ISRAEL JOURNAL OF MATHEMATICS, Vol. 24, Nos. 3-4, 1976

THE $\bar{d}$-DISTANCE BETWEEN TWO MARKOV PROCESSES CANNOT ALWAYS BE ATTAINED BY A MARKOV JOINING

BY

MARTIN H. ELLIS

ABSTRACT

For a certain collection of pairs of Markov processes a construction is given which attains the $\bar{d}$-distance for these pairs; it is then shown this distance cannot be attained by any Markov process on their joint atoms.

Given two processes $\mathcal{T}_1$ and $\mathcal{T}_2$ whose partitions are indexed by the same set, the $\bar{d}$-distance between them, $\bar{d}(\mathcal{T}_1, \mathcal{T}_2)$, is the infimum of the partition distances which can be attained between their partitions when embedding both processes in a third process; in fact the infimum is attained by some process (see Ornstein [2] or [3] for further discussion of $\bar{d}$). Given two Markov processes whose partitions are indexed by the same set, the "Markov distance" between them, $M(\mathcal{T}_1, \mathcal{T}_2)$, is the infimum of the partition distances which can be attained between their partitions when embedding both processes in a third process which is a Markov process on the join of their embedded partitions; again, the infimum is attained by a Markov process on the joint partition. Clearly $M(\mathcal{T}_1, \mathcal{T}_2) \geqq \bar{d}(\mathcal{T}_1, \mathcal{T}_2)$ for $\mathcal{T}_1, \mathcal{T}_2$ Markov processes. It was conjectured that $M(\mathcal{T}_1, \mathcal{T}_2)$ always equalled $\bar{d}(\mathcal{T}_1, \mathcal{T}_2)$. In [1] this was shown to be false as follows: the function M was computed for all pairs of two-state Markov processes, and it was shown that M fails to satisfy the Triangle Inequality, hence is not a metric; since $\bar{d}$ is a metric, this shows that M does not always equal $\bar{d}$. The computation of M in [1] is involved.

In this paper we give a collection of pairs of two-state Markov processes and a construction which attains the $\bar{d}$-distance for them, then we calculate the "Markov distance" for them and observe that this distance is greater than the $\bar{d}$-distance, thus showing that $M \neq \bar{d}$. We conclude the paper by mentioning an

Received November 23, 1975

extension of the construction to a larger class of pairs, and by stating precisely when $M = \bar{d}$ for pairs of two-state Markov processes.

We will denote the symmetric two-state Markov process with transition matrix

$$\begin{pmatrix} 1-\kappa & \kappa \\ \kappa & 1-\kappa \end{pmatrix}$$

by (κ), where κ is the probability of leaving a state and going to the other state. We always assume $0 < \kappa < 1$. To avoid confusion when considering a pair $((\alpha),(\gamma))$, we denote the first state of (α) by A, the second state of (α) by B and the first state of (γ) by C, the second state of (γ) by D.

The following proposition gives the $\bar{d}$-distance between two symmetric two-state Markov processes which have the same square.

Proposition 1. *For* $0 < \alpha < 1/2$, $\bar{d}((\alpha),(1-\alpha)) = (1-2\alpha)/2$.

Proof. $\bar{d}((\alpha),(1-\alpha)) \geqq (1-2\alpha)/2$: $(1-\alpha)$ switches the states of $1-\alpha$ of its measure, whereas (α) switches the states of α of its measure, whence in any joint embedding at least $1-2\alpha$ of the measure switches states for the $(1-\alpha)$ process, but not for the (α) process, and since all points in this set are either going into or coming out of disagreement, the distance attained by the embedding must be $\geqq (1-2\alpha)/2$.

$\bar{d}((\alpha),(1-\alpha)) \leqq (1-2\alpha)/2$: It will suffice to show that we can change a generic sequence for (α) into a generic sequence for $(1-\alpha)$ by changing only $(1-2\alpha)/2$ of the letters of the sequence (see Ornstein [2] or [3] for equivalence of this condition to other definitions of $\bar{d}$). Note that the squares of (α) and $(1-\alpha)$ are both equal to $(2\alpha - 2\alpha^2)$. Hence if we take a generic sequence for (α) and consider only the letters in even places it will be a generic sequence for $(1-\alpha)$ considered only in even places. Thus to change a generic sequence for (α) into one for $(1-\alpha)$ we can limit the changes to letters in the odd places.

The following rule will change a generic sequence $\{y_i\}_{i\in Z}$ for (α) into a generic sequence for $(1-\alpha)$: For all $z \in \mathbb{Z}$, independently of whatever other changes are being made, whenever $y_{2z} = y_{2z+1} = y_{2z+2}$, change y_{2z+1} with probability $(1-2\alpha)/(1-2\alpha+\alpha^2)$; otherwise, leave y_{2z+1} unchanged.

The probability that $y_{2z} = y_{2z+1} = y_{2z+2}$ is $1-2\alpha+\alpha^2$, hence this rule changes $1-2\alpha$ of the odd places, so $(1-2\alpha)/2$ of the y_i are changed. To verify that the resultant sequence is indeed a generic sequence for $(1-\alpha)$, note that both letters appear in the resultant sequence with the correct probability (1/2) and check that the probability that a letter in the resultant sequence differs from the letter immediately to its left is $1-\alpha$ irrespective of what letters occur to its left.

Q.E.D.

Note that the rule given above is a description of a process in which both (α) and $(1-\alpha)$ are embedded which attains the partition distance of $(1-2\alpha)/2$ between their partitions: the eight-state Markov process with transition matrix

$$\begin{pmatrix} 0 & 0 & \alpha & 0 & 1-2\alpha & 0 & \alpha & 0 \\ 0 & 0 & 0 & \alpha & 0 & 1-2\alpha & 0 & \alpha \\ \alpha & 1-\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 1-\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & 1-\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 1-\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and distribution $(1/4, 1/4, \alpha/4, \alpha/4, (1-2\alpha)/4, (1-2\alpha)/4, \alpha/4, \alpha/4)$, with $A = 1 \cup 3 \cup 4 \cup 5$, $B = 2 \cup 6 \cup 7 \cup 8$, $C = 1 \cup 3 \cup 4 \cup 6$, $D = 2 \cup 5 \cup 7 \cup 8$.

States 1 and 2 are the outputs at even times, the other states are the outputs at odd times. One can verify that lumping the states into the two atoms A, B yields (α) and lumping the states into the two atoms C, D yields $(1-\alpha)$. However lumping the states into the four atoms $A \cap C$, $A \cap D$, $B \cap C$, $B \cap D$ does not yield a Markov process. This last fact could in fact be inferred from the following proposition, which is a particular case of Theorem 3.1 of [1].

PROPOSITION 2. *For* $0 < \alpha < 1/2$, $M((\alpha), (1-\alpha)) = (1-2\alpha)/(2-2\alpha)$.

PROOF. The Markov process with transition matrix

$$\begin{pmatrix} \alpha & 1-2\alpha & 0 & \alpha \\ 1-\alpha & 0 & 0 & \alpha \\ \alpha & 0 & 0 & 1-\alpha \\ \alpha & 0 & 1-2\alpha & \alpha \end{pmatrix}$$

and distribution $(1/(4-4\alpha), (1-2\alpha)/(4-4\alpha), (1-2\alpha)/(4-4\alpha), 1/(4-4\alpha))$ attains this distance with $A \cap C = 1$, $A \cap D = 2$, $B \cap C = 3$, $B \cap D = 4$.

Suppose a smaller distance than $(1-2\alpha)/(2-2\alpha)$ can be attained by a Markov process on the joint atoms $1 = A \cap C$, $2 = A \cap D$, $3 = B \cap C$, $4 = B \cap D$. Let

$$\begin{pmatrix} e_{11} & \cdot\cdot & e_{14} \\ \vdots & - & \vdots \\ e_{41} & \cdot\cdot & e_{44} \end{pmatrix}$$

be the transition matrix for this improvement, and let μ be its probability measure. Note that we can assume $e_{13}+e_{14}=\alpha$: If $e_{13}+e_{14}\neq\alpha$, then $e_{41}\mu(4)+e_{31}\mu(3)$ must equal $\alpha\mu(1)$, for if the measure sent from B to A entered A with a distribution differing from the distribution of A (between atoms 1 and 2) then the part of it immediately returning to B would not be equal to α of its measure (it would be more or less depending on whether the overconcentration was in thé atom of A sending more or less than α of its measure to B). Hence if $e_{13}+e_{14}\neq\alpha$, replace the improvement with its inverse to obtain a Markov process on 1, 2, 3, 4 which attains the improved distance and for which $e_{13}+e_{14}=\alpha$.

Since $\mu(2)=\mu(3)$ and $\mu(2)+\mu(3)<(1-2\alpha)/(2-2\alpha)$ we must have

$$\mu(2)<\frac{1-2\alpha}{4-4\alpha},$$

hence

$$\mu(1)=\frac{1}{2}-\mu(2)>\frac{1}{4-4\alpha},$$

so

$$e_{12}\leqq\frac{\mu(2)}{\mu(1)}<1-2\alpha.$$

Thus $e_{12}+e_{13}+e_{14}<1-2\alpha+\alpha=1-\alpha$, so $e_{11}>\alpha$.
But

$$\forall n\in N,\ \mu(1)e_{11}^{n}=\mu\left(\bigcap_{k=0}^{n}T^{k}(1)\right)\leqq\mu\left(\bigcap_{k=0}^{n}T^{k}(C)\right)=\frac{1}{2}\alpha^{n},$$

whence since $\mu(1)>0$, $e_{11}\leqq\alpha$. Contradiction. Q.E.D.

Corollary 3. $M\neq\bar{d}$.

Proof. Fór $0<\alpha<1/2$, $\bar{d}((\alpha),(1-\alpha))=(1-2\alpha)/2<(1-2\alpha)/(2-2\alpha)=M((\alpha),(1-\alpha))$. Q.E.D.

The construction given in Proposition 1 can be extended to pairs of non-symmetriç two-state Markov processes which have the same square (but are not themselves the same): to change a generic sequence for one process into a generic sequence for the other change only letters in odd positions (and make as few changes as necessary). The construction always yields a distance which is less than the "Markov distance" for such pairs. Except for the case covered by Proposition 1, however, the construction does not attain the $\bar{d}$-distance.

We have recently proved that the $\bar{d}$-distance between two-state Markov processes with positive entropy[†] is always less than the "Markov distance" between them unless the "Markov distance" is the partition distance (a paper is in preparation). Thus by Lemmas 2.2 and 2.7 of [1] the pairs of two-state Markov processes with positive entropy for which $\bar{d} = M$ are precisely those given by Propositions 2.3 and 2.5 of [1].

References

1. M. H. Ellis, *Distances between two-state Markov processes attainable by Markov joinings*, submitted for publication.

2. D. S. Ornstein, *An application of ergodic theory to probability theory*, Ann. Probability **1** (1973), 43–58.

3. D. S. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems*, Yale University Press, New Haven, 1974.

State University of New York at Albany
1400 Washington Avenue
Albany, N.Y. 12222, USA

[†] If either or both processes has entropy zero it is not hard to show that $M = \bar{d}$.