The Annals of Statistics
1982, Vol. 10, No. 1, 226–232

BINARY EXPERIMENTS, MINIMAX TESTS AND 2-ALTERNATING CAPACITIES[1]

BY TADEUSZ BEDNARSKI

Mathematical Institute of the Polish Academy of Sciences, Wrocław

The concept of Choquet's 2-alternating capacity is explored from the viewpoint of Le Cam's experiment theory. It is shown that there always exists a least informative binary experiment for two sets of probability measures generated by 2-alternating capacities. This result easily implies the Neyman-Pearson lemma for capacities. Moreover, its proof gives a new method of construction of minimax tests for problems in which hypotheses are generated by 2-alternating capacities. It is also proved that the existence of least informative binary experiments is sufficient for a set of probability measures to be generated by a 2-alternating capacity. This gives a new characterization of 2-alternating capacities, closely related to that of Huber and Strassen.

1. Introduction. Robust test problems between two approximately known probability measures P and Q are usually formalized as minimax test problems between neighborhoods of P and Q. It was proved by Strassen (1964, 1965) for finite spaces and then by Huber and Strassen (1973) for Polish spaces that the Neyman-Pearson lemma generalizes to Choquet's 2-alternating capacities. This result in particular implies that if neighborhoods of P and Q can be described in terms of 2-alternating capacities, then the minimax tests have a simple Neyman-Pearson structure (see Huber, 1969 and Huber and Strassen, 1973). Fortunately, all neighborhoods used to formalize inaccuracies in specification of underlying distributions can be described by 2-alternating capacities.

Let Ω be a Polish space and let $\mathscr{B}$ stand for the Borel σ-field on Ω. By $\mathscr{M}$ we denote the set of all probability measures on $\mathscr{B}$. The concept of alternating capacity of order n was introduced by G. Choquet (1953, 1959) and resulted from problems of potential theory. The 2-alternating capacity used by Huber and Strassen (1973) can be defined as a set function v from $\mathscr{B}$ to $[0, 1]$ which is the upper probability of a weakly compact set of probability measures, and it satisfies the condition $v(A \cup B) + v(A \cap B) \le v(A) + v(B)$ for all $A, B \in \mathscr{B}$. A set $\mathscr{P}$ of all probability measures majorized by v, i.e. $\mathscr{P} = \{P \in \mathscr{M} : P(A) \le v(A), \text{ for all } A \in \mathscr{B}\}$, is said to be generated by v.

In this paper we examine connections between 2-alternating capacities and binary experiments.

In Section 2 we state basic facts concerning finite experiments and 2-alternating capacities. Also, some preliminary results are proved. The existence of least informative binary experiments in $\mathscr{P}_0 \times \mathscr{P}_1$, where $\mathscr{P}_i$ are subsets of $\mathscr{M}$ generated by 2-alternating capacities, is proved in Section 3. This result is closely related to the main result of Huber and Strassen (1973) in the sense that each one of the results can be easily derived from the other. The proof we give here is based on quite elementary considerations and it implies a general method of construction of minimax solutions for hypotheses generated by 2-alternating capacities. In Section 4 we prove a necessity of 2-alternating capacities, namely, if $\mathscr{P} \subset \mathscr{M}$ is convex and weakly compact and for every probability measure $Q \in \mathscr{M}$ and every finite field $\mathscr{A} \subset \mathscr{B}$ there exists a least informative binary experiment in the closure of $\mathscr{P} \times \{Q\} \mid \mathscr{A}$, then $\mathscr{P}$ has to be generated by a 2-alternating capacity.

Received November 1977; revised August 1980

[1] This research was partly supported by NIH Grant No. GM-10525, National Institutes of Health, Public Health Service and National Science Foundation Grants GP 31091X2 and MPS74-18967.

AMS 1970 *subject classification*. Primary 62G35; secondary 62B15.

Key words and phrases. Robust testing, minimax testing, capacities, binary experiments.

2. Notation and basic facts. Notions and facts concerning the finite statistical experiment can be found in Blackwell (1951), Le Cam (1964, 1969, 1972) and Torgersen (1970). Below we recall some of them.

Let Θ be a finite set. An experiment E indexed by Θ is defined by a σ-field $\mathscr{B}$ and a map $\theta \to P_\theta$ from Θ to the set of all probability measures on $\mathscr{B}$. When Θ is a two-element set then experiments indexed by Θ are called binary experiments. Let $E = (P_0, P_1)$ and $F = (Q_0, Q_1)$ be binary experiments. The deficiency $\delta(E, F)$ of E with respect to F can be evaluated by the following formula proved by Torgersen (1970) (see also Le Cam, 1977).

Define f by $fd(P_0 + P_1) = dP_0$. For each $\alpha \in [0, 1]$, let $m_\alpha(u)$ be the function defined on $[0, 1]$ by $m_\alpha(u) = \min[\alpha u, (1 - \alpha)(1 - u)]$ and let $T_\alpha(P_0, P_1) = \int m_\alpha(f)d(P_0 + P_1)$. The formula says that

$$\delta(E, F) = \sup\{T_\alpha(P_0, P_1) - T_\alpha(Q_0, Q_1) : \alpha \in [0, 1]\}. \tag{2.1}$$

The integral $T_\alpha(P_0, P_1)$ is the Bayes risk for the problem of testing P_0 against P_1 with 0, 1 loss function and with prior probabilities α, $1 - \alpha$, i.e.

$$T_\alpha(P_0, P_1) = \inf\{\alpha P_0(A) + (1 - \alpha)P_1(A^c) : A \in \mathscr{B}\}, \tag{2.2}$$

where $\mathscr{B}$ is a σ-field on which P_0, P_1 are defined.

If $\delta(E, F) = 0$ we say that E is more informative or better than F. The distance $\Delta(E, F)$ between two experiments E and F is defined as $\max\{\delta(E, F), \delta(F, E)\}$. If $\Delta(E, F) = 0$ we say that the experiments are equivalent. The equivalence class of an experiment E is called the type of E. For a given set Θ, the experiment types induced by Θ form a set which is a compact metric space.

The testing affinity of the pair (P_0, P_1) corresponds to an average Bayes risk for testing P_0 against P_1 and is defined by

$$a(P_0, P_1) = (\tfrac{1}{2}) \int (1 - f) fd(P_0 + P_1). \tag{2.3}$$

Let Ω be a separable, complete and metrizable space and let $\mathscr{B}$ stand for its Borel σ-field. Let $\mathscr{M}$ be the set of all probability measures on $\mathscr{B}$. Huber and Strassen (1973) give the following definition.

A set function v from $\mathscr{B}$ to $[0, 1]$ is called a 2-alternating capacity if the following conditions are satisfied:

(a) for every $A \subset B$, $0 = v(\varnothing) \le v(A) \le v(B) \le v(\Omega) = 1$,
(b) for every sequence of measurable sets $A_n \nearrow A$ we have $v(A_n) \nearrow v(A)$,
(c) for every sequence of closed sets $F_n \searrow F$ we have $v(F_n) \searrow v(F)$,
(d) for every $A, B \in \mathscr{B}$ we have $v(A \cup B) + v(A \cap B) \le v(A) + v(B)$.

We shall say that a family $\mathscr{P} \subset \mathscr{M}$ is generated by a 2-alternating capacity v if $\mathscr{P}$ is the largest set of probability measures majorized by v, i.e. $\mathscr{P} = \{P \in \mathscr{M} : P(A) \le v(A)$ for all $A \in \mathscr{B}\}$.

Below we give some important relations between $\mathscr{P}$ and v. The proofs can be found in Huber and Strassen (1973) and in Choquet (1953, 1959).

(2.4) v is regular, i.e. for every $A \in \mathscr{B}$

$$v(A) = \sup\{v(K) : K \text{ compact}, A \supset K\} = \inf\{v(G) : G \text{ open}, A \subset G\}.$$

(2.5) The set $\mathscr{P}$ generated by v is convex and weakly compact.
(2.6) For every $A \in \mathscr{B}$ there is $P \in \mathscr{P}$ such that $P(A) = v(A)$.
(2.7) For every weakly compact set $\mathscr{P}_0 \subset \mathscr{M}$, its upper probability satisfies (a), (b) and (c).
(2.8) For every monotone sequence of closed sets $F_1 \subset F_2 \subset \cdots \subset F_n$, there exists $P \in \mathscr{P}$ such that $P(F_i) = v(F_i)$ for $i = 1, \cdots, n$.

The above properties imply that the 2-alternating capacity can be equivalently defined as the upper probability of a weakly compact set of probability measures that satisfies condition (d).

Let $\mathscr{A}$ be a finite field of some subsets of Ω and let v be a monotone set function from $\mathscr{A}$ to [0, 1], such that $v(\varnothing) = 0$ and $v(\Omega) = 1$. The 2-alternating property of v is characterized by the following lemma proved by Le Cam (unpublished manuscript).

LEMMA 2.1. *Under the above assumptions, the following two conditions are equivalent:*

(i) *for every $A, B \in \mathscr{A}$ we have $v(A \cup B) + v(A \cap B) \le v(A) + v(B)$,*
(ii) *for every sequence of $\mathscr{A}$ measurable sets $A_1 \subset A_2 \subset \cdots \subset A_n$, there exists a probability measure P on $\mathscr{A}$ such that $P(A_i) = v(A_i)$ for $i = 1, \cdots, n$ and $P(A) \le v(A)$ for every A in $\mathscr{A}$.*

PROOF. Every finite field is the Borel σ-field of a Polish space. Therefore, by (2.8), (ii) is a consequence of (i). For the converse, it is enough to take P majorized by v and such that $P(A \cap B) = v(A \cap B)$, $P(A \cup B) = v(A \cup B)$. □

LEMMA 2.2. *Let $\mathscr{P}$ be a closed and convex set of probability measures defined on a finite field $\mathscr{A}$. Then $\mathscr{P}$ is generated by a 2-alternating capacity* iff *for every sequence of $\mathscr{A}$ measurable sets $A_1 \subset A_2 \subset \cdots \subset A_n$ there is $P_0 \in \mathscr{P}$ such that $P_0(A_i) = \sup\{P(A_i) : P \in \mathscr{P}\}$.*

PROOF. Lemma 2.1 guarantees the existence of P_0 if $\mathscr{P}$ is generated by a 2-alternating capacity. For the converse, let us first notice that Lemma 2.1 implies that the upper probability $u(A) = \sup\{P(A) : P \in \mathscr{P}\}$ is a 2-alternating capacity. By contradiction, assume that there is $P_0 \notin \mathscr{P}$ such that $P_0(A) \le u(A)$ for all $A \in \mathscr{A}$. Since $\mathscr{P}$ is convex and closed, there exists a hyperplane separating P_0 and $\mathscr{P}$. Thus there is a positive $\mathscr{A}$-measurable function h such that $\sup\{\int h dP : P \in \mathscr{P}\} < \int h dP_0$. Since $\int h dP = \int_0^\infty P(h > t)\, dt$ for every P on $\mathscr{A}$, Lemma 2.1 yields a contradiction. □

Let $\mathscr{P}_v$ denote the subset of $\mathscr{M}$ generated by a 2-alternating capacity v, and let $\mathscr{A}$ be a finite subfield of $\mathscr{B}$. The restriction of $\mathscr{P}$ and v to the field $\mathscr{A}$ is further denoted by $\mathscr{P} | \mathscr{A}$ and $v | \mathscr{A}$, respectively. The subset of $\mathscr{M} | \mathscr{A}$ generalized by $v | \mathscr{A}$ will be denoted by $\mathscr{P}_{v|\mathscr{A}}$. Relations between $\mathscr{P}_v$, $\mathscr{P}_v | \mathscr{A}$ and $\mathscr{P}_{v|\mathscr{A}}$ are summarized by the following lemma.

LEMMA 2.3. *Assume $\mathscr{P} \subset \mathscr{M}$ is convex and weakly compact. Then $\mathscr{P}$ is generated by a 2-alternating capacity* iff *for every finite subfield $\mathscr{A} \subset \mathscr{B}$, the set $\overline{\mathscr{P} | \mathscr{A}}$ (closure of $\mathscr{P} | \mathscr{A}$ for pointwise convergence on atoms) is generated by a 2-alternating capacity.*

PROOF. Suppose $\mathscr{P} \subset \mathscr{M}$ is generated by a 2-alternating capacity v. Lemma 2.2 says that, for every sequence $A_1 \subset A_2 \subset \cdots \subset A_n$ of $\mathscr{A}$ measurable sets, there is $P \in \mathscr{P}_{v|\mathscr{A}}$ such that $P(A_i) = v | \mathscr{A}(A_i)$ for $i = 1, \cdots, n$. The same lemma implies that the convex set generated by such probability measures is $\mathscr{P}_{v|\mathscr{A}}$. By the regularity of v and by (2.8), the convex set $\mathscr{P} | \mathscr{A}$ has "approximately" the same property in the sense that for every $\varepsilon > 0$ there is $P \in \mathscr{P} | \mathscr{A}$ such that $P(A_i) + \varepsilon \ge v(A_i)$ for $i = 1, \cdots, n$. Hence we have $\overline{\mathscr{P} | \mathscr{A}} = \mathscr{P}_{v|\mathscr{A}}$.

Now let u be the upper probability of $\mathscr{P}$. Lemma 2.1 and (2.7) imply that u is a 2-alternating capacity on $\mathscr{B}$. It remains to show that $\mathscr{P}_u \subset \mathscr{P}$. Let then $P \in \mathscr{P}_u$. Consider a base $\mathscr{G} = \{G_1, G_2, \cdots\}$ of open sets which is closed under finite intersections, and denote by $\mathscr{A}_n$ the field generated by $\{G_1, \cdots, G_n\}$. Our assumptions about $\mathscr{P}$ imply that $\overline{\mathscr{P} | \mathscr{A}_n} = P_{u|\mathscr{A}_n}$ and consequently that $P | \mathscr{A}_n \in \overline{\mathscr{P} | \mathscr{A}_n}$. Hence for every $\varepsilon > 0$ there exists $P_{\varepsilon,n} \in \mathscr{P}$ so that $| P_{\varepsilon,n}(G_i) - P(G_i) | \le \varepsilon$ for $i = 1, \cdots, n$. Taking $P_n = P_{\varepsilon_n,n}$ for a sequence $\{\varepsilon_n\}$ of positive numbers converging to 0, we get by Theorem 2.2 of Billingsley (1968) that P is a weak limit of the sequence $\{P_n\}$. Since $\mathscr{P}$ is weakly compact, we obtain $P \in \mathscr{P}$. □

3. Existence of least informative binary experiments in case of 2-alternating capacities. First we shall consider the case of finite $\Omega = \{\omega_1, \cdots, \omega_n\}$. Let $\mathscr{M}$ stand for the set of all probability measures on Ω. For $P, Q \in \mathscr{M}$ the testing affinity $a(P, Q)$ can be written as

$$\frac{1}{2}\sum_{i=1}^{n} f_i(1 - f_i)(P + Q)(\omega_i), \qquad \text{where } f_i = Q(\omega_i)/(P + Q)(\omega_i).$$

THEOREM 3.1. *If $\mathscr{P}_0 \subset \mathscr{M}$ and $\mathscr{P}_1 \subset \mathscr{M}$ are generated by 2-alternating capacities, then there exists a least informative experiment in $\mathscr{P}_0 \times \mathscr{P}_1$.*

PROOF. Suppose first that for all $P \in \mathscr{P}_j$, $j = 0, 1$ we have $P(\omega) > 0$ for every ω in Ω. Following Huber's (1969) argument, we select a pair $(P, Q) \in \mathscr{P}_0 \times \mathscr{P}_1$ such that

$$a(P, Q) = \sup\{a(Z, U) : (Z, U) \in \mathscr{P}_0 \times \mathscr{P}_1\}$$

and then by differentiation of $a[(1 - \beta)P + \beta Z, Q]$ with respect to β we obtain that the density $f = dQ/d(P + Q)$ is stochastically largest for P in $\mathscr{P}_0$. Similarly, f is stochastically smallest for Q in $\mathscr{P}_1$. Since $f = 1/[1 + dP/dQ]$, we also obtain

$$T_\alpha(P, Q) = \alpha v_0\{\alpha P < (1 - \alpha)Q\} + (1 - \alpha)v_1\{\alpha P \geq (1 - \alpha)Q\}$$

for all $\alpha \in [0, 1]$, where v_0, v_1 are 2-alternating capacities generating $\mathscr{P}_0$ and $\mathscr{P}_1$ respectively. Hence $T_\alpha(P, Q) \geq T_\alpha(Z, U)$ for every $(Z, U) \in \mathscr{P}_0 \times \mathscr{P}_1$, and by (2.1) the pair (P, Q) is least informative.

For arbitrary $\mathscr{P}_0$ and $\mathscr{P}_1$, for $j = 0, 1$, we define subsets $\mathscr{P}_j^m$ of $\mathscr{M}$ by $\mathscr{P}_j^m = (1 - \beta_m)\mathscr{P}_j + \beta_m K$, where the sequence $\{\beta_m\} \subset (0, 1)$ converges to 0 and K is a probability measure defined by $K(\omega_i) = 1/n$ for $i = 1, \cdots, n$. For every $m \geq 1$, the sets $\mathscr{P}_0^m$ and $\mathscr{P}_1^m$ are generated by 2-alternating capacities, and they satisfy the assumptions from the first part of the proof. Thus for every m there exists a least informative experiment (P_m, Q_m) in $\mathscr{P}_0^m \times \mathscr{P}_1^m$. For a subsequence $\{m'\}$ of $\{m\}$, the measures $P_{m'}$, $Q_{m'}$ converge on subsets of Ω to some probability measures P and Q respectively. It is easily seen that $(P, Q) \in \mathscr{P}_0 \times \mathscr{P}_1$ and it forms there the least informative binary experiment. □

Let Ω, $\mathscr{B}$, $\mathscr{M}$ denote, respectively, a Polish space, its Borel σ-field and the set of all probability measures on $\mathscr{B}$.

THEOREM 3.2. *Let $\mathscr{P}_0 \subset \mathscr{M}$ and $\mathscr{P}_1 \subset \mathscr{M}$ be generated by 2-alternating capacities v_0 and v_1. Then there exists a least informative binary experiment in $\mathscr{P}_0 \times \mathscr{P}_1$.*

PROOF. For every $(Z, U) \in \mathscr{P}_0 \times \mathscr{P}_1$ and for all $\alpha \in [0, 1]$ by (2.2) we have

$$M_\alpha := \inf\{\alpha v_0(A) + (1 - \alpha)v_1(A^c) : A \in \mathscr{B}\} \geq T_\alpha(Z, U).$$

Because of (2.1), it will be sufficient to find a pair $(P, Q) \in \mathscr{P}_0 \times \mathscr{P}_1$ for which $M_\alpha = T_\alpha(P, Q)$ for every $\alpha \in [0, 1]$.

Let $\mathscr{G} = \{G_1, G_2, \cdots\}$ be a base of open sets in Ω which is closed under finite unions, and let for every $n \geq 1$, $\mathscr{A}_n$ denote the field generated by $\{G_1, \cdots, G_n\}$. Moreover, we assume that for every $\varepsilon > 0$, there is a compact set K such that $K^c \in \mathscr{G}$ and $v_i(K^c) \leq \varepsilon$ for $i = 0, 1$. This assumption can be made here since v_0 and v_1 are, by (2.5) and (2.6), the upper probabilities of weakly compact sets.

Lemma 2.3 and Theorem 3.1 imply that for every n there exists a least informative binary experiment (P_n, Q_n) in $\overline{\mathscr{P}_0 \mid \mathscr{A}_n} \times \overline{\mathscr{P}_1 \mid \mathscr{A}_n}$. Let $\{A_1, \cdots, A_{i(n)}\}$ be the set of all atoms of $\mathscr{A}_n$ and take a sequence of points $\omega_1, \cdots, \omega_{i(n)} \in \Omega$ such that $\omega_j \in A_j$ for $j = 1, \cdots, i(n)$. The probability measures P_n^*, Q_n^* defined by $P_n^*(\omega_j) = P_n(A_j)$, $Q_n^*(\omega_j) = Q_n(A_j)$ form on $\mathscr{B}$ an experiment equivalent to (P_n, Q_n) on $\mathscr{A}_n$.

As we shall see, every cluster point (in the topology of weak convergence) of the sequences $\{P_n^*\}$ and $\{Q_n^*\}$ forms a least informative binary experiment in $\mathscr{P}_0 \times \mathscr{P}_1$.

The definition of $\mathscr{G}$ ensures tightness of the sequences $\{P_n^*\}$ and $\{Q_n^*\}$. Let P and Q be, respectively, weak limits of $P_{n'}^*$ and $Q_{n'}^*$. Pormanteau's Lemma yields $P(G) \le \liminf_{n'} P_{n'}(G) \le v_0(G)$ for every $G \in \mathscr{G}$. Since $\mathscr{G}$ is closed under formation of finite unions, the regularity of P and v_0 implies $P(A) \le v_0(A)$ for every $A \in \mathscr{B}$. Thus $P \in \mathscr{P}_0$ and in a similar way we infer that $Q \in \mathscr{P}_1$.

Since, for every n, $\mathscr{A}_n \subset \mathscr{A}_{n+1}$ we have $T_\alpha(P_n, Q_n) \ge T_\alpha(P_{n+1}, Q_{n+1})$. Thus the compactness of all binary experiments in Δ and (2.1) imply the convergence of (P_n, Q_n) (defined on $\mathscr{A}_n$) and hence of (P_n^*, Q_n^*) (defined on $\mathscr{B}$) to a binary experiment. According to Proposition 6 of Le Cam (1972), (P, Q) forms an experiment which is less informative or equivalent to the one formed by the limit of (P_n^*, Q_n^*) in Δ. Therefore (2.1) and (2.2) imply that $M_\alpha \le \lim_n T_\alpha(P_n^*, Q_n^*) \le T_\alpha(P, Q)$ for every $\alpha \in [0, 1]$. On the other hand, we also have $(P, Q) \in \mathscr{P}_0 \times \mathscr{P}_1$. Thus $M_\alpha = T_\alpha(P, Q)$ for every $\alpha \in [0, 1]$. □

The following corollary explains the relation between least informative binary experiments and Huber's least favourable pairs of distributions.

Corollary 3.1. *Every least informative binary experiment (P, Q) in $\mathscr{P}_0 \times \mathscr{P}_1$ forms Huber's least favourable pair of distributions, i.e. there is a version $q = dQ/dP$ such that $P(q > t) = v_0(q > t)$ and $Q(q \le t) = v_1(q \le t)$, for all $t \ge 0$. Also, each least favourable pair forms a least informative binary experiment in $\mathscr{P}_0 \times \mathscr{P}_1$.*

Proof. For each $\alpha \in (0, 1)$, let t be defined by $\alpha = t/(1 + t)$. Lemma 3.2 of Huber and Strassen (1973) says that there exists a decreasing family of measurable sets A_s, $s \ge 0$, such that

$$A_t = \cup_{s>t} A_s \quad \text{and} \quad M_\alpha = \alpha v_0(A_t) + (1 - \alpha) v_1(A_t^c).$$

Since (P, Q) is least informative, we also have $\mathrm{T}_\alpha(P, Q) = M_\alpha$ for all $\alpha \in [0, 1]$. This in turn implies that the function q defined by $q(x) = \inf\{t : x \notin A_t\}$ satisfies, for all $\alpha \in [0, 1]$, the equality

$$T_\alpha(P, Q) = \alpha P(q > t) + (1 - \alpha) Q(q \le t)$$

and, consequently, q is a version of dQ/dP.

On the other hand, if a pair (P, Q) is least favourable in the sense of Huber, by (2.2) we then have

$$T_\alpha(P, Q) = \alpha v_0(A_t) + (1 - \alpha) v_1(A_t^c) = M_\alpha$$

and by (2.1), (P, Q) forms the least informative binary experiment in $\mathscr{P}_0 \times \mathscr{P}_1$. □

Let $\mathscr{G} = \{G_1, G_2, \cdots\}$ be a base of open sets in Ω. Assume that $\mathscr{G}$ is closed under the formation of finite unions and that for every $\varepsilon > 0$ there exists a compact set K such that $K^c \in \mathscr{G}$ and $v_i(K^c) \le \varepsilon$ for $i = 0, 1$. Let $\mathscr{A}_n$ be the field generated by $\{G_1, \cdots, G_n\}$ and let (P_n, Q_n) be a least informative experiment in $\overline{\mathscr{P}_0 \mid \mathscr{A}_n} \times \overline{\mathscr{P}_1 \mid \mathscr{A}_n}$. The final part of the proof of the theorem gives the following.

Corollary 3.2. *The experiments (P_n, Q_n) defined on $\mathscr{A}_n$ converge in Δ to the least informative binary experiment (P, Q) in $\mathscr{P}_0 \times \mathscr{P}_1$.* □

The corollary says that every least informative experiment (P_n, Q_n) on $\mathscr{A}_n$ has "approximately" the same properties, in the sense of risk, as any least informative experiment (P, Q) in $\mathscr{P}_0 \times \mathscr{P}_1$. The next theorem shows that also the significance level and minimum power of the minimax test for testing $\mathscr{P}_0 \mid \mathscr{A}_n$ against $\mathscr{P}_1 \mid \mathscr{A}_n$ approximates the significance level and minimum power of the minimax test in $\mathscr{P}_0$ against $\mathscr{P}_1$. In this sense, (P_n, Q_n) forms an approximate solution of the general minimax problem $\mathscr{P}_0$ against $\mathscr{P}_1$.

For every $n \ge 1$, let $q_n = dQ_n/dP_n$ be a version such that $P_n(q_n > t) = v_0(q_n > t)$ and

$Q_n(q_n \le t) = v_1(q_n \le t)$. Let $q_0 = dQ/dP$ be defined in the same way. We shall consider a sequence of tests $\varphi_{t,n}$ given by

$$\varphi_{t,n} = \begin{cases} 1 & \text{if } q_n > t \\ \gamma & \text{if } q_n = t \\ 0 & \text{if } q_n < t, \quad \text{for } t \ge 0 \text{ and } n = 0, 1, 2, \cdots. \end{cases}$$

THEOREM 3.3. *Let α and β be the significance level and the minimum power of the test $\varphi_{t,0}$ for the test problem $\mathscr{P}_0$ against $\mathscr{P}_1$. Then, for every continuity point t of the distributions of q under P and Q, we have* $\lim_{n\to\infty} E_{P_n}\varphi_{t,n} = \alpha$ *and* $\lim_{n\to\infty} E_{Q_n}(1 - \varphi_{t,n}) = \beta$.

PROOF. The convergence of the experiments (P_n, Q_n) to the experiment (P, Q) implies the weak convergence of the canonical measures for the experiments (Le Cam, 1972). This and formula (2.1) easily imply that the distributions of q_n under P_n and Q_n converge weakly to the distributions of q under P and Q respectively. □

4. Necessity of 2-alternating capacities. As before, we shall begin with the finite case. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ be a three-element set and let $\mathscr{M}$ stand for the set of all probability measures on Ω. The following lemma easily implies the remaining results of this section.

LEMMA 4.1. *Let $\mathscr{P} \subset \mathscr{M}$ be convex and closed. Assume that, for every $Q \in \mathscr{M}$, there exists $P_Q \in \mathscr{P}$ such that the experiment (P_Q, Q) is least informative in $\mathscr{P} \times \{Q\}$. Then $\mathscr{P}$ is generated by a 2-alternating capacity.*

PROOF. The sequence of sets $B_1 = \{\omega_1\}$, $B_2 = \{\omega_1, \omega_2\}$, $B_3 = \Omega$ defines a probability measure Q by $Q(B_i) = \sup\{P(B_i) : P \in \mathscr{P}\}$, for $i = 1, 2, 3$. We shall prove that $Q \in \mathscr{P}$. The arbitrariness of the sequence $B_1 \subset B_2 \subset B_3$ and Lemma 2.2 will imply that $\mathscr{P}$ is generated by the 2-alternating capacity $v(\cdot) = \sup\{P(\cdot) : P \in \mathscr{P}\}$.

Let (P_Q, Q) form a least informative experiment in $\mathscr{P} \times \{Q\}$. Suppose by contradiction that $P_Q \ne Q$. (If $\mathscr{P}$ is generated by a 2-alternating capacity, then obviously $P_q = Q$.) This implies that there exists a number $\alpha \in (0, 1)$ such that the set $A_\alpha = \{\omega : \alpha P_Q(\omega) < (1 - \alpha)Q(\omega)\}$ has the following two properties: (i) $P_Q(A_\alpha) < Q(A_\alpha)$, and (ii) $[P_Q + Q] \cdot (\{\omega : \alpha P_Q(\omega) = (1 - \alpha)Q(\omega)\}) = 0$. If A_α is a singleton, say $A_\alpha = \{\omega_0\}$, then we take $P \in \mathscr{P}$ such that $P(\omega_0) = Q(\omega_0)$. It is not difficult to see that the definition of Q and assumptions about $\mathscr{P}$ ensure that for every $\omega \in \Omega$ there exists $P \in \mathscr{P}$ so that $P(\omega) = Q(\omega)$. Since $\mathscr{P}$ is convex, $\beta P_Q + (1 - \beta)P \in \mathscr{P}$ for every $\beta \in [0, 1]$ and thus for β sufficiently close to 1 we have $A_\alpha = \{\omega : \alpha[\beta P_Q + (1 - \beta)P](\omega) < (1 - \alpha)Q(\omega)\}$. This in turn implies that $T_\alpha(P_Q, Q) < T_\alpha(\beta P_Q + (1 - \beta)P, Q)$, which by (2.1) contradicts the assumption that (P_Q, Q) is least informative.

If A_α^c is a singleton, then the same argument as before with $P(A_\alpha^c) = Q(A_\alpha^c)$, $P \in \mathscr{P}$ gives the contradiction. Hence $P_Q = Q$ and $Q \in \mathscr{P}$. □

Assume now that $\Omega = \{\omega_1, \cdots, \omega_n\}$ is a finite set. The symbols $\mathscr{B}$, $\mathscr{M}$ have the usual meaning. As a consequence of the last lemma we have the following.

COROLLARY 4.1. *Let $\mathscr{P} \subset \mathscr{M}$ be convex and closed. If for every $Q \in \mathscr{M}$ and every three-atom subfield $\mathscr{A}$ of $\mathscr{B}$ there exists $P_Q \in \mathscr{P}$ such that the experiment $(P_Q, Q) \mid \mathscr{A}$ is least informative in $\mathscr{P} \mid \mathscr{A} \times \{Q\} \mid \mathscr{A}$, then the set function defined as $v(\cdot) = \sup\{P(\cdot) : P \in \mathscr{P}\}$ is a 2-alternating capacity.* □

THEOREM 4.1. *Let $\mathscr{P} \subset \mathscr{M}$ be closed and convex. Assume that for every $Q \in \mathscr{M}$ and every three-atom field $\mathscr{A} \subset \mathscr{B}$, there exist least informative experiments in $\mathscr{P} \mid \mathscr{A} \times \{Q\} \mid \mathscr{A}$ and in $\mathscr{P} \times \{Q\}$. Then $\mathscr{P}$ is generated by a 2-alternating capacity.*

PROOF. By Corollary 4.1 the set function $v(\cdot) = \sup\{P(\cdot) : P \in \mathscr{P}\}$ is a 2-alternating capacity. In order to prove that $\mathscr{P}$ is generated by v, it is, by Lemma 2.2, sufficient to show that for every monotone sequence of sets $B_1 \subset B_2 \subset \cdots \subset B_n$ there is a measure $P \in \mathscr{P}$ such that $P(B_i) = v(B_i)$. Thus let $B_i = \{\omega_1, \cdots, \omega_i\}$ for $i = 1, \cdots, n$. As before, the measure Q is defined by the conditions $Q(B_i) = v(B_i)$ for $i = 1, \cdots, n$. Let (P_Q, Q) form a least informative binary experiment in $\mathscr{P} \times \{Q\}$. Arguing as in Lemma 4.1, we assume by contradiction that $P_Q \neq Q$. This implies that there exists $\alpha \in (0, 1)$ such that the set $A_\alpha = \{\omega : \alpha P_Q(\omega) < (1 - \alpha)Q(\omega)\}$ satisfies conditions (i) and (ii). Since v is a 2-alternating capacity, Lemma 2.1 implies $Q(A_\alpha) \leq v(A_\alpha)$. Moreover, since v is the upper probability of $\mathscr{P}$, there is $P \in \mathscr{P}$ such that $P(A_\alpha) > P_Q(A_\alpha)$. For $\beta \in (0, 1)$ and β sufficiently close to 1 we obtain $T_\alpha[\beta P_Q + (1 - \beta)P, Q] > T_\alpha(P_Q, Q)$. Thus (P_Q, Q) cannot be least informative in $\mathscr{P} \times \{Q\}$. This completes the argument. □

Let Ω be a Polish space, $\mathscr{B}$ its Borel σ-field, and let $\mathscr{M}$ stand for the set of all probability measures on $\mathscr{B}$. The main result of this section, stated below, is a consequence of Lemma 2.3 and Theorem 4.1.

THEOREM 4.2. *Let $\mathscr{P} \subset \mathscr{M}$ be convex and weakly compact. If for every $Q \in \overline{\mathscr{M}}$ and finite subfield $\mathscr{A} \subset \mathscr{B}$ there exists a least informative binary experiment in $\overline{\mathscr{P}|\mathscr{A}} \times \{Q\} \mid \mathscr{A}$ then $\mathscr{P}$ is generated by a 2-alternating capacity.* □

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BLACKWELL, D. (1951). Comparison of experiments. *Proc. Sec. Berkeley Symp. on Math. Statist. and Probab.* Univ. of Cal. Press, Berkeley and Los Angeles.

CHOQUET, G. (1953, 1954). Theory of capacities. *Ann. Inst. Fourier* **5** 131–292.

CHOQUET, G. (1959). Forme abstraite du théorème de capacitabilité. *Ann. Inst. Fourier* **9** 83–89.

HUBER, P.J. (1969). Théorié de l'nférence statistique robuste. Lecture notes, Université de Montréal.

HUBER, P.J. and STRASSEN, V. (1973). Minimax tests and the Neyman-Pearson Lemma for capacities. *Ann. Statist.* **1** 251–263.

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

LE CAM, L. (1969). *Théorie Asymptotique de la Décision Statistique.* University of Montreal Press, Montreal.

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

LE CAM, L. (1977). On the asymptotic normality of estimates. *Proc. of the Symp. to Honour Jerzy Neyman.* Warszawa PWN, 203–217.

STRASSEN, V. (1964). Messfehler und information. *Z. Wahrsch. verw. Gebeite.* **2** 273–305.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

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

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
UL. KPERNIKA 18, 51-617
WROCŁAW, POLAND