1980, p. 15]) can be invoked if we can show that $T : \Pi \to \Pi$ is (weakly) continuous. A fixed point $p^* \in \Pi$ for T will then satisfy the requirements of the theorem. We now show that T is indeed continuous.

Let V be the subbasic open neighborhood of $q_0 = Tp_0 \in \Pi$ generated by $\varepsilon > 0$ and $f \in C(X)$. Thus

$$V = \{q \in \Pi \mid |\tilde{f}(q - q_0)| < \varepsilon\} = \{q \in \Pi \mid |\int f d(q - q_0)| < \varepsilon\}.$$

For this $f \in C(X)$ define $h : X \to \mathbb{R}$ by

$$h(y) = \int f dp_y$$

It follows that $h \in C(X)$. To see this consider two cases:

(a) If f is a simple function, say $f = \sum_{i=1}^{n} c_i \chi_{E_i}$, then

$$h(y) = \int \left[\sum_{i=1}^{n} c_i \chi_{E_i} \right] dp_y = \sum_{i=1}^{n} c_i p_{E_i}(y)$$

So h is continuous since each p_{E_i} is continuous.

(b) For the general case, since f is continuous and bounded, we can construct a sequence of simple functions $\{f_n\}$ for which

$$\sup |f_n - f| \to 0 \text{ as } n \to \infty$$

Then, define h_n by

$$h_n(y) = \int f_n dp_y$$

It follows that for any $y \in X$,

$$|h_n(y) - h(y)| = \left|\int (f_n - f) dp_y\right| \leq \int |f_n - f| dp_y$$

$$\leq \sup |f_n - f|$$

So $\{h_n\}$ is a sequence of continuous functions converging uniformly to h. Hence h is continuous.

Further, $\forall\, p \in \Pi$, we prove that

$$\tilde{h}(p) = \tilde{f}(T_p). \qquad (*)$$

Again, it suffices to show that (*) is true for simple functions.

Thus, let $f = \sum_{i=1}^{n} c_i \chi_{E_i}$. Then

$$\tilde{h}(p) = \int h dp = \int \left[\int f dp_y \right] dp$$

$$= \int \left[\int \sum c_i \chi_{E_i} dp_y \right] dp$$

$$= \sum c_i \left[\int \left[\int \chi_{E_i} dp_y \right] dp \right]$$

$$= \sum c_i \left[\int p_{E_i} dp \right] = \sum c_i Tp(E_i)$$

$$= \int \sum c_i \chi_{E_i} dTp = \int f dTp = \tilde{f}(Tp).$$

Let $U = \{p \in \Pi \mid |\tilde{h}(p - p_0)| < \varepsilon\}$, a subbasic open neighborhood of p. Then

$$p \in U \to |\tilde{h}(p - p_0)| < \varepsilon \to |\tilde{h}(p) - \tilde{h}(p_0)| < \varepsilon$$

$$\to |\tilde{f}(Tp) - \tilde{f}(Tp_0)| < \varepsilon \to |\tilde{f}(Tp - q_0)| < \varepsilon$$

$$\to Tp \in V.$$

Thus $T : \Pi \to \Pi$ is continuous and the proof is complete.

Q.E.D.

3. Comparison of Processes in Different Spaces.

The above theorem can be extended to the case where the Markov processes $\bar{p}$ and $\bar{q}$ reside in different spaces. Of particular interest is the case when $\bar{p}$ is a Markov process over an arbitrary measurable topological space Y, while $\bar{q}$ is a Markov process over some measurable subspace X of $\mathbb{R}$. Let $\underline{Y}$ and $\underline{X}$ denote the Borel sets on Y and X respectively. We assume there is a $\underline{Y}$ measurable function $K : Y \to X$ such that for all $c \in \mathbb{R}$, $\{y \in Y \mid K(y) \leq c\}$ is compact. Let $\underline{M}(X)$ and $\underline{M}(Y)$ denote the sets of probability measures on X and Y, respectively. Also set $I : X \to X$ be the identity function. Now, given any $p \in \underline{M}(Y)$, p induces a measure $p \circ K^{-1} \in \underline{M}(X)$ in the natural way: i.e. $(p \circ K^{-1})(A) = p(K^{-1}(A))$. We can now define stochastic dominance.

For $p \in \underline{M}(Y)$, and $q \in \underline{M}(X)$, we say that q stochastically dominates p with respect to K, written $q \gg p$ iff

$$q \gg p \circ K^{-1} \text{ with respect to } I.$$

(Note that both q and $p \circ K^{-1}$ reside in $\underline{M}(X)$, so this is the same use of stochastic dominance as before.) Next, given two Markov processes, $\bar{p}$ and $\bar{q}$ over Y and X respectively, we say that $\bar{q}$ stochastically dominates $\bar{p}$ with respect to K, written $\bar{q} \gg \bar{p}$ iff, for all $y \in Y$,

$$\bar{q}_{K(y)} \gg \bar{p}_y$$

The definition of stochastic increasing Markov processes remains unchanged since all the action takes place within a single space.

However, since $X \subseteq \mathbb{R}$, we can use the identity function, I, for the mapping L. Then we say, for a Markov process $\bar{q}$ over X, that $\bar{q}$ is stochastically increasing if it is stochastically increasing with respect to I.

We can now prove a generalized version of Lemma 2.

Lemma 3. Let $\bar{p}$ and $\bar{q}$ be Markov processes on Y and X respectively where $X \subseteq \mathbb{R}$, and let $p \in \underline{M}(Y)$ and $q \in \underline{M}(X)$. Further, assume

(a) $\bar{q}$ is stochastically increasing

(b) $\bar{q} \gg \bar{p}$

(c) $q \gg p$

Then if $p^1 : \underline{Y} \to \mathbb{R}$, and $q^1 : \underline{X} \to \mathbb{R}$ are defined by, $\forall A \in \underline{Y}$, $B \in \underline{X}$

$$p^1(A) = \int \bar{p}_A dp$$

$$q^1(B) = \int \bar{q}_B dp$$

it follows that $q^1 \gg p^1$.

Proof: Setting $L = I$, we have for any $c \in \mathbb{R}$, the sets S_c are of the the form

$$S_c = \{x \in X \mid x \leq c\}.$$

We must show, for all c, that

$$p^1 \circ K^{-1}(S_c) \geq q^1(S_c)$$

or, setting $R_c = \{y \in Y \mid K(y) \leq c\}$, we must show

$$p^1(R_c) \geq q^1(S_c)$$

for all c. But now $\bar{q} >> \bar{p} \Rightarrow \bar{q}_{K(y)} >> \bar{p}_y$ ($\forall$ y ε Y).

$\Rightarrow \bar{q}_{K(y)} >> \bar{p}_y \circ K^{-1}$ ($\forall$ y ε Y).

$\Rightarrow \bar{q}_{K(y)}(S_c) \leq \bar{p}_y \circ K^{-1}(S_c) = \bar{p}_y(R_c)$ ($\forall$ y ε Y).

Or, written differently, for all c, and all y,

$$\bar{p}_{R_c}(y) \geq \bar{q}_{S_c}(K(y)) = \bar{q}_{S_c} \circ K(y)$$

Hence

$$p^1(R_c) = \int \bar{p}_{R_c} dp \geq \int \bar{q}_{S_c} \circ K\, dp = \int \bar{q}_{S_c} d(p \circ K^{-1})$$

$$\geq \int \bar{q}_{S_c} dq = q^1(S_c)$$

The next to last step follows from Lemma 1 using the fact that $q >> p \circ K^{-1}$ with respect to I, and the fact (from assumption (a)), that $\bar{q}_{S_c}$ is monotonic decreasing with I. Hence $p \circ K^{-1} << q$ with respect to $\bar{q}_{S_c}$.

Q.E.D.

This gives us the following corollary to Theorem 1.

Corollary 1. Let $\bar{p}$ be a Markov process on Y, and $\bar{q}$ be a Markov process on $X \subseteq \mathbb{R}$ such that

(a) $\bar{q}$ is stochastically increasing

(b) $\bar{q} >> \bar{p}$

(c) For all A ε $\underline{Y}$, $\bar{p}_A$ is a continuous function on X.

Then if $\bar{q}$ has a stationary distribution, q*, then $\bar{p}$ has a stationary distribution, p*, and $q^* >> p^*$.

Proof: The argument is exactly the same as that of Theorem 1. Lemma 3 establishes that the mapping T maps Π into Π.

4. An Example and an Application.

Our initial intuition led us to believe that Theorem 1 would be true without condition (a). We close with an example illustrating that the result is not true without this condition.

Let $X = \underline{N}$, the natural numbers. Write $\bar{p}(i,j)$ for $\bar{p}(\{i\},j)$, and similarly for $\bar{q}(i,j)$. Then define $\bar{p}$ and $\bar{q}$ as follows. For j odd, set

$$\bar{r}(i,j) = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{otherwise} \end{cases}$$

and for j even, set

$$\bar{r}(i,j) = \begin{cases} \alpha_i & \text{if } i = j + 2 \\ (1-\alpha_i) & \text{if } i = j + 3 \\ 0 & \text{otherwise} \end{cases}$$

Then let $\bar{q}$ be defined to be the process $\bar{r}$ generated when $\alpha_i = 1/2$ for all i, and let $\bar{p}$ be any process generated when $\alpha_i > 1/2$ for all i and when $\prod_{i=0}^{\infty} \alpha_i > 0$. Clearly $\bar{q} \gg \bar{p}$ with respect to the identity function, and $\bar{q}$ has a stationary distribution q [specifically, writing q(i) for q({i}) and so forth, $q(2i) = 2^{-i}\ 3^{-1}$, $q(1) = 0$, and $q(2i+3) = 2^{-(i+1)}\ 3^{-1}$], but it will be shown that $\bar{p}$ has no stationary distribution.

Suppose that $\bar{p}$ has a stationary distribution p, i.e., that $Tp = p$ where $T : \Pi \to \Pi$ is defined by $Tr(A) = \sum_{i=0}^{\infty} \bar{p}(A|i)r(i)$. Denote the even numbers by $2\underline{N}$ and the odd numbers by $2\underline{N}+1$. It is evident that if $r(2\underline{N}+1) = 1$, then $Tr(2\underline{N}+1) = 0$. Thus $p(2\underline{N}) > 0$, so $p(2j) > 0$ for some j. For some $k > j$, $p(2k) < \prod_{i=0}^{\infty} \alpha_i p(2j)$. However,

$$p(2k) = T^{2(k-j)}p(2k) = \prod_{i=2j}^{2k-1} \alpha_i p(2j) \geq \prod_{i=0}^{\infty} \alpha_i p(2j),$$ a contradiction.

Since $\bar{p}$ and $\bar{q}$ satisfy hypotheses (b) and (c) of Theorem 1 (continuity of $\bar{p}_A$ is trivial because $\underline{N}$ is a discrete topological space), the indispensability of assumption (a) for the theorem is established.

The motivation for the ideas developed in this paper emerged from investigation of a stochastic model of committee voting behavior (Ferejohn, McKelvey, and Packel 1981). We conclude with a brief description of how Theorem 1 applies in this setting.

The underlying space X is Euclidean and represents the set of alternatives, one of which must be selected by a finite set of voters operating under majority rule. Each voter has an ''ideal point'' in X, with voter preference for any given alternative decreasing monotonically with its Euclidean distance from the voter's ideal point. A stepwise selection procedure is modelled as a discrete time Markov process $\bar{p}$ over X, whose stationary distribution p* provides the desired solution concept. When X is a compact subset of $\mathbb{R}^m$ convergence follows readily, but the noncompact case is theoretically important and leads to a number of difficulties. By studying a more tractable process $\bar{q}$, a Markov chain over the positive integers, which satisfies the conditions of Corollary 1, we are able to establish the existence of p* in a fairly general setting. Corollary 1 also ensures that the stationary distribution q* for $\bar{q}$ satisfies $q^* \gg p^*$, providing useful bounds on the concentration of p* in terms of the voter's ideal points.

It seems reasonable to expect the Theorem and Corollary of this paper to have application in a variety of practical and theoretical situations of this sort. In this role, the result serves as a ''dominated convergence theorem'' for Markov processes.

REFERENCES

Ferejohn, J., R. McKelvey, and E. Packel. ''Limiting Distributions for Continuous State Markov Voting Models.'' Social Science Working Paper #394.

Futia, Carl A. ''A Stochastic Approach to Economic Dynamics.'' Unpublished manuscript, Bell Telephone Laboratories, 1980.

Smart, D. R. Fixed Point Theorems. Cambridge: Cambridge University Press, 1980.