THE SUPPORT OF DISCRETE EXTREMAL MEASURES WITH GIVEN MARGINALS

J. L. Denny

Let X and Y be countable sets. Let P be a probability on $X \times Y$ with marginals (projections) P_X on X and P_Y on Y. The set $M(P)$ of all probability measures on $X \times Y$ with marginals P_X and P_Y is convex. The purpose of this note is to characterize the supports of the extremal measures of $M(P)$, that is, the extreme points of $M(P)$. We show that each support of an extremal measure is the union of the graphs of two functions f and g, f mapping a subset of X into Y and g mapping a subset of Y into X, so that for each integer n the composition $(g \circ f)^n$ has no fixed points. Conversely, each probability P supported on the union of such graphs is an extremal measure of $M(P)$.

Now assume only that μ is a signed measure on $X \times Y$ and define μ_X, μ_Y, and $M(\mu)$ analogously. It is also true that each extremal $\nu \in M(\mu)$ assigns all its mass to the union of two such graphs. In this situation there may not be any extremal measures. However, even when extremal measures do exist the above converse is not true, for nonfinite positive μ.

As an application we characterize the extreme points of the set of Markov matrices with a given stationary invariant probability measure. When the Markov matrix is finite ($N \times N$) and doubly stochastic the extreme points are the permutation matrices. This result is known as the Birkhoff-von Neumann theorem (see, for example, p. 189 of [9]) and is a corollary to results in this paper. As pointed out by Letac [5], when P is a probability measure $M(P)$ is compact in the weak* topology and so the Choquet representation is applicable (see [2] or [8]). For other applications we refer to the survey paper [7] and for the case of non-discrete measures we mention [1] and [10]. This paper uses a method of Letac [5] which he employed to obtain another characterization of the supports of discrete measures.

The following is the usual definition of composition of functions. Some notation is needed. Let $A \subset X$ and $B \subset Y$ with at least one of A or B nonempty. Let $f: A \to Y$ and $g: B \to X$. Let $D_1 = A \cap f^{-1}(B)$. If D_1 is nonempty set $(g \circ f)^1(x) = g(f(x))$, $x \in D_1$. Assuming D_{n-1} and $(g \circ f)^{n-1}$ are defined set

$$D_n = D_{n-1} \cap ((g \circ f)^{n-1})^{-1}[D_1]$$

and $(g \circ f)^n(x) = (g \circ f)^1((g \circ f)^{n-1}(x))$, $x \in D_n$.

Definition. The pair of functions (f,g) is aperiodic if for each n, $x \in D_n$ implies $(g \circ f)^n(x) \neq x$.

THEOREM 1. *Let P be a probability on $X \times Y$ and let $Q \in M(P)$. The following two assertions are equivalent:*

Received May 1, 1978.

Michigan Math. J. 27 (1980).

(i) *Q is an extreme point of $M(P)$;*

(ii) *There exists a pair of aperiodic functions (f,g) such that $Q(G) = 1$, where* $G = \{(x, f(x)) : x \in A\} \cup \{(g(y),y) : y \in B\}$.

THEOREM 2. *Let μ be a signed measure on* $X \times Y$. *If v is an extreme point of $M(\mu)$ then there is a pair of aperiodic (f,g) such that $v(G^c) = 0$.*

Before proving the theorems, we note that (f,g) is aperiodic if either A or B is empty and that the graphs of aperiodic (f,g) are disjoint subsets of $X \times Y$. To prove that (ii) implies (i) in Theorem 1 we use a special case of a theorem discovered independently by R. G. Douglas [3] and J. Lindenstrauss [6]:

THEOREM. *Let* $L = \left\{\phi + \psi : \phi : X \to R, \psi : Y \to R, \int |\phi| P_X < \infty, \int |\psi| P_Y < \infty\right\}$. *Then, $Q \in M(P)$* is *an extreme point if and only if L is norm-dense in $L_1(Q)$.*

LEMMA 1. *Let (f,g) be an aperiodic pair with A non-empty. Let $x_0 \in A$ and define $A_0 = \{x_0\}$. For each integer $i \geq 1$ define $B_i = g^{-1}(A_{i-1}) \cap B$ and $A_i = f^{-1}(B_i) \cap A$. Then $A_m \cap A_k$ is empty* if $m > k$.

Proof. If $x \in A_m \cap A_k$ then $(g \circ f)^m(x) = (g \circ f)^k(x) = x_0$. This implies $(g \circ f)^{m-k}(x_0) = x_0$ which contradicts the aperiodicity.

To prove that (ii) implies (i) we prove that

$$\int |h - (\phi + \psi)| Q \tag{1}$$

can be made arbitrarily small for $h \in L_1(Q)$, $\phi + \psi \in L$. By standard L_1 facts it is enough to prove this for $h = I_B$, the indicator function of a finite subset $B \in X \times Y$. Since L is a linear space it is enough to prove this for $B = \{z\}$, $z \in X \times Y$. We can assume $z \in G$. If A is empty and $z = (g(y),y)$, define $\phi \equiv 0$ and $\psi = I_{\{y\}}$. Then (1) equals zero. Otherwise assume $z = (x_0, f(x_0))$. Define $\psi_0 \equiv 0$, $\phi_i = I_{A_i}$, $\psi_i = I_{B_i}$ for $i \geq 1$ where A_i and B_i are defined in Lemma 1. Since the A_i are pairwise disjoint the B_i are pairwise disjoint and consequently the subsets of $X \times Y$, $B_i' = \{(g(y),y) : y \in B_i\}$, are pairwise disjoint. By an easy computation we have

$$\int \left| h - \left(\sum_{i=0}^{n} (\phi_i - \psi_i) \right) \right| Q = Q(B'_{n+1}). \tag{2}$$

Since $Q(B'_{n+1}) \to 0$ as $n \to \infty$ it follows that (ii) implies (i).

We show (i) implies (ii). The proof of Theorem 2 is the same. The following notation is used: $(g \circ f)^0(x) = x$, $G(f) = \{(x, f(x)) : x \in A\}$, and

$$G(g) = \{(g(y),y) : y \in B\}.$$

LEMMA 2. *Let Q be an extreme point of $M(P)$ and let $U = \{(x,y) : Q(\{x,y\}) > 0\}$. Let $f : A \to Y$ and $g : B \to X$ satisfy* (i) *$G(f) \cup G(g) \subset U$ and* (ii) *$G(f) \cap G(g)$ is empty. Then for each $n \geq 1$, $x \in D_n$ implies*

$$(g \circ f)^m(x) \neq (g \circ f)^p(x)$$

if $0 \leq p < m \leq n$. *In particular,* $f((g \circ f)^m(x)) \neq f((g \circ f)^p(x))$ if $0 \leq p < m \leq n - 1$.

Proof. For $n = 1$ this follows from the disjointness of $G(f)$ and $G(g)$. Assume true for n and let $x \in D_{n+1}$. Since $D_{n+1} \subset D_n$ it is enough to prove

$$(g \circ f)^{n+1}(x) \neq (g \circ f)^p(x),$$

$0 \leq p \leq n$.If $(g \circ f)^{n+1}(x) = (g \circ f)^p(x)$ for $1 \leq p \leq n$ then as

$$(g \circ f)^p(x) \in D_{n+1-p}$$

we have

$$(g \circ f)^{n+1-p}((g \circ f)^p(x)) = (g \circ f)^p(x),$$

which contradicts the inductive hypothesis. It is therefore enough to prove that $(g \circ f)^{n+1}(x) = x$ contradicts the extremality of Q (the following argument essentially appears on p. 501 of [5]).

Define $z_k = (x_k, y_k)$ for $k = 0, \ldots, 2n + 1$ by

$$z_{2p} = ((g \circ f)^p(x), f((g \circ f)^p(x))),$$
$$z_{2p+1} = ((g \circ f)^{p+1}(x), f((g \circ f)^p(x))).$$

The inductive hypothesis implies:

(i) for $1 \leq k \leq 2n - 1$, $k < j \leq 2n$, $x_k = x_j$ if and only if k is odd and $j = k + 1$;

(ii) for $0 \leq k \leq 2n - 2$, $k < j \leq 2n - 1$, $y_k = y_j$ if and only if k is even and $j = k + 1$;

(iii) $x_0 \neq x_k$, $1 \leq k \leq 2n$ and $y_{2n} \neq y_k$, $0 \leq k \leq 2n - 1$.

If $(g \circ f)^{n+1}(x) = x$ then

(iv) $x_{2n+1} = x_0$, $y_{2n+1} = y_{2n}$.

In the terminology of Letac [5], $C = \{z_0, \ldots, z_{2n+1}\}$ is a cycle. Let

$$\alpha = \min \{Q(\{z_k\}) : k = 0, \ldots, 2n + 1\},$$

which is positive since $C \subset U$. Define the signed measure m_1 on

$$C: m_1(\{z_k\}) = (-1)^k \alpha / 2, k = 0, \ldots, 2n + 1.$$

Let $m_2 = -m_1$. The marginals of the m_i are identically zero, by (i) – (iv). Let $Q_i = Q + m_i$, a probability measure with the same marginals as Q. Since $Q = 1/2(Q_1 + Q_2)$ we have the desired contradiction and the proof is completed.

LEMMA 3. *Let extremal* $Q \in M(P)$ *and let* $U = \{(x,y): Q(\{x,y\}) > 0\}$. *For each finite subset* $V \subset U$ *there is* $\bar{f}: \bar{A} \to Y$, $\bar{g}: \bar{B} \to X$ *so that*

$$V = G(\bar{f}) \cup G(\bar{g}).$$

Proof. This is clear if V contains one element. We proceed by induction and let $V = \{(x_k,y_k): k = 1, \ldots, n\} = G(f) \cup G(g)$ where $f: A \to Y$ and $g: B \to X$. We want a representation for $V \cup \{x_{n+1}, y_{n+1}\}$.

If $x_{n+1} \in A^c$ define $\bar{A} = A \cup \{x_{n+1}\}$, $\bar{f} = f$ on A, $\bar{f}(x_{n+1}) = y_{n+1}$, $\bar{B} = B$, and $\bar{g} = g$. Make the analogous definition if $y_{n+1} \in B^c$.

Otherwise, let $D_0 = A$ and let D_m be defined as above, the domain of $(g \circ f)^m$. If $m \geq 0$ and $x_{n+1} \in D_m \sim D_{m+1}$ then either

$$(g \circ f)^m (x_{n+1}) \in A^c \tag{3}$$

or else

$$f((g \circ f)^m (x_{n+1})) \in B^c. \tag{4}$$

In addition, if $x_{n+1} \in D_m$, $m \geq 1$, Lemma 2 implies that the $((g \circ f)^p(x_{n+1}), f((g \circ f)^p(x_{n+1})))$, $0 \leq p \leq m - 1$ are distinct elements of $G(f)$. Since $G(f)$ is finite we assume that for some $m \geq 0$ $x_{n+1} \in D_m \sim D_{m+1}$.

If $x_{n+1} \in D_0 \sim D_1$ then (4) holds. Define $\bar{B} = B \cup \{y_{n+1}\}$, $\bar{g} = g$ on B, $\bar{g}(y_{n+1}) = x_{n+1}$, $\bar{A} = A$, and $\bar{f} = f$.

Assume $x_{n+1} \in D_m \sim D_{m+1}$, $m \geq 1$. We consider only case (3), for (4) is analogous. Let

$$A_1 = \{(g \circ f)^p(x_{n+1}): p = 0, \ldots, m\} \subset A \cup \{(g \circ f)^m(x_{n+1})\},$$
$$B_1 = \{f((g \circ f)^p(x_{n+1})): p = 0, \ldots, m - 1\} \subset B.$$

Define $\bar{A} = A \cup \{(g \circ f)^m(x_{n+1})\}$ and $\bar{B} = B$. Define $\bar{f}$ on A_1 by $\bar{f}(x_{n+1}) = y_{n+1}$, $\bar{f}((g \circ f)^p(x_{n+1})) = f((g \circ f)^{p-1}(x_{n+1}))$ for $1 \leq p \leq m$. Define $\bar{f}$ on $A \sim A_1$ by $\bar{f} = f$. Similarly, define $\bar{g}$ on B_1 by

$$\bar{g}(f((g \circ f)^p(x_{n+1}))) = (g \circ f)^p(x_{n+1}), \qquad 0 \leq p \leq m - 1.$$

Define $\bar{g}$ on $B \sim B_1$ by $\bar{g} = g$. By Lemma 2, $\bar{f}$ and $\bar{g}$ are well-defined.

We show that $V \cup \{x_{n+1}, y_{n+1}\} = G(\bar{f}) \cup G(\bar{g})$. Suppose $(x,y) \in G(\bar{f})$. If $x \in A_1$, $x \neq x_{n+1}$, then $(x,y) = ((g \circ f)^p(x_{n+1}), \bar{f}((g \circ f)^p(x_{n+1}))) = ((g \circ f)^p(x_{n+1}), f((g \circ f)^{p-1}(x_{n+1})))$, for some $1 \leq p \leq m$, which belongs to $G(g) \subset V$. Suppose $(x,y) \in G(f)$. If $x \in A_1 \cap A$ then $(x,y) = ((g \circ f)^p(x_{n+1}),$

$$f((g \circ f)^p(x_{n+1}))) = (\bar{g}(f((g \circ f)^p(x_{n+1}))),$$

$f((g \circ f)^p(x_{n+1})))$, for some $0 \leq p \leq m - 1$, which belongs to $G(\bar{g})$. The other cases are similar, and the proof is completed.

The straightforward proof of the next lemma is omitted.

LEMMA 4. *Let graphs* $G(f_i) \cup G(g_i) \subset G(f_{i+1}) \cup G(g_{i+1})$, i = 1, 2, ... *Then there exist* $f: A \to Y$ *and* $g: B \to X$ *so that* $\bigcup_i \{G(f_i) \cup G(g_i)\} = G(f) \cup G(g)$.

Lemmas 2, 3, and 4 show that (ii) implies (i), since there is clearly no loss of generality in assuming $G(f) \cap G(g)$ is empty, where $G(f) \cup G(g) = U$.

We characterize the extreme points of the set of Markov matrices with a given stationary invariant probability measure. Let P be a probability on the nonnegative integers Z^+, say, and let $p_i = P(i) > 0$. Denote by $[p_{ij}]$ a Markov matrix for which P is a stationary invariant measure: $\sum_i p_i p_{ij} = p_j$, $j \in Z^+$. Let $\mathscr{P} = \{[p_{ij}]$: P is a stationary invariant measure for $[p_{ij}]\}$. The convex set $\mathscr{P}$ is linearly isomorphic to $M(P \times P)$, the set of probability measures on $Z^+ \times Z^+$ with marginals P, and we have

COROLLARY 1. $[p_{ij}] \in \mathscr{P}$ *is an extreme point if and only if there is aperiodic* (f,g) *so that* $p_{ij} > 0$ *is equivalent to* $(i,j) = (x,f(x))$ *or* $(i,j) = (g(y),y)$.

Recall that a Markov matrix is doubly stochastic if each column sums to one.

COROLLARY 2. (Birkhoff-von Neumann Theorem). *The permutation matrices are the extreme points of the* $N \times N$ *doubly stochastic matrices.*

Proof. Let $X = Y = \{1, \ldots, N\}$. Identify the doubly stochastic matrices with positive measures on $X \times Y$. If Q is extremal and U is the support of Q then $U = G(f) \cup G(g)$ for aperiodic (f,g). If either $G(f)$ or $G(g)$ is empty then the other graph is clearly a permutation matrix. We claim there is no loss of generality in assuming the domain A of f is X. For if $x \in X \sim A$, $g^{-1}\{x\}$ is nonempty since Q is doubly stochastic. Choose $y \in g^{-1}\{x\}$, set $f(x) = y$, and restrict g to $B \sim \{y\}$. Proceeding in this way we get aperiodic (f',g') with the domain of f' equal X and $U = G(f') \cup G(g')$. Next, we verify the range of f' must equal Y. For if $y \in (f'(X))^c$, then $Q(g'(y),y) = 1$ and $Q(g'(y),f'(g'(y))) = 0$ since Q is doubly stochastic. The latter equality contradicts the definition of U. The domain B of g' cannot equal Y since (f',g') is aperiodic and X is finite. We show B is empty, and the result follows since f' is one-to-one. If $y \in B$, $y = f'(x)$, then there is a smallest positive k so that $f'((g' \circ f')^k(x)) \in B^c$, using aperiodicity. Since f' is onto $Q(g'(y),y) < 1$, and this implies that $Q((g' \circ f')^k(x), f'((g' \circ f')^k(x))) < 1$. This last inequality contradicts the row sums each being equal to one.

D. Kendall shows, with an addendum by J. Kiefer, that the Birkhoff-von Neumann theorem is true for infinite doubly stochastic matrices [4]. Since each infinite permutation matrix is the graph of one function it can easily be shown that the converse to Theorem 2 does not hold.

REFERENCES

1. J. R. Brown and R. C. Shiflett, *On extreme doubly stochastic measures.* Michigan Math. J. 17 (1970), 249–254.

2. G. Choquet and P. A. Meyer, *Existence et unicité des représentations intégrales dans les convexes compacts quelconques.* Ann. Inst. Fourier (Grenoble) 13 (1963) no. 1, 139–154.

3. R. G. Douglas, *On extremal measures and subspace density.* Michigan Math. J. 11 (1964), 243–246.

4. D. G. Kendall, *On infinite doubly-stochastic matrices and Birkhoff's problem* 111. J. London Math. Soc. 35 (1960), 81–84.

5. G. Letac, *Représentation des mesures de probabilité sur le produit de deux espaces dénomberables, de marges données.* Illinois J. Math. 10 (1966), 497–507.

6. J. Lindenstrauss, *A remark on extreme doubly stochastic measures.* Amer. Math. Monthly 72 (1965), 379–382.

7. L. Mirsky, *Results and problems in the theory of doubly-stochastic matrices.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 1 (1962/1963), 319–334.

8. R. R. Phelps, *Lectures on Choquet's Theorem.* D. van Nostrand Co. Inc., Princeton, N.J., 1966.

9. G.-C. Rota and L. H. Harper, *Matching theory, an introduction.* Advances in Probability 1, Marcel Dekker, Inc., New York (1971), 169–215.

10. T. L. Seethoff and R. C. Shiflett, *Doubly stochastic measures with prescribed support.* Z. Wahrscheinlichskeitstheorie und Verw. Gebiete, 41 (1978), 283–288.

Department of Mathematics
The University of Arizona
Tucson, Arizona 85721