MIXTURES OF MARKOV PROCESSES[1]

By David A. Freedman[2]

Princeton University

1. Introduction. Kriloff and Bogoliouboff proved in [4] that stationary stochastic processes may be represented as mixtures of metrically transitive processes. It follows easily (as in [3]) from the strong law of large numbers that this representation is unique. A more illuminating proof is presented here for mixtures of Markov processes (even without assuming stationarity). All processes under discussion are natural number valued, and their time parameter ranges over the natural numbers.

The proof is based on the cycles of a process, which are the (finite) sequences of states beginning and ending at the same state. There is a function on the space of cycles which is associated with each process; namely, assign to a cycle the conditional probability of observing it given its initial state. In Section 2 it will be shown that by and large this association is $1 - 1$ for Markov processes, so that the transition probabilities may be determined from the cycle probabilities. Hence a probability distribution over a family of Markov processes may be thought of as a distribution in the space of cycle probabilities. If the mixture of these processes with respect to some distribution is known, so are the probabilities of repeating given cycles a given number of times. But these are the moments of the distribution of cycle probabilities, and determine that distribution. Thus the mixing distribution can be recovered from the mixture, and if a process can be represented as a mixture of Markov processes, there is essentially only one way to do this. Section 2 below discusses the cycles, and Section 3 gives the uniqueness theorem. The terminology and theory of [2, Chapter 15] is used.

2. Cycles. The symbol $\mathbf{c}$ is reserved for specific cycles, and c for a variable whose domain is some set of cycles. If $\mathbf{c}$ is the sequence $(i_j : 1 \leqq j \leqq n + 1)$, where $i_1 = i_{n+1}$, it is said to be of length n, and its cycle probability is defined as $k_{\mathbf{c}}(p) = p_{i_1 i_2} \cdots p_{i_n i_{n+1}}$, where $p \in [0, 1]^{Z \times Z}$ and Z is the set of natural numbers. Here and throughout the paper, subscripts on a point in a product space indicate its coordinates. Usually, p is a matrix of transition probabilities associated with a Markov process, with the understanding that if i or j is not in the state space of the process, $p_{ij} = 0$. The domains of c are C_i, the set of cycles beginning and ending at i; C_i^n, the subset of C_i of cycles of length n; and C_{ij}^n, the subset of C_i^n of cycles whose first transition is (i, j); ${}_jC_{ij}^n$ is the subset of C_{ij}^n of cycles which pass through j only once. Then

Theorem 1. *The following features of a stochastic matrix are determined by its*

Received May 3, 1961.

[1] The author wishes to acknowledge the very generous financial support provided by McGill University and the RAND Corporation.

[2] Now at the University of California, Berkeley.

cycle probabilities:

(i) *the classification of states into transient, persistent null, and persistent non-null*;

(ii) *the partitioning of the persistent states into irreducible classes*;

(iii) *the transition probabilities within each class of persistent states.*

Moreover, if (i) *and* (ii) *are known, then* (iii) *follows from a knowledge of the cycle probabilities through any one state of each irreducible class.*

PROOF.

(i) The classification of the state j depends only on the asymptotic behavior of $p_{jj}^{(n)}$. But

$$p_{jj}^{(n)} = \sum_{c \varepsilon C_j^n} k_c . \tag{1}$$

(ii) The states i and j belong to the same irreducible class if and only if there is a $\mathbf{c} \varepsilon C_i$ passing through j with $k_{\mathbf{c}} > 0$.

(iii) It suffices to consider an irreducible stochastic matrix of persistent states.

Suppose the k_c with $c \varepsilon C_1$ are given. Let $\mathbf{c} \varepsilon C_i$ but not εC_1. There is a $\mathbf{c}_0 \varepsilon C_1$ passing through i, with $k_{\mathbf{c}_0} > 0$. But $k_{\mathbf{c}_0}$ and $k_{\mathbf{c}_0} \cdot k_{\mathbf{c}}$ together determine $k_{\mathbf{c}}$, and these two quantities are known. Hence all the cycle probabilities are fixed.

If the class is non-null, let (Ces, 1) indicate the first Cesaro method of limitation, and denote the unique stationary initial probability distribution by p_i^*. Then

$$\begin{aligned} p_i^* &= (\text{Ces}, 1) \lim\nolimits_{n\to\infty} p_{ii}^{(n)} > 0 \\ p_{ij} &= p_i^{*-1} (\text{Ces}, 1) \lim\nolimits_{n\to\infty} p_{ij} p_{ji}^{(n)}, \end{aligned} \tag{2}$$

and $p_{ii}^{(n)}$ and $p_{ij}p_{ji}^{(n)}$ can be calculated from the cycle probabilities:

$$p_{ij}p_{ji}^{(n)} = \sum_{c \varepsilon C_{ij}^{n+1}} k_c . \tag{3}$$

If the class is null, the calculation is similar but less elementary, using Derman's stationary initial infinite "probability" distribution. The notation and theory of [1, Section I.9] will be used. From equation (18) of the reference,

$$_jp_{ji}^* = \lim_{N\to\infty} \frac{\sum_{n=0}^{N} p_{ii}^{(n)}}{\sum_{n=0}^{N} p_{jj}^{(n)}},$$

with $0 < {}_jp_{ji}^* < \infty$. Moreover, from the definition of $_jp_{ji}^*$,

$$p_{ij} = ({}_jp_{ji}^*)^{-1} \sum_{n=1}^{\infty} p_{ij}\, {}_jp_{ji}^{(n)} \tag{4}$$

and

(5) $$p_{ij}\,{}_jp_{ji}^{(n)} = \sum_{c\,\varepsilon\,{}_jC_{ij}^{n+1}} k_c .$$

Hence the transition probabilities have been computed from the cycle probabilities, finishing the proof.

The precise form of the theorem needed in Section 3 is stated as a

COROLLARY. *If a sub-stochastic matrix p has only one irreducible class E of persistent states and if* $i \varepsilon E$ *then* $\{k_c(p) : c \varepsilon C_i\}$ *determines*

(ii′) *whether* $j \varepsilon E$ *or* $j \varepsilon' E$, *for any* $j \varepsilon Z$;

(iii′) *the* p_{jk} *for* $j, k \varepsilon E$.

PROOF. Enlarge the matrix by adding the state ∞, with $p_{i\infty} = 1 - \sum_j p_{ij}$, $p_{\infty i} = 0$, $p_{\infty\infty} = 1$.

The new matrix is stochastic, and Theorem 1 applied to it gives (iii′). The argument for part (ii) implies part (ii′).

3. The uniqueness theorem. Let $\mathfrak{S} = Z^Z$; $B(\mathfrak{S})$ be the smallest σ-algebra of subsets of $\mathfrak{S}$ containing the cylinder sets.

By the Kolmogorov consistency theorem [5, p. 93], the marginal distribution functions of each process induce a probability on $B(\mathfrak{S})$. Let Λ be some family of processes and let $\{P_\lambda : \lambda \varepsilon \Lambda\}$ be the corresponding family of probabilities induced on $B(\mathfrak{S})$. Take $B(\Lambda)$ to be the smallest σ-algebra of subsets of Λ for which all the functions $\{P_\lambda(E) : E \varepsilon B(\mathfrak{S})\}$ are measurable. If μ is any probability on $B(\Lambda)$, the functional P_μ on $B(\mathfrak{S})$ is defined by

$$P_\mu(E) = \textstyle\int_\Lambda P_\lambda(E)\, d\mu : E \varepsilon B(\mathfrak{S}).$$

Then P_μ is itself a probability, and is called a mixture of the probabilities of P_λ, while μ is called the mixing distribution. Theorem 2 below will recapture μ from P_μ, when Λ is a family of Markov processes with no transient states; a restriction assumed in the balance of the paper.

Simple examples show that this is not always possible if some member of Λ contains transient states. Moreover, the way in which the irreducible components of the chains in Λ are combined to form general chains, or the way in which the initial distributions corresponding to each irreducible component in Λ are distributed (provided their average is fixed) obviously cannot affect P_μ. However, apart from these inevitable ambiguities, μ is uniquely determined by P_μ.

To state this precisely, let $M' = [0, 1]^Z$ and let $L' = [0, 1]^{Z \times Z}$, with $B(M')$ and $B(L')$ the products of the σ-algebras of Borel subsets of [0, 1]. Give L' the product topology. Let $M = \{p : p \varepsilon M' \text{ and } \sum_i p_i = 1\}$, while L is the set of all p in L' such that

(i) $\sum_j p_{ij} = 0$ or $1 : i \varepsilon Z$;

(ii) if $\sum_j p_{ij} = 0$ for a point p, then $\sum_j p_{ji} = 0$;

(iii) a Markov process with transition probabilities p_{ij} has no transient states, and only one irreducible class of persistent states; as usual, if $\sum_j p_{i_0 j} = 0$, the state i_0 is to be deleted from the state space.

Define $B(M) = M \cap B(M')$ and $B(L) = L \cap B(L')$.

LEMMA 1.

(i) *$M \varepsilon B(M')$ and $L \varepsilon B(L')$;*

(ii) *any probability μ on $B(\Lambda)$ induces a probability m on $B(L)$ and a measurable transformation $f: L \to M$ in such a way that*

$$(6) \qquad P_\mu(s:s \varepsilon \mathfrak{S}, s_i = j_i :1 \leqq i \leqq n) = \int_L f(p)_{j_1} \prod_{i=1}^{n-1} p_{j_i j_{i+1}} \, dm(p).$$

Moreover, f has the property that $\sum_j p_{i_0 j} = 0$ implies $f(p)_{i_0} = 0$.

PROOF. Since the proof is tedious but straightforward, it will not be argued in detail. For part (i), p_i is a continuous and hence measurable function of $p \varepsilon M'$. Thus $\sum_i p_i$ is a measurable function, and M is a measurable set. A similar argument shows that conditions (i) and (ii) on L restrict p to a measurable subset of L'. To deal with (iii), $p_{ij}^{(n)}$ is a measurable function of p, and hence $\sum_n p_{ii}^{(n)} = \infty$ and $\sup_n p_{ij}^{(n)} > 0$ define measurable subsets of L'. The details of part (ii) will be found in [3]. The procedure is to expand Λ to $\Lambda \times 2^Z$ by splitting each process into its irreducible components (λ, z) with $z \varepsilon 2^Z$ and (λ, z) having z as its (one) irreducible class. This produces a probability μ' in $\Lambda \times 2^Z$. Let T map $\Lambda \times 2^Z$ into L, assigning to each process its matrix of transition probabilities, with $p_{ij} = p_{ji} = 0$, all $j \varepsilon Z$, if i is not in the state space z. Then T is measurable, $m = \mu' T^{-1}$, and $f(p)_i = E(P_{(\lambda,z)}(s_1 = i) \mid T)$, the conditional expectation being with respect to the probability μ'. The reader unwilling to accept this argument may restrict his attention to families of Markov processes indexed by their matrices of transition probabilities.

The uniqueness result is

THEOREM 2. *The mixture P_μ uniquely determines m and f* a.e. $[m]$.

PROOF. Let $L_i = \{p: p \varepsilon L \text{ and } \sum_j p_{ij} = 1\}$ and put

$$m_i(E) = \int_E f(p)_i \, dm = \int_{L_i \cap E} f(p)_i \, dm, \qquad \text{for } E \varepsilon B(L).$$

Let $B(L_i) = L_i \cap B(L)$. Since $\sum_i f(p)_i = 1$, it follows that $m = \sum_i m_i$ and $f(p)_i = dm_i / dm$ a.e. $[m]$, so that it suffices to compute m_i on $B(L_i)$ from P_μ to complete the proof.

To do this, it is advantageous to pass to the space of cycle probabilities. Let $\mathbf{L}_i' = \prod\{[0, 1]_c : c \varepsilon C_i\}$ and define u to be the natural map from L' to $\mathbf{L}_i'$, which associates with every matrix its cycle probabilities, i.e., $u(p)_c = k_c(p)$, $p \varepsilon L'$, $c \varepsilon C_i$. Put $u_i = u$ restricted to L_i, $\mathbf{L}_i = u_i(L_i)$; define $B(\mathbf{L}_i')$ as the product of the σ-algebras of the Borel subsets in [0, 1], $B(\mathbf{L}_i) = \mathbf{L}_i \cap B(\mathbf{L}_i')$. Give $\mathbf{L}_i'$ the product topology.

But then u_i is a $1 - 1$ map of $(L_i, B(L_i))$ onto $(\mathbf{L}_i, B(\mathbf{L}_i))$ which is measurable in both directions. Indeed, it is onto by the definition of $\mathbf{L}_i$ and $1 - 1$ by Theorem 1. The measurability of u_i follows from the continuity of u. The measurability of u_i^{-1} follows from the equations (2) and (4) and the fact that the right-hand sides of (1), (3), and (5) are continuous, hence measurable, functions on $\mathbf{L}_i'$.

Let $N = \{n_c : c \varepsilon C_i\}$ be a set of non-negative integers, all but a finite number vanishing. With the convention that $0^0 = 1$,

$$(7) \qquad \int_{\mathbf{L}_i} \prod_{c \varepsilon C_i} k_c^{n_c} \, dm_i u_i^{-1}(k_c) = \int_{I_i} \prod_{c \varepsilon C_i} (k_c(p))^{n_c} \, dm_i(p) = P_\mu(F_N),$$

where F_N is the cylinder set of $s \varepsilon \mathfrak{S}$ whose coordinates, starting with $s_1 = i$, give rise to n_c cycles c in some specified order. Hence P_μ fixes the left-hand side of (2), for any N. But knowing the left-hand side of (7) as a function of N is equivalent to knowing $m_i u_i^{-1}$. For example, $\mathbf{L}_i'$ is compact by Tychonoff's Theorem [6, Theorem 5D] and the algebra generated by $\prod_{c \varepsilon C_i} k_c^{n_c}$ as N varies separates points in $\mathbf{L}_i'$. Hence the Stone-Weierstrass theorem applies [6, Theorem 4E], and the Riesz representation theorem [6, Section 16A] gives the desired result. That is to say, $m_i u_i^{-1}$, and hence m_i, may be computed from P_μ, and this completes the proof.

4. Acknowledgments. This paper is based on the author's doctoral dissertation, written under the very helpful guidance of Professor William Feller.

REFERENCES

[1] Kai-lai Chung, *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin, 1960.

[2] William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd ed., John Wiley and Sons, New York, 1957.

[3] David A. Freedman, "Mixtures of stochastic processes," unpublished doctoral dissertation, Princeton, 1960.

[4] N. Kriloff and N. Bogoliouboff, "La théorie générale de la mesure dans son application a l'étude des systèmes dynamique de la mécanique non-linéaire," *Ann. Math.*, 2nd Ser., Vol. 38 (1937), pp. 65–113.

[5] Michel Loève, *Probability Theory*, D. Van Nostrand, New York, 1955.

[6] Lynn H. Loomis, *An Introduction to Abstract Harmonic Analysis*, D. Van Nostrand, New York, 1955.