Common Priors and Markov Chains

by

Dov Samet

September 1996

ABSTRACT

The type function of an agent, in a type space, associates with each state a probability distribution on the type space. Thus, a type function can be considered as a Markov chain on the state space. A common prior for the space turns out to be a probability distribution which is invariant under the type functions of all agents. Using the Markovian structure of type spaces we show that a necessary and sufficient condition for the existence of a common prior is that for each random variable it is common knowledge that all its joint averagings converge to the same value.

§1. Introduction

Ever since Harsanyi's (1967-1968) work on games with incomplete information, type spaces have been the most important tool for the study of such games. In the vast majority of the applications of type spaces to economics, it is assumed that players' beliefs are derived from a common prior. It is therefore of interest to characterize, in terms of players' beliefs, the existence of a common prior. Aumann (1975), in his agreement theorem, gave a necessary condition for the existence of a common prior: If there is a common prior, then it is impossible to agree to disagree, i.e., to have common knowledge of differences in the posterior beliefs of any given event. By extending the notion of disagreement to the expectation of a general random variable, several authors (Morris (1995), Feinberg (1995, 1996), and Bonanno and Nehring (1996)) were able to formulate a necessary and sufficient condition for the existence of a common prior in terms of agreement to disagree. This condition is proved using various derivatives of the separation theorem for convex sets. Without stating it explicitly, all these works depend on the simple observation that the set of priors of an agent is the convex hull of his types. Samet (1996) shows how the condition follows directly from this observation.

In this work we give a new characterization of the existence of a common prior, based on the following simple observation, which is related to the stochastic nature of types and priors, rather than their convex structure.

> *The type function of an agent in a type space, namely the function that associates with each state the agent's probability distribution over the type space, at the state, is by definition a Markov chain, and the set of priors of the agent is the set of invariant probability distributions of this Markov chain.*

Thus, a common prior is a probability distribution which is invariant under the Markov chains of all agents. It is the interaction of several simple Markov chains, one for each agent, which makes the stochastic structure of type spaces non-trivial.

Using the stochastic structure of type spaces we are able to show the following results, concerning joint averaging, which we define next. Let $i_1, i_2, \ldots$ be an infinite sequence of names of agents, such that each agent's name appears infinitely many times. Let f be a random variable (i.e., a function on the state space). Agent i_1 evaluates, in each state, the expected value of f. Denote this by $E_{i_1} f$. This evaluation itself is a random variable, and $E_{i_2} E_{i_1} f$ is its evaluated expected value for i_2, and so on. We call the sequence of the random variables thus obtained a **joint averaging** of the random variable f. Different I-sequences give rise to different joint averagings, but we prove:

> *Each joint averaging of a random variable converges, and its value is always common knowledge. Moreover, there exists a common prior, if and only if, for each random variable it is always common knowledge that all its joint averagings converge to the same value.*

We present type spaces, and state the main results in the next section. The interpretation of type functions as Markov chains, and all the proofs are given in Section 3.

§2. Type spaces, priors and common priors

Let I and Ω be finite sets of sizes n and m respectively. The elements of I are **agents**, and those of Ω are **states**. Subsets of Ω are **events**. For each $i \in I$, Π_i is a partition of Ω. For $\omega \in \Omega$ we denote by $\Pi_i(\omega)$ the element of Π_i containing ω. For each $i \in I$ let $t_i(\omega)$ be a probability distribution on Ω, such that:

(a) $t_i(\omega)\big(\Pi_i(\omega)\big) = 1$;

(b) for each $\omega' \in \Pi_i(\omega)$, $t_i(\omega') = t_i(\omega)$. The function t_i is i's type function and $t_i(\omega)$ is i's type at ω.

The tuple $< I, \Omega, (\Pi_i, t_i)_{i\in I} >$ is called a **type space**.

We assume, for simplicity, and without relinquishing any important feature of type spaces, that for each i and ω, $t_i(\omega)\big(\{\omega\}\big) > 0$.

The **meet** of $(\Pi_i)_{i\in I}$ is the partition Π of Ω which is the finest among all partitions that are coarser than Π_i for each $i \in I$. For an event A, the event that A is **common knowledge** is the union of all the elements of Π contained in A. We observe that any element P in the meet Π forms a type space when the partitions Π_i and the types t_i are restricted to it, and the meet of this space is $\{P\}$. We identify any probability distribution on P with its natural extension to a probability distribution over all of Ω.

We consider probability distributions on Ω as row vectors in the m-dimensional space R^Ω. A **random variable** is a real valued function on Ω, which we consider as a column vector in R^Ω. For a probability distribution p and a random variable f on Ω, the expectation of f with respect to p is the scalar product $pf = \sum_w p(\omega)f(\omega)$. For each agent i, and random variable f on Ω, i's **expectation** of f, denoted $E_i f$, is the random variable, $(E_i f)(\omega) = t_i(\omega)f$.

A **prior** for agent i, is a probability distribution p on Ω, such that for each state ω, if $p\big(\Pi_i(\omega)\big) > 0$, then i's type at ω is the conditional probability distribution defined by p on $\Pi_i(\omega)$. That is, p is a prior for i, if for each event A and state ω, $p\big(A \mid \Pi_i(\omega)\big) = t_i(\omega)(A)$, whenever the conditional probability distribution is defined. The probability distribution p is a **common prior** if it is a prior for each agent i.

We present here necessary and sufficient conditions for the existence of a common prior which are the result of the stochastic nature of type spaces explored in the next section.

First, we show in the following proposition that the question of existence of a common prior on Ω can be reduced to the question of the existence of common prior on the elements of the meet.

Proposition 1. *For each $P \in \Pi$ there exists at most one common prior on P. The set of common priors on Ω is the convex hull of the common priors on the elements P in Π.*

We need the following definitions. We call a sequence $s = (i_1, i_2, \ldots)$, of elements of I, an **I-sequence** if for each agent i, $i = i_k$ for infinitely many k's. The **joint averaging** of a random variable f with respect to the I-sequence s is the sequence of random variables $(E_{i_k} \cdots E_{i_1} f)_{k=1}^{\infty}$.

Proposition 2. *For each random variable f on Ω and I-sequence s, the limit of the joint averaging of f with respect to s exists, and its value is common knowledge in each state. That is, it is constant on each element P in Π.*

In view of Proposition 1, there exists a common prior on Ω, iff there exists a common prior on at least one of the elements of the meet. Thus it is enough to characterize the existence of a common prior for the case that the meet consists of only one element.

Theorem 1. *Suppose $\Pi = \{\Omega\}$. Then there exists a common prior, iff for each random variable f it is common knowledge in each state that the joint averaging of f, with respect to all I-sequences s, converge to the same limit. Moreover, if p is the common prior, then this limit is pf.*

Two remarks are in order. First, note that each random variable is a linear combination of the random variables $\chi_{\{\omega\}}$ — the characteristic functions of single states. As the expectation operators E_i are also linear, Theorem 1 holds true if the the random variables f, in this theorem, are restricted to the functions $\chi_{\{\omega\}}$, or more generally to characteristic functions of events, χ_E.

Second, we observe that Theorem 1 generalizes Aumann's agreement theorem as follows. Suppose the event $(E_1\chi_E = \alpha) \cap (E_2\chi_E = \beta)$ is common knowledge i.e., $E_1\chi_E$ and $E_2\chi_E$ have fixed values α and β, correspondingly, on Ω. Then, obviously, all joint averagings of χ_E for I-sequences that start with 1 are constantly α, while for those sequences that start with 2, they are constantly β. If there is common prior, then, by Theorem 1, $\alpha = \beta$.

§3. Type spaces as Markov chains

For a given type space $< I, \Omega, (\Pi_i, t_i)_{i\in I} >$ define for each agent i a stochastic matrix M_i in R^{Ω^2}, by $M_i(\omega,\omega') = t_i(\omega)\big(\{\omega'\}\big)$. Then, M_i is a matrix of a Markov chain on Ω. The following proposition follows directly from the definitions of type spaces and of Markov chains.

Proposition 3.

a) *Each element of the partition Π_i is an irreducible class of M_i.*
b) *A probability distribution p on Ω is a prior for i, iff it is a stationary probability distribution for M_i, that is, $pM_i = p$.*
c) *For each random variable f, $M_i f = E_i f$.*

For any permutation σ of I, we write,

$$M_\sigma = M_{\sigma(1)} \cdots M_{\sigma(1)}.$$

Proposition 4. *For any permutation σ of I, the meet, Π, is a partition of Ω into irreducible, aperiodic, classes of M_σ. Thus, the restriction of M_σ to any $P \in \Pi$ is ergodic and therefore has a unique invariant probability distribution p^P_σ on P.*

Proof: Let If $\omega, \omega' \in P \in \Pi_{\sigma(i)}$. Then

$$M_\sigma(\omega,\omega') \geq M_{\sigma(1)}(\omega,\omega)\cdots M_{\sigma(i-1)}(\omega,\omega)M_{\sigma(i)}(\omega,\omega')M_{\sigma(i+1)}(\omega',\omega')M_{\sigma(n)}(\omega',\omega') > 0.$$

Therefore, any two states in the same element of a partition communicate. Hence, if ω is in an equivalence class of states, then $\Pi_i(\omega)$, for each i, is a subset of this class. This means that each class is a union of elements of Π. Also, for each $P \in \Pi$, the probability of $\omega \in P$ to stay in P under M_σ is 1, and therefore P is an equivalence class. The Markov chain M_σ is aperiodic since for each σ, $M_\sigma(\omega, \omega) > 0$. ∎

Proposition 5. *The following conditions are equivalent.*

(1) p is a common prior on Ω.

(2) p is an invariant probability distribution of the Markov chain M_i for each $i \in I$.

(3) p is an invariant probability distribution of the Markov chain M_σ for each permutation σ.

Proof: Clearly (1) and (2) are equivalent by Proposition 3 (b), and (2) implies (3). Suppose (3) is true, and let p be the invariant probability distribution in (3). Thus,

$$pM_1 M_2 \cdots M_n = p.$$

Multiplying this equality by M_1, from the right, yields

$$pM_1 M_2 \cdots M_n M_1 = pM_1.$$

Therefore, pM_1 is an invariant probability distribution of $M_2 \cdots M_n M_1$. But by (3), p is an invariant probability distribution of this Markov chain, and by Proposition 4, $M_2 \cdots M_n M_1$ has a unique invariant probability distribution. Thus, $pM_1 = p$, and similarly, $pM_i = p$ for each $i \in I$. ∎

Proof of Proposition 1: By Proposition 5, if p is a common prior on P, then it is an invariant probability distribution of the restriction of M_σ to P, which is unique by Proposition 4. Let p be a common prior, and denote by p^P the conditional probability distribution of p to $P \in \Pi$, when it exists. Then, clearly, p is a convex combination of the measures p^P. It is enough, now, to show that each p^P is an invariant probability distribution on P. Indeed, by Proposition 3 (b), p is an invariant probability distribution of M_i for each i, and therefore by (a), p^P is an invariant probability distribution of the restriction of M_i to P, for each i. Hence, by Proposition 5, p^P is a common prior on P. ∎

To prove Proposition 2, and Theorem 1, we first prove a variant of these claims. Let σ be a permutation of I. Denote by E_σ the operator, which is defined for each f by

$$E_\sigma f = E_{\sigma(1)} \cdots E_{\sigma(n)} f.$$

The **joint averaging** of a random variable f with respect to σ is the sequence $(E_\sigma^k f)_{k=1}^\infty$.

Proposition 2'. *For each random variable f on Ω and permutation σ, the limit of the joint averaging of f with respect to σ exists, and it is measurable with respect to Π, i.e., it is constant on each element P in Π.*

Note that the joint averaging of f with respect to a permutation σ is a subsequence of the joint averaging of f with respect to the I-sequence, $\sigma(1), \ldots, \sigma(n), \sigma(1), \ldots, \sigma(n), \ldots$, and therefore the claim of Proposition 2' is weaker that that of Proposition 2.

Theorem 1'. *Suppose* $\Pi = \{\Omega\}$. *Then there exists a common prior, iff for each random variable* f*, the joint averaging of* f*, with respect to all permutations* σ*, converge to the same limit. Moreover, if* p *is the common prior, then this limit is* pf*.*

The claim that convergence of the joint averaging of f, with respect to all I-sequences s, to the same limit, implies the existence of a common prior, is weaker than that of Theorem 1. But the claim of converse direction is stronger here.

Proof of Proposition 2': By Proposition 3 (c), $E_\sigma^k f = M_\sigma^k f$, for each f and k. As, by Proposition 4, M_σ is ergodic on P, $\lim_{k\to\infty} M_\sigma^k f = p_\sigma^P f$, over P.

Proof of Theorem 1': As in the proof of Proposition 2',

$$\lim_{k\to\infty} E_\sigma^k f = \lim_{k\to\infty} M_\sigma^k f = p_\sigma f,$$

where p_σ is the unique invariant probability distribution of M_σ on Ω. Thus, for each f, the limits for all σ are the same, iff for each f, $p_\sigma f$ are the same for all σ, which is true, iff there is a probability distribution p such $p_\sigma = p$ for all σ. This amounts, by the equivalence of (1) and (3) in Proposition 5, to saying that p is a common prior. ∎

We turn now to prove Proposition 2 and Theorem 1. We first prove Proposition 6, which generalizes a theorem concerning the convergence of the powers of an ergodic stochastic matrix to the case in which different matrices are multiplied. We then prove Lemma 1, which shows that the conditions of Proposition 6 hold in our case.

We say that a stochastic matrix A is **bounded by** ε if all its positive entries are bounded from below by ε, that is, if for each row r and column c, either $A(r,c) = 0$ or $A(r,c) > \varepsilon$. We say that A is **positive** if all its entries are positive.

Proposition 6. *Let* $A_1, A_2, \ldots, A_k, \ldots$ *be a sequence of stochastic matrices of the same dimension. Denote* $A^{(k)} = A_k A_{k-1} \cdots A_1$. *Let* $1 = k_1 < k_2 < \cdots < k_l < \cdots$ *be an increasing sequence of indices and denote by* B_l*, for* $l = 1, \ldots, \infty$*, the block* $B_l = A_{k_{l+1}-1} \cdots A_{k_l}$. *If there exists* $\varepsilon > 0$ *such that* B_l *is positive and bounded by* ε*, for each* l*, then there exists a matrix* A*, all the rows of which are identical, such that* $\lim_{k\to\infty} A^{(k)} \to A$. *Moreover, if there exists a probability distribution* p *which is invariant for* A_k *for all* k*, then all the rows of* A *are* p*.*

Proof: It is enough to show that for each column vector x, $A^{(k)}x$ converges to a vector all the component of which are identical. Indeed, if we prove this, then substituting unit vectors for x shows the existence of the limit matrix A with the desired property.

For a vector x write $\max x$ for the maximal coordinate x_i and $\min x$ for the minimal one. If A is a stochastic matrix, and $y = Ax$, then $\max y \le \max x$ and $\min y \ge \min x$. Moreover, if A is positive and bounded by $\varepsilon > 0$, then $\max y \le \varepsilon \min x + (1-\varepsilon) \max x$, and $\min y \ge \varepsilon \max x + (1-\varepsilon) \min x$, and therefore, $\max y - \min y \le (1-2\varepsilon)(\max x - \min x)$.

Thus, if $y^{(k)} = A^{(k)}x$, then $\max y^{(k)}$ is a decreasing sequence and $\min y^{(k)}$ is an increasing one. We need to show that $\max y^{(k)} - \min^{(k)} \to 0$. This is indeed true, as for each l, $y^{(k_{l+1}-1)} = B_l \cdots B_1 x$, and therefore, $\max y^{(k_{l+1}-1)} - \min y^{(k_{l+1}-1)} \le (1-2\varepsilon)^l (\max x - \min x)$.

If p is an invariant probability distribution of each A_i, then $pA^{(k)} = p$ for each k and therefore $pA = p$. But as all the rows of A are identical, pA is a row of A. ∎

Note that if all the matrices A_k are the same matrix M, then the condition concerning the uniform boundness of the positive blocks B_l, is equivalent to the ergodicity of M.

If we define $\varepsilon_0 = \min_{i,\omega} M_i(\omega,\omega)$, then $\varepsilon_0 > 0$, and all the matrices M_i are bounded by ε_0.

Lemma 1. *Let $B = M_{i_l} \cdots M_{i_1}$ and $C = M_i B$, and suppose B is bounded by ε. Then,*
(a) all the positive entries of B are also positive in C;
(b) if C has no positive entries other than those of B, then C is also bounded by ε;
(c) in any case, C is bounded by $\varepsilon_0\varepsilon$.

Proof: To prove (a), observe that if $B(\omega,\omega') > 0$, then $C(\omega,\omega') \geq M_i(\omega,\omega)B(\omega,\omega') > 0$. To prove (b), suppose the assumption of (b) holds and $C(\omega,\omega') > 0$. Then $B(\omega,\omega') > 0$. But for any $\bar{\omega} \in \Pi_i(\omega)$, the row $M_i(\omega,\cdot)$ is the same as the row $M_i(\bar{\omega},\cdot)$. Therefore, for each such $\bar{\omega}$, $C(\bar{\omega},\omega') = C(\omega,\omega') > 0$ and hence, by the assumption in (b), $B(\bar{\omega},\omega') > 0$. Now, by the definitions of C and M_i,

$$C(\omega,\omega') = \sum_{\bar{\omega}:\bar{\omega}\in\Pi_i(\omega)} M_i(\omega,\bar{\omega})B(\bar{\omega},\omega').$$

But as $B(\bar{\omega},\omega')$ is positive, it is bounded from below by ε, and hence

$$C(\omega,\omega') \geq \sum_{\bar{\omega}:\bar{\omega}\in\Pi_i(\omega)} M_i(\omega,\bar{\omega})\varepsilon = \varepsilon$$

.

Obviously, (c) is true as each positive entry of C is the sum of products of positive entries from M_i and B, each of which is bounded from below by $\varepsilon_0\varepsilon$. ∎

Proof of Proposition 2: Let $s = (i_1, i_2, \ldots)$ be an I-sequence. Fix $P \in \Pi$. For each k, define A_k to be the restriction of M_{i_k} to P. We show that the sequence $A^{(k)}$ (in the notation of Proposition 6) converges. For each permutation σ of I, M_σ is ergodic on P, and therefore there is a whole number ν_σ such that for all $\mu \geq \nu_\sigma$, the restriction of M_σ^μ to P is positive. Let $\nu = \max_\sigma \nu_\sigma$. As there are finitely many permutations of I, there must be a permutation σ, and blocks, $B_l = A_{k_{l+1}-1} \cdots A_{k_l}$, such that for each block B_l, there are indices $j_1, \ldots, j_{n\nu}$ which satisfy $k_l \leq j_1 < j_2 < \ldots < j_{n\nu} \leq k_{l+1} - 1$ and the restriction of M_σ^ν to P is $A_{j_{nq}} \cdots A_{j_1}$. It is clear that all the entries that are positive in A_{j_1} are positive also in $A_{j_1} \cdots A_{k_l}$. The block B_l is obtained from this matrix by multiplying it from the left repeatedly. Such multiplications can only enlarge the set of positive entries, by Proposition 3 (a). Thus, the set of positive entries in $A_{j_2} \cdots A_{j_1} \cdots A_{k_1}$ is not smaller than the set of entries that are positive in $A_{j_2}A_{j_1}$. Finally, remembering that M_σ^ν is positive on P, we conclude that B_l is positive. Moreover, the number of times that the set of positive entries can be enlarged is, of course, at most m^2. Thus, by Proposition 3, (b) and (c), all blocks B_l are uniformly bounded by $\varepsilon_0^{m^2}$. Thus, by Proposition 6, $(A^{(k)}$ converges to a matrix A, the rows of which are identical. Obviously, for each f and ω, $E_{i_k} \cdots E_{i_1} f = A^{(k)} f$, and the latter converges to the constant function pf, where p is a row in A. ∎

Proof of Theorem 1: That the convergence in the theorem implies the existence of a common prior follows from Theorem 1'. Indeed, joint averaging of f with respect to a permutation σ is a subsequence of the joint averaging of f with respect to the I-sequence, $\sigma(1), \ldots, \sigma(n), \sigma(1), \ldots, \sigma(n), \ldots$. If there is a common prior p then convergence holds by Proposition 2. ∎

References

Aumann, R. (1976), *Agreeing to Disagree,* The Annals of Statistics, 4, pp 1236-1239.

Bonanno, G. and K. Nehring, (1996), *Fundamental Agreement: A New Foundation for the Harsanyi Doctrine,* Working Paper No. 96-02, University of California, Davis.

Feinberg, Y. (1995), *Converse to the Agreement Theorem,* Discussion Paper No. 83, Center for Rationality and Interactive Decision Theory, The Hebrew University.

Harsanyi, J.C. (1967-68), *Games with Incomplete Information Played by Bayesian Players,* parts I,II, & III, Man. Sc., 14, pp. 159-182, 320-334, 486-502.

Morris, S. (1995), *Trade with Heterogenous Prior Beliefs and Asymmetric Information,* Econometrica, 62, pp. 1327-1347.

Samet, D. (1996), *Common Priors as the Intersection of Convex Sets,* a manuscript.