9-1990

Bayes' Theorem for Choquet Capacities

Larry A. Wasserman
Carnegie Mellon University

Joseph B. Kadane
Carnegie Mellon University, kadane@stat.cmu.edu

Published In
The Annals of Statistics, 18, 3, 1328-1339.

The Annals of Statistics
1990, Vol. 18, No. 3, 1328–1339

BAYES' THEOREM FOR CHOQUET CAPACITIES

BY LARRY A. WASSERMAN[1] AND JOSEPH B. KADANE[2]

Carnegie Mellon University

We give an upper bound for the posterior probability of a measurable set A when the prior lies in a class of probability measures $\mathscr{P}$. The bound is a rational function of two Choquet integrals. If $\mathscr{P}$ is weakly compact and is closed with respect to majorization, then the bound is sharp if and only if the upper prior probability is 2-alternating. The result is used to compute bounds for several sets of priors used in robust Bayesian inference. The result may be regarded as a characterization of 2-alternating Choquet capacities.

1. Introduction. Sets of probability measures arise naturally in classical robustness, Bayesian robustness and group decision making. Some sets of probability measures give rise to upper probabilities that are 2-alternating Choquet capacities. These upper probabilities are pervasive in the robustness literature; see Huber (1973), Huber and Strassen (1973), Buja (1984, 1985, 1986), Bednarski (1982) and Rieder (1977), for example. The purpose of this paper is to prove a version of Bayes' theorem for sets of prior probabilities that have the 2-alternating property and to see how these sets of probabilities may be exploited in Bayesian robustness. This result was proved in Wasserman (1988, 1990) for infinitely alternating capacities (also known as belief functions; Shafer, 1976). Since infinitely alternating capacities are also 2-alternating, the present result generalizes that theorem. However, the proof in the infinitely alternating case uses an argument that depends on properties that infinitely alternating capacities possess that are not shared by 2-alternating capacities in general. Also, in this paper, the conditions given are both necessary and sufficient. A proof of sufficiency when the parameter space is finite is given in Walley (1981).

Section 2 of this paper states and proves the main result. In Section 3, we apply the result to derive explicit bounds for the posterior probability of a measurable set using various classes of priors. Section 4 contains a discussion.

2. Main result. Let Θ be a Polish space, that is, the topology for Θ is complete, separable and metrizable, and let $\mathscr{B}$ be the Borel σ-algebra for Θ. Let $\mathscr{X}$ be the set of all bounded, nonnegative, real-valued, measurable functions defined on Θ. Let $\{P_\theta;\ \theta \in \Theta\}$ be a set of probability measures on a sample space Y with σ-algebra $\mathscr{A}$. Assume that each P_θ has a density $f(y|\theta)$

Received October 1988; revised May 1989.

[1]Research supported by a post-doctoral fellowship from the Natural Sciences and Engineering Research Council of Canada.

[2]Research supported by NSF grant DMS-87-05646.

AMS 1980 *subject classifications*. Primary 62F15; secondary 62F35.

Key words and phrases. 2-alternating Choquet capacities, robust Bayesian inference, posterior bounds.

with respect to some σ-finite dominating measure and let $L(\theta) = f(y|\theta)$ be the likelihood function for θ having observed $y \in Y$. We assume that $L \in \mathscr{X}$ for each $y \in Y$.

Let $\mathscr{P}$ be a nonempty set of prior probability measures on $\mathscr{B}$ and define the upper and lower prior probability functions by

$$\overline{P}(A) = \sup_{P \in \mathscr{P}} P(A) \quad \text{and} \quad \underline{P}(A) = \inf_{P \in \mathscr{P}} P(A).$$

Since $\overline{P}$ and $\underline{P}$ are related by the equation $\overline{P}(A) = 1 - \underline{P}(A^c)$, it suffices to concentrate on $\overline{P}$. In what follows, there is no loss in generality by assuming that $\mathscr{P}$ is convex. For each $X \in \mathscr{X}$, define the upper expectation of X by $\overline{E}(X) = \sup_{P \in \mathscr{P}} P(X)$, where $P(X) = \int X(\theta) P(d\theta)$. The lower expectation of X, denoted by $\underline{E}(X)$, is defined in an analogous way. Let $\mathscr{P}_y$ be the class of posteriors arrived at by applying Bayes' theorem to each prior P in $\mathscr{P}$ for which $P(L) > 0$. To avoid triviality, we assume there is at least one such P. Let $\overline{P}_y$ be the upper probability generated by $\mathscr{P}_y$. The posterior corresponding to P is denoted by P_y.

We say that $\mathscr{P}$ is *2-alternating* if for each $A, B \in \mathscr{B}$, $\overline{P}(A \cup B) \le \overline{P}(A) + \overline{P}(B) - \overline{P}(A \cap B)$. We say that $\mathscr{P}$ generates a *Choquet capacity* if $\overline{P}(F_n) \downarrow \overline{P}(F)$ for each sequence of closed sets $F_n \downarrow F$. It can be shown that $\mathscr{P}$ generates a Choquet capacity if and only if $\{P;\ P \le \overline{P}\}$ is weakly compact. Here, $P \le \overline{P}$ means that $P(A) \le \overline{P}(A)$ for every $A \in \mathscr{B}$. See Choquet (1953) and Huber and Strassen (1973) for details on capacities. We say that $\mathscr{P}$ is *closed with respect to majorization*, or is *m-closed*, if $P \le \overline{P}$ implies that $P \in \mathscr{P}$. It is common practice in the literature to assume that $\mathscr{P}$ is m-closed. We will show (Section 3, Example 1) that replacing $\mathscr{P}$ with its m-closure has nontrivial consequences.

For each $X \in \mathscr{X}$, the *upper Choquet integral* of X is defined by

$$\overline{C}(X) = \int_0^\infty \overline{P}(X_t)\, dt,$$

where $X_t \equiv \{\theta \in \Theta;\ X(\theta) > t\}$. Similarly, the *lower Choquet integral* $\underline{C}(X)$ is defined by

$$\underline{C}(X) = \int_0^\infty \underline{P}(X_t)\, dt.$$

It follows easily that $\underline{C}(X) \le \underline{E}(X) \le \overline{E}(X) \le \overline{C}(X)$. If $\mathscr{P}$ consists of a single probability measure P, then $\overline{C}(X) = \underline{C}(X) = \overline{E}(X) = \underline{E}(X) = P(X)$; if this probability measure is our prior probability, then the posterior probability of a subset A, after observing y and applying Bayes' theorem, may be expressed as $P(L_A)/(P(L_A) + P(L_{A^c}))$, where $L_A(\theta) = L(\theta) I_A(\theta)$ and $I_A(\theta)$ is the indicator function for A. Note that $\overline{C}$ and $\underline{C}$ may be extended to the set of all bounded, measurable functions by adding and subtracting a sufficiently large positive constant.

THEOREM. *Let $\mathscr{P}$ be a nonempty set of prior probabilities on $\mathscr{B}$ and let $\mathscr{P}_y$ be the corresponding class of posterior probabilities. Then, for each $A \in \mathscr{B}$,*

$$\overline{P}_y(A) \leq \frac{\overline{E}(L_A)}{\overline{E}(L_A) + \underline{E}(L_{A^c})} \leq \frac{\overline{C}(L_A)}{\overline{C}(L_A) + \underline{C}(L_{A^c})},$$

when the ratios are well defined. If $\mathscr{P}$ generates a Choquet capacity and if $\mathscr{P}$ is m-closed, then the following three statements are equivalent:

(i) *$\overline{P}$ is 2-alternating.*
(ii) *The first inequality is an equality for each $A \in \mathscr{B}$ and each $L \in \mathscr{X}$.*
(iii) *The second inequality is an equality for each $A \in \mathscr{B}$ and each $L \in \mathscr{X}$.*

Before proving the theorem, we shall state a few lemmas. The first lemma is from Huber and Strassen (1973).

LEMMA 1. *If $\mathscr{P}$ is m-closed and generates a 2-alternating Choquet capacity, then for each upper semicontinuous $X \in \mathscr{X}$, there exists a $P \in \mathscr{P}$ such that $P(X) = \overline{E}(X)$ and $\overline{P}(X_t) = P(X_t)$, for each real number t.*

The next two lemmas are from Bednarski (1982).

LEMMA 2. *If Θ is finite, then the following two conditions are equivalent:*

(i) *$\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B) - \overline{P}(A \cap B)$ for each $A, B \subset \Theta$.*
(ii) *For every sequence of sets $A_1 \subset A_2 \subset \cdots \subset A_n$, there exists $P \leq \overline{P}$ such that $P(A_i) = \overline{P}(A_i)$ for $i = 1, 2, \ldots, n$.*

LEMMA 3. *Let $\mathscr{P}$ be weakly compact and suppose that $\mathscr{P}$ is m-closed. Then $\mathscr{P}$ is 2-alternating if and only if, for each finite field $\mathscr{A}$, $\overline{\mathscr{P}|\mathscr{A}}$ is 2-alternating where $\mathscr{P}|\mathscr{A}$ is the restriction of $\mathscr{P}$ to $\mathscr{A}$ and $\overline{M}$ is the closure of M with respect to pointwise convergence on atoms.*

The last two lemmas may be found in Buja (1984).

LEMMA 4. *Let $\mathscr{P}$ generate a 2-alternating Choquet capacity. For each $X \in \mathscr{X}$ and each $\varepsilon > 0$, there exists a nonnegative, upper semicontinuous function $h \leq X$ such that $\overline{C}(X) - \varepsilon < \overline{C}(h) \leq \overline{C}(X)$. Furthermore, if $\mathscr{P}$ is m-closed, then for each $X \in \mathscr{X}$, $\overline{E}(X) = \overline{C}(X)$.*

LEMMA 5. *$\mathscr{P}$ is 2-alternating if and only if $\overline{C}(X + Y) \leq \overline{C}(X) + \overline{C}(Y)$ for each $X, Y \in \mathscr{X}$.*

PROOF OF THEOREM. Note that for every $P \in \mathscr{P}$ for which $P(L) > 0$,

$$P_y(A) = \frac{P(L_A)}{P(L_A) + P(L_{A^c})}.$$

Now, $P_y(A)$ is of the form

$$h(P) = \frac{1}{(1 + g(P)/f(P))},$$

where $f(P) = P(L_A)$ and $g(P) = P(L_{A^c})$. $h(P)$ is maximized where $g(P)/f(P)$ is minimized. But

$$\frac{g(P)}{f(P)} \geq \frac{\underline{E}(L_{A^c})}{\overline{E}(L_A)},$$

which proves the first inequality. The second inequality follows from that fact that $\overline{E}(X) \leq \overline{C}(X)$ and $\underline{E}(X) \geq \underline{C}(X)$.

Now suppose $\mathscr{P}$ is m-closed and generates a Choquet capacity. We begin by showing that (i) implies (ii). Assume that $\overline{P}$ is 2-alternating. We need to show that the first inequality is an equality for each $A \in \mathscr{B}$ and for each $L \in \mathscr{L}$. If $\overline{E}(L_A) = 0$, this follows easily. Now suppose $\overline{E}(L_A) > 0$. Let $b = \sup_{\theta \in A^c} L(\theta)$. Define a function α on Θ by $\alpha(\theta) = L(\theta) + b$ on A and $\alpha(\theta) = -L(\theta) + b$ on A^c. For notational convenience, define

$$B = \frac{\overline{E}(L_A)}{\overline{E}(L_A) + \underline{E}(L_{A^c})}.$$

We will show that for each $\varepsilon > 0$, there exists $P \leq \overline{P}$ such that

$$B - \varepsilon < \frac{P(L_A)}{P(L_A) + P(L_{A^c})} \leq B.$$

For every $\delta > 0$, there exists, by Lemma 4, a nonnegative, bounded, upper semicontinuous function h such that $h \leq \alpha$ and $\overline{E}(\alpha) - \delta \leq \overline{E}(h) \leq \overline{E}(\alpha)$. Set $k = \overline{E}(L_A) + \underline{E}(L_{A^c})$. Note that $k > 0$. Choose $\delta < \varepsilon k$. By Lemma 1, there exists $P \in \mathscr{P}$ such that $P(h) = \overline{E}(h)$ and $P(h > t) = \overline{P}(h > t)$, for each t. It is easily verified that $P(L_A) \geq \overline{E}(L_A) - \delta$ and $P(L_{A^c}) \leq \underline{E}(L_{A^c}) + \delta$. Thus,

$$\begin{aligned} P_y(A) &= \frac{P(L_A)}{P(L_A) + P(L_{A^c})} \geq \frac{\overline{E}(L_A) - \delta}{\overline{E}(L_A) - \delta + \underline{E}(L_{A^c}) + \delta} \\ &= B - \frac{\delta}{\overline{E}(L_A) + \underline{E}(L_{A^c})} = B - \frac{\delta}{k} > B - \varepsilon. \end{aligned}$$

Since this holds for each $\varepsilon > 0$, $\sup_{P \in \mathscr{P}} P_y(A) = B$ so that (ii) holds. Further, (i) implies (iii) because of Lemma 4.

To prove that (ii) implies (i), first assume that Θ is finite. Suppose that $\overline{P}_y(A) = \overline{E}(L_A)/(\overline{E}(L_A) + \underline{E}(L_{A^c}))$ for every subset A and every $L \in \mathscr{L}$. $\mathscr{P}$ is closed, with respect to the Euclidean topology, so there exists, for each A, a $P \in \mathscr{P}$ such that $P_y(A) = \overline{P}_y(A)$. Therefore,

$$\frac{\overline{E}(L_A)}{\overline{E}(L_A) + \underline{E}(L_{A^c})} = \frac{P(L_A)}{P(L_A) + P(L_{A^c})},$$

from which it follows that

$$(1) \qquad P(L_A) = \overline{E}(L_A) \quad \text{and} \quad P(L_{A^c}) = \underline{E}(L_{A^c}).$$

This holds for each A and each $L \in \mathscr{X}$. Let $A_1 \subset A_2 \subset \cdots \subset A_n$ be an increasing sequence of subsets. For $i = 1, 2, \ldots, n-1$, define L_i on Θ to be equal to 1 on A_i and A_{i+1}^c and 0 otherwise. From (1) we deduce that there exists $P_i \in \mathscr{P}$ such that $P_i(A_i) = \overline{P}(A_i)$ and $P_i(A_{i+1}) = \overline{P}(A_{i+1})$ for $i = 1, 2, \ldots, n-1$. Define $A_0 = \varnothing$ and let $g_i = A_i - A_{i-1}$. Let $\mathscr{G}$ be the field generated by $\{g_1, \ldots, g_n\}$. Define an additive set function S on $\mathscr{G}$ by

$$\begin{aligned} S(g_1) &= P_1(A_1), \\ S(g_2) &= P_2(A_2) - P_1(A_1), \\ &\vdots \\ S(g_{n-1}) &= P_{n-1}(A_{n-1}) - P_{n-2}(A_{n-2}), \\ S(g_n) &= P_{n-1}(A_n) - P_{n-1}(A_{n-1}) \end{aligned}$$

and $S(A_n^c) = 1 - S(A_n)$. Then, $S(A_i) = \overline{P}(A_i)$ for $i = 1, 2, \ldots, n$. Also, $S(g_1) \geq 0$ and $S(g_i) = P_i(A_i) - P_{i-1}(A_{i-1}) = \overline{P}(A_i) - P_{i-1}(A_{i-1}) = P_{i-1}(A_i) - P_{i-1}(A_{i-1}) \geq 0$, so that S is a probability measure on $\mathscr{G}$ which maximizes the probability of the A_i's, and since S is dominated by $\overline{P}$ on $\mathscr{G}$, S may be extended by the Hahn–Banach theorem to 2^{Θ}. By Lemma 2, $\overline{P}$ satisfies the 2-alternating condition. Therefore $\mathscr{P}$ is 2-alternating.

Now we show that (iii) implies (i) and we continue with the assumption that Θ is finite. Let the second inequality be an equality for each $A \in \mathscr{B}$ and each $L \in \mathscr{X}$. Then $\overline{E}(X) = \overline{C}(X)$ for each bounded function X. For suppose $\overline{E}(X) < \overline{C}(X)$ for some X not identically equal to zero. Let $a = \min\{X(\theta); \theta \in \Theta\}$ and set $A = \{\theta, X(\theta) > a\}$. We claim that $\overline{E}(X_A) < \overline{C}(X_A)$. To see this, note that if $a = 0$, then $\overline{E}(X_A) = \overline{E}(X) < \overline{C}(X) = \overline{C}(X_A)$. If $a > 0$, then by the definition of $\overline{C}$, $\overline{C}(X_A) = \overline{C}(X) - a\underline{P}(A^c) > \overline{E}(X_A) + \Delta$, where $\Delta = \overline{E}(X) - \overline{E}(X_A) - a\underline{P}(A^c)$. Now, for each $P \in \mathscr{P}$,

$$P(X) = P(X_A) + P(X_{A^c}) = P(X_A) + aP(A^c) \geq P(X_A) + a\underline{P}(A^c),$$

so that $\overline{E}(X) \geq \overline{E}(X_A) + a\underline{P}(A^c)$. Hence, $\Delta \geq 0$ which implies that $\overline{E}(X_A) < \overline{C}(X_A)$. But this, together with the fact that $\underline{E}(L_{A^c}) = a\underline{P}(A^c) = \underline{C}(L_{A^c})$ leads to the second inequality being strict. This contradicts our assumption. Therefore, $\overline{E}(X) = \overline{C}(X)$. It follows that $\overline{C}$ is subadditive and from Lemma 5 we deduce that $\overline{P}$ is 2-alternating.

Now remove the restriction that Θ is finite. Suppose that either (ii) or (iii) holds. Let $\mathscr{A}$ be a finite subfield and without loss of generality, assume $\mathscr{A}$ is generated by a finite partition $H = \{h_1, h_2, \ldots, h_k\}$. It follows easily that (ii) and (iii) hold on $\overline{\mathscr{P}|\mathscr{A}}$. By the previous argument, if either (ii) or (iii) holds, then $\overline{\mathscr{P}|\mathscr{A}}$ is 2-alternating. By Lemma 3, $\mathscr{P}$ is 2-alternating. □

We close this section with a few remarks. First, note that if $\mathscr{P}$ is not m-closed, we still have that (ii) implies (i) and also (iii) implies (i), since

m-closure was not used in those parts of the proof. Second, if $\mathscr{P}$ is not convex, the theorem is still true since replacing $\mathscr{P}$ with its convex closure does not change the upper and lower probabilities or the upper and lower expectations. Third, it may be shown that if the prior is 2-alternating, then the posterior is 2-alternating. This is proved for finite Θ in Walley (1981). The proof is easily extended to the case where Θ is a Polish space. However, if the set of priors $\mathscr{P}$ is m-closed, it does not follow in general that the set of posteriors is m-closed. Consider the following example.

Let $\Theta = \{1, 2, 3, 4\}$ and let $\mathscr{P}$ be the convex hull of P and Q which have coordinates $(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{2})$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0)$, respectively. Set $B = \{1, 2, 3\}$. Now $\mathscr{P}$ is m-closed but the set of probabilities $\mathscr{P}_B$ obtained by conditioning each $P \in \mathscr{P}$ is not. For, $R \leq \overline{P}_B$ but $R \notin \mathscr{P}_B$, where R has coordinates $(0.3, 0.4, 0.3, 0)$ and $\overline{P}_B$ is the upper probability generated by $\mathscr{P}_B$. Thus, m-closure is not preserved by conditioning and hence is not, in general, preserved by Bayes' theorem. An interesting problem is to determine when an m-closed class of priors remains m-closed under the action of Bayes' theorem.

The implication of the example above is that the Choquet integral of a bounded random variable X, based on the posterior upper probability, will not, in general, equal the upper posterior expectation of X. Nonetheless, the Choquet integral still provides a conservative bound for the upper posterior expectation of X. Thus, denoting the Choquet integral of X with respect to the upper posterior probability by $\overline{C}_y$, we have that

$$\overline{E}(X|Y=y) \leq \overline{C}_y(X) = \int_0^\infty \overline{P}_y(X_t)\,dt = \int_0^\infty \frac{dt}{1 + g(t)/f(t)},$$

where

$$f(t) = \int_0^\infty \overline{P}(X_t \cap L_u)\,du \quad \text{and} \quad g(t) = \int_0^\infty \overline{P}(X_t^c \cap L_u)\,du.$$

Of course, $\underline{C}_y(X)$ may be calculated in a similar way. Note that if $\overline{C}_y(X) - \underline{C}_y(X)$ is small, then it is unnecessary to compute $\overline{E}(X|Y=y)$.

3. Examples.

EXAMPLE 1. A set that is not m-closed. To see that closure with respect to majorization is important, here is an example of a set of probabilities that generates a 2-alternating upper probability but that is not m-closed and for which the equalities in the theorem arė strict. Let $\Theta = \{1, 2, 3\}$, let the vector of likelihood values be $(1, 6, 2)$ and let $\mathscr{P}$ be the set of probability measures whose coordinates on Θ are given by $(p/2, p/2, (1-p))$, where $p \in [0, 1]$. The upper probability is 2-alternating but $\mathscr{P}$ is not m-closed. For example, if P has coordinates $(0.2, 0.3, 0.5)$, then $P \leq \overline{P}$ but P is not in $\mathscr{P}$. Let $A = \{1\}$. Then

$$\overline{P}_y(A) = \frac{1}{7} < \frac{\overline{E}(L_A)}{\overline{E}(L_A) + \underline{E}(L_{A^c})} = \frac{1}{5} < \frac{\overline{C}(L_A)}{\overline{C}(L_A) + \underline{C}(L_{A^c})} = \frac{1}{3}.$$

This shows that if $\mathscr{P}$ is not m-closed, then the 2-alternating condition is not strong enough to imply the result. Note that enlarging $\mathscr{P}$ to be m-closed has a nontrivial effect on the upper posterior probability.

EXAMPLE 2. A set that is not 2-alternating. Now we give an example of a set of probabilities that is m-closed but is not 2-alternating and we show that the inequalities are strict. A similar example was considered in Wasserman (1990). Let $\Theta = \{1,2,3,4\}$ and let the vector of likelihood values on Θ be (a,b,c,d). Suppose that $a > b > c > d > 0$ and $b > c + d$. Now let P be a probability measure with values $(\frac{1}{2},\frac{1}{2},0,0)$ on Θ and let Q have values $(0,0,\frac{1}{2},\frac{1}{2})$. Finally, let $\mathscr{P}$ be the convex hull of P and Q. It is easily verified that $\mathscr{P}$ is m-closed and that $\overline{P}$ is not 2-alternating. Let $A = \{1\}$. Then

$$\overline{P}_y(A) = \frac{a}{a+b} < \frac{\overline{E}(L_A)}{\overline{E}(L_A) + \underline{E}(L_{A^c})}$$

$$= \frac{a}{a+c+d} < \frac{\overline{C}(L_A)}{\overline{C}(L_A) + \underline{C}(L_{A^c})} = \frac{a}{a+c},$$

so that the inequalities in the theorem are strict.

The following are sets of probability measures that arise naturally in the study of Bayesian robustness. For a general overview of Bayesian robustness, see Berger (1984, 1987).

EXAMPLE 3. ε-Contaminated priors. Let P be a fixed probability measure on $\mathscr{B}$, let $\varepsilon \in [0,1]$ be fixed and assume that Θ is compact. Define $\mathscr{P}$ by

$$\mathscr{P} = \{Q; Q = (1-\varepsilon)P + \varepsilon R, \text{ where } R \text{ is any probability measure on } \mathscr{B}\}.$$

This is the class of ε-contaminated priors that have been studied by Huber (1973), Berger (1985) and Berger and Berliner (1986). It follows easily that $\overline{P}(A) = (1-\varepsilon)P(A) + \varepsilon$, $A \neq \varnothing$ and $\underline{P}(A) = (1-\varepsilon)P(A)$ for $A \neq \Theta$. $\mathscr{P}$ is m-closed and generates a 2-alternating Choquet capacity. [Buja (1986), shows that the compactness assumption may be dropped.] For every $X \in \mathscr{X}$,

$$\overline{C}(X) = (1-\varepsilon)P(X) + \varepsilon \sup_{\theta\in\Theta} X(\theta)$$

and

$$\underline{C}(X) = (1-\varepsilon)P(X) + \varepsilon \inf_{\theta\in\Theta} X(\theta).$$

Applying the theorem, we have

$$\overline{P}_y(A) = \frac{(1-\varepsilon)P_y(A) + \varepsilon a/P(L)}{(1-\varepsilon) + \varepsilon a/P(L)},$$

where $a = \sup_{\theta\in A} L(\theta)$. This result was stated in Huber (1973).

EXAMPLE 4. Total variation neighbourhoods. As in Example 3, fix P and ε and let Θ be compact. Define $\mathscr{P}$ by

$$\mathscr{P} = \{Q; \Delta(P,Q) \le \varepsilon\},$$

where $\Delta(P,Q) = \sup_{A \in \mathscr{B}} |P(A) - Q(A)|$. This is the ε total-variation class generated by P. Let $x \vee y$ denote the maximum of x and y and $x \wedge y$ the minimum of x and y. Then $\overline{P}(A) = (P(A) + \varepsilon) \wedge 1$, $A \ne \varnothing$, and $\underline{P}(A) = (P(A) - \varepsilon) \vee 0$, $A \ne \Theta$. Again, $\mathscr{P}$ is m-closed and generates a 2-alternating Choquet capacity. For every $X \in \mathscr{X}$, define $t_0(X) = \sup\{t; P(X_t) \ge 1 - \varepsilon\}$ and $s_0(X) = \inf\{t; P(X_t) \le \varepsilon\}$. Also, let

$$J(X, t_1, t_2) = \int_{t_1}^{t_2} P(X_t)\, dt.$$

Then

$$\overline{C}(X) = P(X) + \varepsilon \sup X + (1 - \varepsilon) t_0(X) - J(X, 0, t_0(X))$$

and

$$\underline{C}(X) = P(X) + \varepsilon \inf X - \varepsilon s_0(X) - J(X, s_0(X), \sup X).$$

Let $t_0 = t_0(L_A)$ and $s_0 = s_0(L_{A^c})$. Then,

$$\overline{P}_y(A) = \frac{P(L_A) + (1-\varepsilon)t_0 + \varepsilon a - J(L_A, 0, t_0)}{P(L) + (1-\varepsilon)t_0 + \varepsilon a - J(L_A, 0, t_0) - \varepsilon s_0 - J(L_{A^c}, s_0, b)},$$

where $a = \sup_{\theta \in \Theta} L_A(\theta)$ and $b = \sup_{\theta \in \Theta} L_{A^c}(\theta)$.

Bednarski (1981) and Buja (1986) considered an upper probability that contains ε-contamination and total variation neighbourhoods as special cases. It may be shown that this generalized class is an m-closed, 2-alternating Choquet capacity as well.

EXAMPLE 5. Density bounded classes. Let L and U be two σ-finite measures on $\mathscr{B}$ such that $L \le U$; that is, $L(A) \le U(A)$ for each $A \in \mathscr{B}$. Lavine (1987) considers classes of probability measures of the form $\mathscr{P} = \{P; L \le P \le U\}$. He calls this a *density bounded class*. We shall assume that $L(\Theta) \le 1 \le U(\Theta) < \infty$. Without loss of generality, we also assume that L and U have densities l and u with respect to some σ-finite dominating measure ν. Then $P \in \mathscr{P}$ if and only if P has density p with respect to ν and for ν-almost all θ, $l(\theta) \le p(\theta) \le u(\theta)$. DeRobertis and Hartigan (1981) consider the class DH of all σ-finite measures between L and U. The class DH is also discussed in Berger (1987) and DasGupta and Studden (1988a, 1988b).

For the class $\mathscr{P}$, it follows that $\overline{P}(A) = \min\{U(A), 1 - L(A^c)\}$ and $\underline{P}(A) = \max\{L(A), 1 - U(A^c)\}$. Further, it may be shown that $\mathscr{P}$ generates a 2-alternating, m-closed Choquet capacity. For $X \in \mathscr{X}$, set $t_0(X) = \inf\{t; U(X_t) + L(X_t^c) \le 1\}$ and $s_0(X) = \sup\{t; U(X_t^c) + L(X_t) \le 1\}$. Let $J(X_t, c, d) = \int_c^d D(X_t)\, dt$, where $D = U - L$. Finally, let $d_L = 1 - L(\Theta)$ and $d_U = U(\Theta) - 1$. Then, $\overline{C}(X) = U(X) + t_0(X)\, d_L - J(X, 0, t_0(X))$ and $C(X) =$

$L(X) - s_0(X)\, d_U + J(X, 0, s_0(X))$. Here, $U(X) = \int X(\theta)U(d\theta)$ and $\overline{L}(X) = \int X(\theta)L(d\theta)$. Therefore,

$$\overline{P}_y(A) = \frac{U(L_A) + t_0 d_L - J(L_A, 0, t_0)}{U(L_A) + t_0 d_L - J(L_A, 0, t_0) + L(L_{A^c}) - s_0\, d_U + J(L_{A^c}, 0, s_0)},$$

where $t_0 \equiv t_0(L_A)$ and $s_0 \equiv s_0(L_{A^c})$.

EXAMPLE 6. Probabilities with partition restrictions. Moreno and Cano (1988) consider the following set of priors. Let P be a fixed probability measure on $\mathscr{B}$, ε a fixed number in $[0, 1]$ and $H = \{h_1, \dots, h_k\}$ a measurable partition of Θ. Assume that Θ is compact. Let $p = (p_1, \dots, p_k)$ be such that each $p_i \geq 0$ and $\sum p_i = 1$. Then define

$$\mathscr{P} = \{Q; Q = (1 - \varepsilon)P + \varepsilon S, S \in \mathscr{S}\}.$$

Here $\mathscr{S}$ is the set of all probability measures that satisfy $S(h_i) = p_i$, $i = 1, 2, \dots, k$. The interpretation is the same as the ε-contaminated class of Example 3, except that we are putting constraints on the possible contaminations by forcing them to have fixed probabilities on some partition. This is a simple, reasonable way to restrict the size of the possible contaminating distributions.

This class contains two important special cases. By taking $H = \Theta$, we get back the general ε-contaminated model of Example 3. By taking $\varepsilon = 1$, we get the set of all probability measures with fixed probabilities on a partition. This latter class has been studied by Berliner and Goel (1986).

Define $I_A = \{i; h_i \subset A\}$ and $J_A = \{i; h_i \cap A \neq \varnothing\}$. Then

$$\underline{P}(A) = (1 - \varepsilon)P(A) + \varepsilon \sum_{i \in I_A} p_i$$

and

$$\overline{P}(A) = (1 - \varepsilon)P(A) + \varepsilon \sum_{i \in J_A} p_i.$$

It may be shown that $\overline{P}$ is n-alternating for every n and hence is 2-alternating. Further, $\mathscr{P}$ generates an m-closed, Choquet capacity. It follows that

$$\underline{C}(X) = (1 - \varepsilon)P(X) + \varepsilon \sum \underline{X}_i p_i$$

and

$$\overline{C}(X) = (1 - \varepsilon)P(X) + \varepsilon \sum \overline{X}_i p_i,$$

where $\underline{X}_i = \inf_{\theta \in h_i} X(\theta)$ and $\overline{X}_i = \sup_{\theta \in h_i} X(\theta)$. Therefore,

$$\overline{P}_y(A) = \frac{(1 - \varepsilon)P(L_A) + \varepsilon \sum p_i \overline{L}_{A_i}}{(1 - \varepsilon)P(L) + \varepsilon \sum p_i \left[\overline{L}_{A_i} + \underline{L}_{A_i^c}\right]}.$$

EXAMPLE 7. Symmetric, unimodal contaminations. Sivaganesan and Berger (1989) consider the following modification of the ε-contaminated class for Θ

equal to the real line. For fixed P and ε, set

$$\mathscr{P} = \{Q = (1 - \varepsilon)P + \varepsilon S, S \in \mathscr{S}\},$$

where $\mathscr{S}$ is the set of all probability measures that possess densities that are unimodal and symmetric about a fixed point θ_0. Without loss of generality, we take $\theta_0 = 0$. Now, $\mathscr{P}$ generates a 2-alternating upper probability if and only if $\mathscr{S}$ generates a 2-alternating upper probability. Hence, we confine attention to the class $\mathscr{S}$. As Sivaganesan and Berger (1989) remark, when computing extrema, one need only consider the smaller class $\mathscr{S}'$ consisting of uniform probability measures that are centered at the origin. A point mass at the origin must be included as well. Let $\overline{S}$ be the upper probability generated by $\mathscr{S}$. We claim that $\overline{S}$ is not 2-alternating and that $\mathscr{S}$ is not m-closed.

To see that $\overline{S}$ is not 2-alternating, set $A = [-3, -1] \cup [1, 2]$ and $B = [-2, -1] \cup [1, 3]$. Then, $\frac{2}{3} = \overline{S}(A \cup B) > \overline{S}(A) + \overline{S}(B) - \overline{S}(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{2}$. To see that $\mathscr{S}$ is not m-closed, let $R = \frac{1}{2}\delta_0 + \frac{1}{2}u$, where δ_0 is a point mass at 0 and u is the uniform $[0, 1]$ distribution. Now, if $0 \in A$, $R(A) \le 1 = \overline{S}(A)$. Otherwise, $R(A) = R(A \cap [0, 1]) = u(A \cap [0, 1])/2 = u'(A \cap [0, 1]) \le u'(A) \le \overline{S}(A)$, where u' is uniform $[-1, 1]$. Thus, $R \le \overline{S}$ but $R \notin \mathscr{S}$, so $\mathscr{S}$ is not m-closed.

As a result, the Choquet integral fails miserably as an approximation to $\overline{E}$. For example, let X_n be defined by

$$X_n(\theta) = \begin{cases} \theta, & \text{if } -n \le \theta \le n, \\ n, & \text{if } \theta > n, \\ -n, & \text{if } \theta < -n. \end{cases}$$

Then $\overline{E}(X_n) = \underline{E}(X_n) = 0$, but it is easily shown that $\overline{C}(X_n) = n/2$. Hence, $\lim_{n \to \infty}(\overline{C}(X_n) - \overline{E}(X_n)) = \infty$. This is an extreme example that shows how important the 2-alternating and m-closure conditions are. It would be interesting to investigate how much of the difference between $\overline{E}$ and $\overline{C}$ is accounted for by the failure of the 2-alternating condition and how much is because of the lack of m-closure.

4. Discussion. The main result of this paper helps to unify the process of finding bounds for posterior probabilities in robust Bayesian inference. Also, this theorem may be thought of as a characterization of 2-alternating Choquet capacities. Other authors have given different characterizations. Huber and Strassen (1973) showed that the 2-alternating structure is necessary and sufficient for generalizing the Neyman–Pearson lemma to sets of probabilities. Buja (1984, 1985) generalized this result from the case of two hypotheses to a finite number of hypotheses. Similarly, Bednarski (1982) showed that this condition is necessary and sufficient for the construction of least informative binary experiments in the sense of Blackwell (1951), Le Cam (1964, 1969, 1972) and Torgersen (1970). These authors were concerned with uncertainty with respect to sampling distributions, as represented by sets of probability measures dominated by 2-alternating Choquet capacities. We have focused on uncertainty in the prior distribution. Other properties of 2-alternating capacities may be found in Walley (1981).

Acknowledgments. We are grateful to the referees for their insightful comments and suggestions.

REFERENCES

BEDNARSKI, T. (1981). On solutions of minimax test problems for special capacities. *Z. Wahrsch. Verw. Gebiete* **58** 397–405.

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

BERGER, J. O. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness in Bayesian Statistics* (J. Kadane, ed.) 63–144. North-Holland, Amsterdam.

BERGER, J. O. (1985). *Statistical Decision Theory*, 2nd ed. Springer, New York.

BERGER, J. O. (1987). Robust Bayesian analysis: sensitivity to the prior. Technical Report No. 87-10. Dept. Statist., Purdue Univ.

BERGER, J. O. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

BERLINER, L. M. and GOEL, P. K. (1986). Incorporating partial prior information: ranges of posterior probabilities. Technical Report No. 357, Dept. Statist., Ohio State Univ.

BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Probab.* Univ. California, Berkeley.

BUJA, A. (1984). Simultaneously least favourable experiments. I: Upper standard functionals and sufficiency. *Z. Wahrsch. Verw. Gebiete* **65** 367–384.

BUJA, A. (1985). Simultaneously least favourable experiments. II: Upper standard loss functions and their applications. *Z. Wahrsch. Verw. Gebiete* **69** 387–420.

BUJA, A. (1986). On the Huber–Strassen theorem. *Probab. Theory Related Fields* **73** 367–384.

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

DASGUPTA, A. and STUDDEN, W. J. (1988a). Robust Bayesian analysis and optimal designs in normal linear models with many parameters. I. Technical Report No. 88-14, Dept. Statist., Purdue Univ.

DASGUPTA, A. and STUDDEN, W. J. (1988b). Robust Bayesian analysis and optimal designs in normal linear models with many parameters. II. Technical Report No. 88-34C, Dept. Statist., Purdue Univ.

DEROBERTIS, L. and HARTIGAN, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9** 235–244.

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. L. (1987). Prior influence in Bayesian statistics. Ph.D. thesis, Univ. Minnesota.

LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.

LE CAM, L. (1969). *Théorié Asymptotique de la Décision Statistique*. Univ. Montreal Press, Montreal.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

MORENO, E. and CANO, J. A. (1988). Robust Bayesian analysis with ε-contaminations partially known. Report, Univ. de Granada, Spain.

RIEDER, H. (1977). Least favourable pairs for special capacities. *Ann. Statist.* **5** 909–921.

SHAFER, G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, N.J.

SIVAGANESAN, S. and BERGER, J. O. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17** 868–889.

TORGERSEN, E. N. (1970). Comparison of experiments when the parameter space is finite. *Z. Wahrsch. Verw. Gebiete* **16** 219–249.

WALLEY (1981). Coherent lower (and upper) probabilities. Statistics research report, Univ. Warwick, Coventry.

WASSERMAN, L. A. (1988). Some applications of belief functions to statistical inference. Ph.D. thesis, Univ. Toronto.

WASSERMAN, L. A. (1990). Prior envelopes based on belief functions. *Ann. Statist.* **18** 454–464.

DEPARTMENT OF STATISTICS
232 BAKER HALL
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213-3890