April 12, 1995

Symmetric, Coherent, Choquet Capacities

Joseph B. Kadane[1]
and
Larry Wasserman[2]
Carnegie Mellon University

Choquet capacities are a generalization of probability measures that arise in robustness, decision theory and game theory. Many capacities that arise in robustness are symmetric or can be transformed into symmetric capacities. We characterize the extreme points of the set of upper distribution functions corresponding to coherent, symmetric Choquet capacities on [0,1]. We also show that the set of 2-alternating capacities is a simplex and we give a Choquet representation of this set.

1. Introduction. A *Choquet capacity* on a measurable space $(\Omega, \mathcal{B})$ is a mapping $C : \mathcal{B} \to [0,1]$ such that $C(\emptyset) = 0$. C is *coherent* if there exists a non-empty set of probability measures M such that $C(A) = \sup_{P \in M} P(A)$ for every $A \in \mathcal{B}$. Coherent capacities are also called *upper probabilities* (Walley 1991, Fine 1988, Dempster 1967, 1968, Smith 1961) or *upper envelopes* (Anger and Lembcke 1985, Denneberg 1994). Let $\Omega = [0,1]$, let $\mathcal{B}$ be the Borel subsets of Ω and let μ be Lebesgue measure. C is *symmetric* if $C(A) = C(B)$ whenever $\mu(A) = \mu(B)$. As we shall show, it is possible to say exactly when a symmetric capacity is coherent.

Many robustness models used in statistics involve symmetric, coherent capacities or can be transformed into the same by a smooth, one-to-one mapping (Buja 1986, Huber and Strassen 1973, Wasserman and Kadane 1990, Fortini and Ruggeri 1994). For example, the upper probability for an ϵ-contamination neighborhood around a probability measure P

[1]Research supported by NSF grants DMS-9303557, SES-8900025 and ONR contract N00014-89-J-1851.

[2]Research supported by NSF grants NSF grant DMS-9303557 and DMS-9357646 and NIH grant R01-CA54852-01.

AMS 1980 subject classifications. Primary 60A05.

Key words and phrases. Bayesian robustness, Choquet capacities, equimeasurability, lower probabilities, majorization, 2-alternating capacities.

(Berger 1984, 1990, 1994, Huber 1973, 1981) generates a symmetric capacity once the set of probabilities is transformed to the unit interval under the inverse integral transform corresponding to P. This is true for many neighborhoods. Capacities are also used in decision theory (Gilboa 1987, Schmeidler 1989) and game theory (Shapley 1971). Symmetric Choquet integrals, which are related to symmetric capacities, have been studied by Armstrong (1990) and Talagrand (1978). Symmetric capacities were studied in Wasserman and Kadane (1992) under the additional assumption that M consisted of non-atomic probabilities with bounded densities. Despite the ubiquity of capacities, there is little in the way of simple characterizations for capacities as there are for probabilities. This paper is concerned with such characterizations. In particular, we are interested in the following question: what are the extreme points in the set of all distribution functions corresponding to symmetric capacities?

A capacity is *2-alternating* if

$$C(A \cup B) \leq C(A) + C(B) - C(A \cap B) \tag{1}$$

for all $A, B \in \mathcal{B}$. Many capacities used in statistics are 2-alternating. Furthermore, the 2-alternating condition is crucial for many important results. For example, a particular generalization of the Neyman-Pearson lemma holds if and only if the capacity generated by the underlying models is 2-alternating (Huber and Strassen 1973). Similarly, a particular generalization of Bayes' theorem for capacities holds if and only if the capacity is 2-alternating (Wasserman and Kadane 1990). In game theory, 2-alternating capacities represent certain convex games (Shapley 1971). Most work on coherent capacities have focused on the 2-alternating case. Little is known about the non-2-alternating case. We shall consider the general case in sections 2 and 3 and the 2-alternating case in section 4.

The following is an outline of the paper and serves as a summary of the main contributions of this paper. In section two we give a majorization representation of symmetric capacities (Theorem 2.1) which generalizes a theorem in Wasserman and Kadane (1992). In section three, which is the main section of the paper, we study the distribution functions of symmetric capacities. There we establish (Lemmas 3.1 and 3.2) a one-to-one correspondence between distribution functions of symmetric capacities and functions α taking $[0,1]$ to $[0,1]$ that satisfy

$$\lim_{\omega \to 1} \int_{1/2}^{\omega} \frac{\alpha(u) - u}{u(1-u)} du = -\infty.$$

This correspondence allows us to characterize the extreme points in the set all distribution functions for symmetric capacities (Theorem 3.2). This is the main theorem of the paper and, loosely, it says that F is extreme if and only if the corresponding α function takes values 0 and 1 almost everywhere. In section four we identify the extreme points of the set of distribution functions for 2-alternating capacities (Theorem 4.1) and we give a Choquet representation for this set by identifying the unique mixing measure over the extreme points (Theorem 4.2). Closing remarks are contained in section five.

2. A Characterization of Symmetric Capacities. Let $\mathcal{P}$ be the set of all probability measures on $\mathcal{B}$, let $\mathcal{P}_a$ be the set of all $P \in \mathcal{P}$ that are absolutely continuous with respect to μ and let $\mathcal{P}_s$ be the set of all probability measures that are singular with respect to μ. Suppose that $P, Q \in \mathcal{P}_a$ and let $p = dP/d\mu$ and $q = dQ/d\mu$. We write $p \sim q$ if

$$\mu(\{\omega; p(\omega) > t\}) = \mu(\{\omega; q(\omega) > t\}) \tag{2}$$

for all real t. If (2) holds we say that p and q are *equimeasurable.* We shall also say that P and Q are equimeasurable and we will write $P \sim Q$. Given any $P \in \mathcal{P}_a$ with $p = dP/d\mu$, there exists a unique, non-increasing, right continuous function p^*, called the *decreasing rearrangement* of p, such that $p \sim p^*$ (Ryff 1965). We call the corresponding probability measure P^* the decreasing rearrangement of P.

Every $P \in \mathcal{P}$ may be written in terms of its Lebesgue decomposition $P = \alpha P_a + \overline{\alpha} P_s$ where $P_a \in \mathcal{P}_a$, $P_s \in \mathcal{P}_s$, $\alpha \in [0,1]$ and $\overline{\alpha} = 1 - \alpha$. We define the decreasing rearrangement of P by $P^* = \alpha P_a^* + \overline{\alpha}\delta_0$ where δ_0 is a point mass at 0. Note that if $P \in \mathcal{P}_a$ then this agrees with the earlier definition of decreasing rearrangement. We say that P is *majorized* by Q, denoted by $P \prec Q$, if $P^*([0,t]) \leq Q^*([0,t])$ for every real t. Majorization has been studied in discrete settings (Marshall and Olkin 1979) and continuous settings (Ryff 1963, 1965, 1967, 1970). Our definition is a natural generalization of previous definitions for probabilities with both absolutely continuous and singular components.

Let $M \subset \mathcal{P}$ be nonempty and let $C(A) = \sup_{P \in M} P(A)$. Throughout the rest of the paper we restrict attention to symmetric, coherent capacities. We say that P is *dominated by* C, written $P \lhd C$, if $P(A) \leq C(A)$ for all $A \in \mathcal{B}$.

THEOREM 2.1. *The following two statements are equivalent:*

(i) C is symmetric.

(ii) $P \lhd C$ and $Q \prec P$ imply that $Q \lhd C$.

Before proving the theorem, we establish some Lemmas.

LEMMA 2.1. *For every $P \in \mathcal{P}_a$ and every $t \in [0,1]$ there exists A_t such that (i) $\mu(A_t) = t$, (ii) $P^*([0,t]) = P(A_t)$ and (iii) $P(A_t) \geq P(B)$ for every B for which $\mu(B) = t$.*

PROOF. The lemma is obvious for $t = 0$ and $t = 1$ so assume $0 < t < 1$. Let $p = dP/d\mu$ and $p^* = dP^*/d\mu$. Define $A_0 = \{\omega; p(\omega) > p^*(t)\}$ and $A_{00} = \{\omega; p(\omega) \geq p^*(t)\}$. Then

$\mu(A_0) \le t \le \mu(A_{00})$ and $A_0 \subset A_{00}$. Let $u = \mu(A_0)$. Choose B such that $A_0 \subset B \subset A_{00}$ and $\mu(B) = t - u$. Let $A_t = A_0 \cup B$. Hence, $\mu(A_t) = t$ and $P(A_t) = P(A_0) + P(B) = P^*([0,u]) + p^*(u)(t-u) = P^*([0,t])$. Now let K be such that $\mu(K) = t$. Then,

$$\begin{aligned} P(A_t) - P(K) &= P(A_t - K) - P(K - A_t) \\ &\ge \mu(A_t - K)\text{ess inf}_{A_t - K}p(\omega) - \mu(K - A_t)\text{ess sup}_{K - A_t}p(\omega) \\ &= \mu(A_t - K)(\text{ess inf}_{A_t - K}p(\omega) - \text{ess sup}_{K - A_t}p(\omega)) \ge 0. \ \Box \end{aligned}$$

LEMMA 2.2. *For every $P \in \mathcal{P}$ and every $A \in \mathcal{B}$, $P(A) \le P^*([0,\mu(A)])$.*

PROOF. Let $t = \mu(A)$. Then, $P^*([0,\mu(A)]) = \alpha P_a^*([0,\mu(A)]) + \overline{\alpha} = \alpha P_a(A_t) + \overline{\alpha} \ge \alpha P_a(A) + \overline{\alpha} \ge \alpha P_a(A) + \overline{\alpha} P_s(A) = P(A)$ where A_t is as defined in Lemma 2.1. The second equality and the inequality that follows it are both due to Lemma 2.1. $\Box$

LEMMA 2.3. *If C is symmetric and $P \lhd C$ then $P^*([0,t]) \le C([0,t])$ for every t.*

PROOF. Let $P = \alpha P + \overline{\alpha} P_s$ and let S be the support of P_s. $P^*([0,t]) = \alpha P_a^*([0,t]) + \overline{\alpha} = \alpha P_a(A_t) + \overline{\alpha} = \alpha P_a(A_t \cup S) + \overline{\alpha} P_s(A_t \cup S) = P(A_t \cup S) \le C(A_t \cup S) = C([0,t])$. The second equality follows from lemma 2.1. The last equality follows since C is symmetric and $\mu(A_t \cup S) = \mu([0,t])$. $\Box$

PROOF OF THEOREM 2.1. (i) implies (ii). Let $P \lhd C$ and $Q \prec P$. For any A, $Q(A) \le Q^*([0,\mu(A)]) \le P^*([0,\mu(A)]) \le C([0,\mu(A)]) = C(A)$. This follows from, respectively, Lemma 2.2, $Q \prec P$, Lemma 2.3 and (i). Thus, $Q \lhd C$.

(ii) implies (i). Let A and B be such that $\mu(A) = \mu(B)$. There exists $P \lhd C$ such that $P(A) = C(A)$. Write, $P = \alpha P_a + \overline{\alpha} P_s$. Define

$$R_a(\cdot) = \frac{P_a(A)}{\mu(A)}\mu(\cdot \cap B) + \frac{P_a(A^c)}{\mu(A^c)}\mu(\cdot \cap B^c)$$

and define $R = \alpha R_a + \overline{\alpha} R_s$ where R_s is any singular measure such that $R_s(B) = P_s(A)$. By construction, $R(B) = P(A)$. Furthermore, $R \prec P$ so that, by (ii), $R \lhd C$. Thus, $C(A) = P(A) = R(B) \leq C(B)$. By a similar argument, $C(B) \leq C(A)$. □

Now we consider some examples of symmetric capacities.

EXAMPLE 2.1. ϵ-contamination. Let $M = \{(1-\epsilon)\mu + \epsilon Q; Q \in \mathcal{P}\}$ where $\epsilon \in [0,1]$. Then $C(A) = (1-\epsilon)\mu(A) + \epsilon$ if $A \neq \emptyset$. This model is used extensively in robustness (Huber 1973, 1981, Berger 1984).

EXAMPLE 2.2. total-variation. Let $d(P,Q) = \sup_{A \in \mathcal{B}} |P(A) - Q(A)|$ be the total variation distance between P and Q. Let $\{P; d(\mu, P) \leq \epsilon\}$. Then $C(A) = \min\{P(A) + \epsilon, 1\}$. This is also common in robustness.

EXAMPLE 2.3. density bounded class. Let $a < 1 < b$ and let M be the set of all P with densities p (with respect to μ) such that $a \leq p(\omega) \leq b$ for μ-almost all ω. Then $C(A) = \min\{b\mu(A), 1 - a\mu(A^c)\}$. Lavine (1991a, 1991b) uses this class in Bayesian robustness.

EXAMPLE 2.4. density ratio class. Let $k \geq 1$ and let M be the set of all P with densities p (with respect to μ) such that ess sup$p(\omega)$/ess inf$p(\omega) \leq k$. It turns out that $C(A) = k\mu(A)[k\mu(A) + \mu(A^c)]^{-1}$. This is a specialization of a class used by DeRobertis and Hartigan (1981) in Bayesian robustness.

3. Distribution Functions for Capacities. The distribution function for a symmetric capacity C is defined by $F(\omega) = C([0,\omega])$. Conversely, a function $F : [0,1] \to [0,1]$ defines a symmetric capacity by way of $C(A) = F(\mu(A))$. Distribution functions for capacities were used by Buja (1986) and Bednarski (1981) in a different context.

EXAMPLES 2.1-2.4 (continued). The distribution functions for these examples are, respectively, $F(\omega) = (1-\epsilon)\omega + \epsilon$, $F(\omega) = \min\{\omega + \epsilon, 1\}$, $F(\omega) = \min\{b\omega, 1 - a(1-\omega)\}$ and $F(\omega) = k\omega[k\omega + (1-\omega)]^{-1}$.

Let $gr\, F = \{(\omega, y) \in [0,1] \times [0,1]; y \leq F(\omega)\}$. If $a = (a_1, a_2)$ and $b = (b_1, b_2)$ are two points in the plane with $a_1 < b_1$, let $L_{a,b}(\omega) = a_2 + (\omega - a_1)(b_2 - a_2)/(b_1 - a_1)$ and $\mathcal{L}(a,b) = \{(\omega, L_{a,b}(\omega)); a_1 \leq \omega \leq b_1\}$. We say that F is *doubly star-shaped* if $a \in gr\, F$ implies that $\mathcal{L}(\mathbf{0}, a) \subset gr\, F$ and $\mathcal{L}(a, \mathbf{1}) \subset gr\, F$ where $\mathbf{0} = (0,0)$ and $\mathbf{1} = (1,1)$.

In general, it is difficult to know whether a capacity is coherent; see Anger and Lembcke (1985) for example. The next theorem characterizes coherent symmetric capacities in terms of their distribution functions.

THEOREM 3.1. *If C is symmetric and coherent then its distribution F is doubly star-shaped. Conversely, if $F : [0,1] \to [0,1]$ is doubly star-shaped then there exists a symmetric coherent capacity C such that $F(\omega) = C([0,\omega])$ for every $\omega \in \Omega$.*

PROOF. Let C be symmetric. Fix $A = [0,t]$, $t \in (0,1)$. There exists $P \lhd C$ such that $P(A) = C(A)$. Define $R(\cdot) = P(A)\mu(\cdot \cap A)/\mu(A) + P(A^c)\mu(\cdot \cap A^c)/\mu(A^c)$. Then $R \prec P$ hence $R \lhd C$. So, for every $\omega \in \Omega$, $G(\omega) \equiv R([0,\omega]) \leq C([0,\omega]) = F(\omega)$. Since, G is piecewise linear this implies that $\mathcal{L}(\mathbf{0}, a) \in gr\, F$ and $\mathcal{L}(a, \mathbf{1}) \in gr\, F$ where $a = (\omega, F(\omega))$. It follows that $gr\, F$ is doubly star-shaped.

Now we construct a class M that generates a symmetric capacity C with distribution F. For every measurable set A such that $0 < \mu(A) < 1$ define $P_A(\cdot) = F(\mu(A))\mu(\cdot \cap A)/\mu(A) + (1 - F(\mu(A)))\mu(\cdot \cap A^c)/\mu(A^c)$. For nonempty A such that $\mu(A) = 0$ define $P_A(\cdot) = F(0)\delta_{\omega_A} + (1 - F(0))\mu(\cdot)$ where $\omega_A \in A$. For A with $\mu(A) = 1$ define $P_A = \mu$. Let $M = \{P_A; A \in \mathcal{B}, A \neq \emptyset\}$ and $C(A) = \sup_{P \in M} P(A)$. We claim that $C(A) = P_A(A)$. First suppose that $A = [0, t]$, $t \in (0, 1)$. For any C, $P_A(A) \geq P_{\hat{C}}(A) \geq P_C(A)$ where $\hat{C} = [0, \mu(C)]$. The first inequality is from the fact that F is doubly star-shaped. The second follows from how P_A is defined. Hence, $P_A(A) = C(A)$. Now consider any A such that $\mu(A) > 0$. Let C be any set, let $\hat{A} = [0, \mu(A)]$ and let D be such that $\mu(D) = \mu(C)$ and $\mu(D \cap \hat{A}) = \mu(C \cap A)$. Then, $P_A(A) = P_{\hat{A}}(\hat{A}) \geq P_D(\hat{A}) = P_C(A)$. Hence, $P_A(A) = C(A)$. It follows that C is symmetric and that $C([0, \omega]) = F(\omega)$. $\square$

Let $\mathcal{F}$ be the set of doubly star-shaped functions taking $[0, 1]$ to $[0, 1]$. According to Theorem 3.1, the set of symmetric, coherent capacities $\mathcal{C}$ can be put in a one-to-one correspondence with $\mathcal{F}$. The following propositions record some basic properties of doubly star-shaped functions. The proof of proposition 3.1 is straightforward and is omitted.

PROPOSITION 3.1. *If $F \in \mathcal{F}$ then*

(i) $F(1) = 1$,

(ii) $F(\omega) \geq \omega$ for every $\omega \in \Omega$,

(iii) F is strictly increasing on $\{\omega; F(\omega) < 1\}$ and

(iv) F is continuous.

Given a function F and a point $\omega \in (0, 1)$ define

$$\begin{aligned} v^{\omega}(y) &= \left(\frac{1 - F(\omega)}{1 - \omega}\right) y + \left(\frac{F(\omega) - \omega}{1 - \omega}\right), \\ \lambda^{\omega}(y) &= \frac{F(\omega)}{\omega} y \quad \text{and} \\ \xi^{\omega}(y) &= \lambda^{\omega}(y) I_{[0,\omega)}(y) + v^{\omega}(y) I_{[\omega,1]}(y). \end{aligned} \tag{3}$$

When $\omega = 1$ define $v^{\omega}(y) \equiv 1$ and $\lambda^{\omega}(y) = y$.

PROPOSITION 3.2. *Let $F : \Omega \to \Omega$. The following four statements are equivalent:*

(i) $F \in \mathcal{F}$.

(ii) For all $\omega \in (0,1]$ and all $0 \le y \le \omega$, $\lambda^\omega(y) \le F(y) \le \upsilon^\omega(y)$.

(iii) F is a continuous function such that $F(1) = 1$ and is strictly increasing on $\{\omega; F(\omega) < 1\}$. Hence, the derivative exists almost everywhere. Furthermore, for almost all $\omega \in (0,1]$,

$$\frac{1 - F(\omega)}{1-\omega} \le F'(\omega) \le \frac{F(\omega)}{\omega}. \tag{4}$$

(iv) For every $\omega \in \Omega$, $F(\omega) = \sup_{G \in \mathcal{G}} G(\omega)$ where $\mathcal{G}$ is a nonempty set of functions such that each $G \in \mathcal{G}$ satisfies (a) $G : \Omega \to \Omega$, (b) $G(1) = 1$, (c) G is concave and (d) $G(\omega) \le F(\omega)$ for all ω.

REMARK: Statement (ii) shows that λ^x and υ^x provide a λower and υpper bound for F.

PROOF. (i) implies (ii): straightforward.

(ii) implies (iii): straightforward.

(iii) implies (iv): Fix $\omega \in (0,1]$. Suppose there exists $y \in (0,\omega]$ such that $F(y) < \xi^\omega(y)$. Let $y_0 = \inf\{z \in [y,\omega]; F(z) \ge \xi^\omega(z)\}$. Thus, $F(y_0) = \xi^\omega(y_0)$ and $F(x) < \xi^\omega(x)$ for $x \in [y, y_0)$. Hence, $F(\omega)/\omega = (\xi^\omega(y_0) - \xi^\omega(y))/(y_0 - y) < (F(y_0) - F(y))/(y_0 - y)$. From this, together with (4) we derive the following contradiction:

$$\begin{aligned}
\frac{F(\omega)}{\omega} &< \frac{F(y_0) - F(y)}{y_0 - y} = \frac{\int_y^{y_0} F'(s)ds}{y_0 - y} \\
&\le \frac{\int_y^{y_0} F(s)/sds}{y_0 - y} < \frac{\int_y^{y_0} \xi^\omega(s)/sds}{y_0 - y} \\
&= \frac{\int_y^{y_0} F(\omega)/\omega ds}{y_0 - y} = \frac{F(\omega)}{\omega}.
\end{aligned}$$

So we conclude that $F(y) \ge \xi^\omega(y)$ for all $y \in (0,\omega]$. A similar argument shows that $F(y) \ge \xi^\omega(y)$ for all $y \in [\omega, 1]$. Now take $\mathcal{G} = \{\xi^\omega; x \in [0,1]\}$.

(iv) implies (i): For $\omega \in [0,1]$ and $\epsilon > 0$ find $G_\epsilon \in \mathcal{G}$ such that $G(\omega) \ge F(\omega) - \epsilon$. Let

$S_\epsilon(\omega)$ be defined as ξ^ω is defined in (3) except with G_ϵ in place of F. The concavity of G_ϵ implies that $S_\epsilon(y) \le F(y)$ for all y. Note that $S_\epsilon \uparrow \xi^\omega$ as $\epsilon \downarrow 0$. It follows that $\xi^\omega \le F$. This holds for every ω and (i) follows. □

Note that, by Proposition 3.2 (iii), if $F : \Omega \to \Omega$ is in $\mathcal{F}$ then there exists a measurable function α on Ω such that $0 \le \alpha(\omega) \le 1$ almost everywhere and

$$F'(\omega) = \frac{\alpha(\omega)F(\omega)}{\omega} + \frac{(1-\alpha(\omega))(1-F(\omega))}{(1-\omega)} \tag{5}$$

almost everywhere. From this observation, we are motivated to consider the transform from F to the corresponding α.

Define the function α_1 by $\alpha_1(\omega) = 1$ for all $\omega \in \Omega$. Let $\mathcal{A} = \{\alpha_1\} \cup \mathcal{A}_\infty$ where $\mathcal{A}_\infty$ is the set of all measurable functions on Ω such that $0 \le \alpha(\omega) \le 1$ for μ-almost all ω and

$$\lim_{\omega \to 1} \int_{1/2}^{\omega} \frac{\alpha(u) - u}{u(1-u)} du = -\infty. \tag{6}$$

Define the function F_1 by $F_1(\omega) = \omega$. Given a function F define $\alpha = \mathcal{T}(F)$ and $c = \mathcal{U}(F)$ by

$$\alpha(\omega) = \begin{cases} \frac{F'(\omega) - \frac{1-F(\omega)}{1-\omega}}{\frac{F(\omega)}{\omega} - \frac{1-F(\omega)}{1-\omega}} & \text{if } F \ne F_1 \\ 1 & \text{if } F = F_1 \end{cases} \tag{7}$$

and $c = F(1/2) - 1/2$.

LEMMA 3.1. *Let $F \in \mathcal{F}$. If $\alpha = \mathcal{T}(F)$ and $c = \mathcal{U}(F)$ then $\alpha \in \mathcal{A}$ and $0 \le c \le 1/2$. Furthermore, $c = 0$ if and only if $\alpha = \alpha_1$.*

PROOF. If $F = F_1$ then $\alpha = \alpha_1 \in \mathcal{A}$ and $c = 0$. Suppose that $F \ne F_1$. Since $F \in \mathcal{F}$, we have by Proposition 3.2 (iii) that

$$\frac{1-F(\omega)}{1-\omega} \le F'(\omega) \le \frac{F(\omega)}{\omega}$$

for almost all ω. Thus, for some function β such that $0 \leq \beta(\omega) \leq 1$ (almost all ω) we can write

$$F'(\omega) = \beta(\omega)\frac{F(\omega)}{\omega} + (1-\beta(\omega))\frac{1-F(\omega)}{1-\omega}.$$

Solving this equation and comparing it to (7) we see that $\beta(\omega) = \alpha(\omega)$ which shows that $0 \leq \alpha(\omega) \leq 1$ almost everywhere. Now we show that (6) holds. Note that $\alpha(\omega)/(\omega(1-\omega)) = (F'(\omega)-1)/(F(\omega)-\omega)$. Since this last expression is equal to $d\log(F(\omega)-\omega)/d\omega$ it follows that

$$\lim_{\omega\to 1}\int_{1/2}^{\omega}\frac{\alpha(u)-u}{u(1-u)}du = \lim_{\omega\to 1}(\log(F(\omega)-\omega) - \log(F(1/2)-1/2)) = -\infty$$

since $F(1) = 1$ and $F(1/2) > 1/2$. Thus $\alpha \in \mathcal{A}$. The remarks about c follow immediately from the fact that $1/2 \leq F(1/2) \leq 1$ and the fact that $F(1/2) = 1/2$ if and only if $F = F_1$.

$\square$

Given a function α and a real number c define a function $F = \mathcal{R}(\alpha, c)$ by

$$F(\omega) = \omega + c\exp\left\{\int_{1/2}^{\omega}\frac{\alpha(u)-u}{u(1-u)}du\right\}. \tag{8}$$

Also, let

$$c_\alpha = \inf_\omega (1-\omega)\exp\left\{-\int_{1/2}^{\omega}\frac{\alpha(u)-u}{u(1-u)}du\right\}.$$

LEMMA 3.2. *Let $\alpha \in \mathcal{A}$ and $c \in \mathbb{R}$ be given. Let $F = \mathcal{R}(\alpha, c)$. If $0 \leq c \leq c_\alpha$ then $F \in \mathcal{F}$. Furthermore, $\mathcal{T}(F) = \alpha$ and $\mathcal{U}(F) = c$.*

PROOF. If $\alpha = \alpha_1$ then $c = 0$ and the claim follows easily. So assume that $c > 0$. From (6) we deduce that $F(1) = 1$. Clearly $F(\omega) \geq 0$ for ω and F is continuous. That $F(\omega) \leq 1$ for all ω follows from the condition on c. Now differentiate F and use the fact that $0 \le \alpha(\omega) \le 1$ for almost all ω to conclude that

$$\frac{1-F(\omega)}{1-\omega} \le F'(\omega) \le \frac{F(\omega)}{\omega}$$

for almost all ω. Since $F(\omega) \le 1$ for all ω, this implies that F' is non-negative almost everywhere. Furthermore, on the set $\{\omega; F(\omega) < 1\}$, F' is strictly positive wherever the derivative is defined. Thus, F is strictly increasing on this set. By Proposition 3.2 (iii) it follows that $F \in \mathcal{F}$. That $\mathcal{T}(F) = \alpha$ and $\mathcal{U}(F) = c$ follow from direct calculation. □

Finally, we are in a position to characterize of the extreme points of $\mathcal{F}$. Denote the set of extreme points by $\mathcal{E}$.

THEOREM 3.2. (Characterization of Extreme F.) *Suppose that $F \in \mathcal{F}$. Then $F \in \mathcal{E}$ if and only if*

(i) $\mu(\{\omega; 0 < \mathcal{T}(F)(\omega) < 1\}) = 0$ *and*

(ii) $\mathcal{U}(F) \in \{0, c_\alpha\}$.

PROOF. Suppose first that (ii) fails to hold. Choose $0 < \epsilon < \min\{c, c_\alpha - c\}$. Define $c_1 = c+\epsilon$ and $c_2 = c-\epsilon$. Let $F_i = \mathcal{R}(\alpha, c_i)$, $i = 1, 2$. By Lemma 3.2, $F_i \in \mathcal{F}$, $i = 1, 2$. Clearly, F_1, F_2 and F are distinct and $F = (1/2)F_1 + (1/2)F_2$ so F is not extreme. Suppose now that (i) fails to hold. Let A be a set of positive measure such that α_F is strictly between 0 and 1. We may find $\delta > 0$, $\omega_0 \in (0, 1)$ and $a > 0$ such that $\delta < \text{ess inf}\alpha(\omega) \le \text{ess sup}\alpha(\omega) < 1 - \delta$ for $\delta \in I = (\omega_0 - a, \omega_0 + a) \subset (0, 1)$. Fix $\Delta \in I\!R$. Define

$$\alpha_1(\omega) = \alpha(\omega) + \frac{\omega(1-\omega)f'(\omega)}{1+f(\omega)}$$

and

$$\alpha_2(\omega) = \alpha(\omega) - \frac{\omega(1-\omega)f'(\omega)}{1-f(\omega)}$$

where $f(\omega) = \Delta(\omega - \omega_0)$ for $\omega \in I$ and $f(\omega) = 0$ otherwise. Let $F_i = \mathcal{R}(\alpha_i, c)$, $i = 1, 2$. Now

$$F_1(\omega) - F_2(\omega) = 2f(\omega)\exp\left\{\int_{1/2}^{\omega} \frac{\alpha(u) - u}{u(1-u)}\right\}$$

so F_1 and F_2 are distinct. It is easy to see that $F = (1/2)F_1 + (1/2)F_2$. Finally, we claim that for $\Delta > 0$ sufficiently small, $F_1, F_2 \in \mathcal{F}$. For Δ small and from the fact that α is almost surely between δ and $1-\delta$ on I we conclude that α_1 and α_2 are almost surely between 0 and 1. It is easy to verify (6) holds for both α_1 and α_2. Also, for $i = 1, 2$, $c_{\alpha_i} \to c$ as $\Delta \to 0$ so that, for small Δ, $0 \le c \le \min\{c_{\alpha_1}, c_{\alpha_2}\}$. Hence, Lemma 3.2 implies that $F_1, F_2 \in \mathcal{F}$. Hence, F is not extreme.

Now suppose that (i) and (ii) hold. Let $F = pF_1 + (1-p)F_2$ with $p \in (0,1)$. Let $\alpha = \mathcal{T}(F)$ and let $\alpha_i = \mathcal{T}(F_i)$, $i = 1, 2$. Let $A_1 = \{\omega; \alpha(\omega) = 1\}$. Suppose there exists $A \subset A_1$ such that $\mu(A) > 0$ and $\alpha_1(\omega) < 1$ on A. From (5) together with the fact that $F_1'(\omega) \le F_1(\omega)/\omega$ for almost all ω, we conclude that $F_1'(\omega) < F_1(\omega)/\omega$ for almost all $\omega \in A$. Thus, for almost all $\omega \in A$ we have,

$$F'(\omega) = pF_1'(\omega) + (1-p)F_2'(\omega) < pF_1(\omega)/\omega + (1-p)F_2(\omega)/\omega = F(\omega)/\omega.$$

But since $\alpha(\omega) = 1$ on A we have that $F'(\omega) = F(\omega)/\omega$ which is a contradiction. Hence, $\alpha_1(\omega) = 1$ for almost all ω in A_1 and similarly for α_2. By a similar argument, $\alpha_1(\omega) = \alpha_2(\omega) = 0$ for almost all $\omega \in A_0 = \{\omega; \alpha(\omega) = 0\}$. Since $\mu(A_0 \cup A_1) = 1$, $\alpha = \alpha_1 = \alpha_2$ almost everywhere. It follows that $c_{\alpha_1} = c_{\alpha_2} = c_\alpha$. Let $c_i = \mathcal{U}(F_i)$, $i = 1, 2$. Since $c = 0$ or $c = c_\alpha$ it follows that $c = c_1 = c_2$. Thus, $F = F_1 = F_2$. $\square$

Now we consider another characterization of the extreme points. Given $F \in \mathcal{F}$ define

$$\Gamma_0(F) \;=\; \{\omega; \lambda^\omega(y) < F(y) < \upsilon^\omega(y) \;\text{ for all } y < \omega\},$$

$$\begin{aligned}\Gamma_1(F) &= \{\omega; \text{there exists some } y < x \text{ such that } F(z) = \upsilon^\omega(y) \text{ for all } z \in [y, \omega]\} \quad \text{and} \\ \Gamma_2(F) &= \{\omega; \text{there exists some } y < x \text{ such that } F(z) = \lambda^\omega(y) \text{ for all } z \in [y, \omega]\}.\end{aligned}$$

PROPOSITION 3.3. *If $F \in \mathcal{F}$ then $\Gamma_0(F)$, $\Gamma_1(F)$ and $\Gamma_2(F)$ partition* $(0, 1]$.

PROOF. Suppose that $F(y) = \upsilon^\omega(y)$ for some $y < \omega$. Let $a = (y, \upsilon(y))$ and $b = (x, F(x))$. For all $x \in [y, \omega]$, the double star-shaped condition implies that $F(z) \geq \mathcal{L}_{a,b}(z) = \upsilon^\omega(z)$. By Proposition 3.2, part (ii), $F(z) \leq \upsilon^\omega(z)$. Hence, $F(z) = \upsilon^\omega(z)$ and $\omega \in \Gamma_1(F)$. By a similar argument, if $F(y) = \lambda^\omega(y)$ for some $y < \omega$ then it can be shown that $\omega \in \Gamma_2(F)$. If $F(y)$ is never equal to $\upsilon^\omega(y)$ or $\lambda^\omega(y)$ for any $y < z$ then $\omega \in \Gamma_3(F)$. $\square$

COROLLARY 3.1. *F is an extreme point in $\mathcal{F}$ if and only if $\Gamma_0(F) = \emptyset$.*

PROOF. Let $\alpha = \mathcal{T}(F)$. Suppose that $\Gamma_0(F) \neq \emptyset$. Choose $\omega \in \Gamma_0(F)$. Then α is strictly between 0 and 1 in some interval $(\omega - \delta, \omega]$. Hence, by the previous theorem, F is not extreme. Now suppose that F is not extreme. Thus either $0 < c < c_\alpha$ or α strictly between 0 and 1 on a set of positive measure. First suppose that $0 < c < c_\alpha$. It is easy to see that $1 \in \Gamma_0(F)$ so that $\Gamma_0(F) \neq \emptyset$. Now suppose that α strictly between 0 and 1 on a set of positive measure. In particular, this is true for some open interval (a, b). Thus, $b \in \Gamma_0$ so that $\Gamma_0(F) \neq \emptyset$. $\square$

The latter corollary gives a geometric interpretation to the extreme points: they are piecewise linear functions that oscillate between the upper and lower supporting lines υ^ω and λ^ω.

4. The 2-alternating Case. In the previous section we identified the extreme points of $\mathcal{F}$. In this section we identify the extreme points of the subset $\mathcal{F}_2$ of distribution functions corresponding to 2-alternating Choquet capacities. It turns out that $\mathcal{F}_2$ is a simplex so that each $F \in \mathcal{F}_2$ has a unique Choquet representation as a mixture of the extreme points. We identify the mixing distribution explicitly.

For every $t \in (0,1]$ define $F_t(\omega) = \min\{\omega/t, 1\}$. Also, define $F_0(\omega) \equiv 1$. Let $\mathcal{E}_2$ be the extreme points of $\mathcal{F}_2$.

THEOREM 4.1. $\mathcal{E}_2 = \{F_t; 0 \leq t \leq 1\}$.

PROOF. It is easy to establish that each F_t is an extreme point. That this set exhausts all extreme points follows from Theorem 4.2. □

THEOREM 4.2. (Choquet Representation for symmetric, 2-alternating capacities.) *For every $F \in \mathcal{F}_2$ there exists a probability measure Q on $([0,1], \mathcal{B})$ such that, for all $\omega \in \Omega$,*

$$F(\omega) = \int_0^1 F_t(\omega) Q(dt). \tag{9}$$

Furthermore, Q is unique and satisfies $\frac{dQ}{d\nu} = r(\omega)$ where

$$r(\omega) = F(0)I_0 - \omega F''(\omega) I_B(\omega) - \omega(F'(\omega^+) - F'(\omega^-)) I_{B^c}(\omega),$$

$\nu = \mu + \tilde{\mu}$, $\tilde{\mu}$ is counting measure on B^c and B is the set where F'' exists. By convention, we set $F'(1^+) = 0$ and we include 0 and 1 in B^c.

PROOF. Since $F \in \mathcal{F}_2$, it is concave by a straightforward generalization of Theorem 5 of Wasserman and Kadane (1992). Thus F'' exists on a set B such that B^c is countable. Define $G(\omega) = \int_{[0,\omega] \cap B} r(s) \nu(ds)$. Now G is non-negative, non-decreasing and right-continuous and so is a distribution function for a measure Q. We need only show that (9) holds. This will establish that Q is a probability measure since $F(\omega) = 1$.

Let $a(t) = F'(\omega^+) - F'(\omega^-)$. First suppose that B^c is finite. We can thus write $B^c =$

$\{\omega_0, \ldots, \omega_n\}$ where $0 = \omega_0 < \omega_1 < \ldots < \omega_n = 1$. Suppose that $\omega_j < \omega < \omega_{j+1}$, say. Then

$$
\begin{aligned}
\int_0^1 F_t(\omega) Q(dt) &= \int_0^1 F_t(\omega) \frac{dQ}{d\nu}(t)\nu(dt) \\
&= \int_0^1 F_t(\omega) r(t)\mu(dt) + \sum_{t\in B^c} F_t(\omega) r(t) \\
&= -\int_0^1 F_t(\omega) t F''(t)\mu(dt) - \sum_{t\in B^c} F_t(\omega) t a(t) \\
&= -\int_0^\omega t F''(t)\mu(dt) - \omega \int_\omega^1 F''(t)\mu(dt) - \sum_{t\in B^c\cap[0,\omega]} t a(t) - \omega \sum_{t\in B^c\cap(\omega,1]} a(t) \\
&= -\sum_{i=0}^{j-1}(\omega_{i+1}F'(\omega_{i+1}^-) - \omega_i F'(\omega_i^+)) + F(\omega_j) - F(0) - \omega F'(\omega) + \omega_j F'(\omega_j^+) \\
&\quad + F(\omega) - F(\omega_j) - \omega F'(\omega_{j+1}^-) + \omega F'(\omega) - \omega \sum_{i=j+1}^{n-1}(F'(\omega_{i+1}^-) - F'(\omega_i^+)) \\
&\quad - \sum_{t\in B^c\cap[0,\omega]} t a(t) - \omega \sum_{t\in B^c\cap(\omega,1]} a(t) \\
&= F(\omega) - F(0) + \sum_{i=1}^{j} \omega_i (F'(\omega_i^-) - F'(\omega_i^+)) + \omega \sum_{i=j+1}^{n}(F'(\omega_i^-) - F'(\omega_i^+)) \\
&\quad - \sum_{t\in B^c\cap[0,\omega]} t a(t) - \omega \sum_{t\in B^c\cap(\omega,1]} a(t) \\
&= F(\omega)
\end{aligned}
$$

A similar calculation can be used when $\omega \in B^c$.

Now suppose that B^c is denumerable. Fix $\epsilon > 0$. Choose $A = \{\omega_0, \ldots, \omega_n\} \subset B^c$ such that $0 = \omega_0 < \omega_1 < \cdots < \omega_n = 1$, and $-\sum_{\omega\in A} \omega a(\omega) > -\sum_{\omega\in B^c} \omega a(\omega) - \epsilon$. Let $W_1 = [0, \omega_1), W_2 = (\omega_1, \omega_2], \ldots, W_n = (\omega_{n-1}, 1]$. We can find $F_\epsilon \in \mathcal{F}_2$ such that $\sup_{\omega\in\Omega} |F_\epsilon(\omega) - F(\omega)| < \epsilon$, $\sup_{\omega\in B} |F_\epsilon''(\omega) - F''(\omega)| < \epsilon$ and

$$\sum_{\omega\in A}([F'(\omega^+) - F'(\omega^-)] - [F_\epsilon'(\omega^+) - F_\epsilon'(\omega^-)]) < \epsilon.$$

Let Q_ϵ be the representing measure for F_ϵ and let $r_\epsilon = dQ_\epsilon/d\nu$. Now $r_\epsilon \to r$ uniformly in ω as $\epsilon \to 0$. It follows that Q_ϵ converges weakly to Q. Now,

$$F(\omega) - \int_0^1 F_t(\omega) Q(dt) = (F(\omega) - F_\epsilon(\omega)) + (F_\epsilon(\omega) - \int_0^1 F_t(\omega) Q_\epsilon(dt))$$

$$+ \quad (\int_0^1 F_t(\omega)Q_\epsilon(dt) - \int_0^1 F_t(\omega)Q(dt)).$$

The first term tends to 0 as $\epsilon \to 0$ since F_ϵ approximates F. The second term is identically 0. The third term tends to 0 as $\epsilon \to 0$ since F_t is bounded and since Q_ϵ converges weakly to Q. Taking the limit of both sides of the equation as $\epsilon \to 0$ establishes (9).

Finally, we show the uniqueness of Q. Suppose there exists another representing probability R. Note that $R << \nu$. Let $q = dQ/d\nu$ and $r = dR/d\nu$. Since, for all ω, $F(\omega) = \int_0^1 F_t(\omega)Q(dt) = \int_0^1 F_t(\omega)R(dt)$ we conclude that $\int_0^1 (1 - F_t(\omega))Q(dt) = \int_0^1 (1 - F_t(\omega))R(dt)$ which implies that

$$\int_0^\omega \frac{q(t) - r(t)}{1 - t^{-1}} d\nu(t) = 0$$

for all ω. Differentiating with respect to ω we conclude that $q(\omega) = r(\omega)$ for almost all ω. $\Box$

Since $\mathcal{F}_2$ is convex and since each F can be represented as a unique mixture of extreme points, it follows that $\mathcal{F}_2$ is a simplex (Choquet 1969, section 28). The class $\mathcal{F}_2$ might be pictured as a triangle inscribed in a circle, being representable as a unique mixture of the extreme points of $\mathcal{F}_2 \subset \mathcal{F}$ but, of course, representable as many different mixtures of the elements of $\mathcal{E}$. Finally, we remark that 2-alternating is equivalent to

$$\alpha'(\omega) \leq \frac{\alpha(\omega)(1 - \alpha(\omega))}{\omega(1 - \omega)}$$

for almost all ω.

5. Conclusion. In this paper, we gave a characterization of the extreme points of the set of distribution functions for symmetric capacities. With the appropriate transform, these correspond to a subset of the 0-1 functions on the unit interval. In the 2-alternating case, each distribution function has a Choquet representation as a unique mixture of the extreme points. Given the special role that 2-alternating capacities play in robustness theory, it would be interesting to develop a measure of the degree to which a capacity fails to be 2-alternating. An open question is whether the characterizations given in this paper can be used to develop such a measure.

REFERENCES

Anger, Bernd and Lembcke, Jörn (1985). Infinitely subadditive capacities as upper envelopes of measures. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **68**, 403-414.

Armstrong, T. (1990). Comonotonicity, simplical subdivision of cubes and non-linear expected utility via Choquet integrals. Dept. of Mathematics and Statistics, University of Maryland.

Bednarski, Tadeusz (1981). On solutions of minimax test problems for special capacities. *Z. Wahr. verw. Gebiete.*, **58**, 397-405.

Bednarski, Tadeusz (1982). Binary experiments, minimax tests and 2-alternating capacities. *Ann. Statist.*, **10**, 226-232.

Berger, J. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness in Bayesian Statistics.* (J. Kadane ed.): North-Holland, Amsterdam.

Berger, J. (1985). *Statistical Decision Theory.* (2nd Edition). Springer-Verlag: N.Y.

Berger, J. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inf.* **25**, 303-328.

Buja, A. (1986). On the Huber-Strassen theorem. *Probab. Th. Rel. Fields,* **73**, 367-384.

Choquet, G. (1953). Theory of capacities. *Ann. Inst. Fourier. (Grenoble)*, **5**, 131-295.

Choquet, G. (1969). *Lectures on Analysis: Volume II.* The Benjamin/Cummings Publishing Company, Inc. Reading, Massachusetts.

Dempster, A.P. (1967). Upper and lower probabilities induced from a multivalued mapping. *Ann. Math. Statist.* **38** 325-339.

Dempster, A.P. (1968). A generalization of Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **30** 205-247.

Denneberg, D. (1994). *Non-additive Measure and Integral.* Kluwer: Dordrecht.

DeRobertis, Lorraine and Hartigan, J.A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.*, **9**, 235-244.

Fine, T. (1988). Lower probability models for uncertainty and nondeterministic processes. *J. Statist. Planning and Inference.* **20** 389-411.

Fortini, S. and Ruggeri, F. (1994). Concentration functions and Bayesian robustness. *Journal of Statistical Planning and Inference,* vol. 40, n.2 & 3, 205-220.

Gilboa, I. (1987). Expected utility theory with purely subjective non-additive probabilities. *Journal of Mathematical Economics*, **16**, 65-88.

Huber, P.J. (1981). *Robust Statistics.* Wiley: New York.

Huber, P.J. (1973). The use of Choquet capacities in statistics. *Bull. Inst. Internat. Statist.*, **45**, 181-191.

Huber, P.J. and Strassen, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.*, **1**, 251-263.

Lavine, M. (1991a). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.* **86**, 396-399.

Lavine, M. (1991b). An approach to robust Bayesian analysis for multidimensional parameter spaces. *J. Amer. Statist. Assoc.* **86**, 400-403.

Marshall, A.W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press: New York.

Ryff, J. (1963). On the representation of doubly stochastic operators. *Pacific J. Math.* **13**, 1379-1386.

Ryff, J. (1965). Orbits of L^1-functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* **117**, 92-100.

Ryff, J. (1967). On Muirhead's theorem. *Pacific J. Math.* **21**, 567-576.

Ryff, J. (1970). Measure preserving transformations and rearrangements. *J. of Math. Anal. Applic.* **31**, 449-458.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica* **57**, 571-587.

Shapley, L. (1971). Cores of convex games. *International Journal of Game Theory.* **1**, 11-26.

Smith, C.A.B. (1961). Consistency in statistical inference and decision (with discussion). *J. Roy. Statist. Soc. Ser. B.* **23** 1-37.

Talagrand, M. (1978). Capacités invariantes extrémales. *Ann. Inst. Fourier. (Grenoble)*, **28**, 79-146.

Walley, P. (1990). *Statistical Reasoning With Imprecise Probabilities.* Chapman and Hall (Monographs on Statistics and Applied Probability), London.

Wasserman, L.A. and Kadane, J. (1990). Bayes' theorem for Choquet capacities. *Ann. Statist.* **18**, 1328-1339.

Wasserman, L.A. and Kadane, J. (1992). Symmetric Upper Probabilities. *Ann. Statist.* **20**, 1720-1736.

Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213