The Annals of Statistics
1973, Vol. 1, No. 2, 251-263

MINIMAX TESTS AND THE NEYMAN-PEARSON LEMMA FOR CAPACITIES

BY PETER J. HUBER[1] AND VOLKER STRASSEN

Swiss Federal Institute of Technology and University of Zürich

Robust test problems between two approximately known simple hypotheses can be formalized as minimax test problems between two composite hypotheses. We show that if the composite hypotheses can be described in terms of alternating capacities of order 2 (in the sense of Choquet), then the minimax tests are ordinary Neyman-Pearson tests between a fixed representative pair of simple hypotheses; moreover, the condition is in a certain sense also necessary. All the neighborhoods customarily used to formalize approximate knowledge happen to have this particular structure.

1. Introduction. Let $\mathscr{M}$ be the set of all probability measures on a complete separable metrizable space Ω, and let $\mathscr{P}_0$, $\mathscr{P}_1$ be two disjoint subsets of $\mathscr{M}$. We show that the minimax tests between $\mathscr{P}_0$ and $\mathscr{P}_1$ have a simple structure if (and in some sense only if) these hypotheses consist of the sets of all probability measures majorized by some alternating capacities v_j of order 2: $\mathscr{P}_j = \{P \in \mathscr{M} \mid P(A) \leqq v_j(A)$ for all Borel sets $A\}$. Then there is a representative pair $(Q_0, Q_1) \in \mathscr{P}_0 \times \mathscr{P}_1$ such that for all fixed sample sizes the Neyman-Pearson tests between the simple hypotheses Q_0, Q_1 constitute a minimal essentially complete class of minimax tests between $\mathscr{P}_0$ and $\mathscr{P}_1$. Conversely, if $\mathscr{P}_0$ has the property that a representative pair (Q_0, P_1) exists for all simple alternatives $\mathscr{P}_1 = \{P_1\}$, then $\mathscr{P}_0$ can be defined by some alternating capacity of order 2.

This generalizes our earlier results ([7], [8], [11], [12]) and is of fundamental importance for a theory of robust statistics: the neighborhoods used to describe inaccuracies in the specifications of the true underlying distributions can all be described in terms of alternating capacities (see below). We would like to thank F. Scholz for his critical remarks.

2. Capacities and upper and lower probabilities. Let Ω be a complete separable metrizable space, $\mathscr{A}$ its Borel-σ-algebra and $\mathscr{M}$ the set of all probability measures on Ω. Every non-empty subset $\mathscr{P} \subset \mathscr{M}$ defines an *upper probability*

$$v(A) = \sup\{P(A) \mid P \in \mathscr{P}\}, \qquad A \in \mathscr{A},$$

and a *lower probability*

$$u(A) = \inf\{P(A) \mid P \in \mathscr{P}\}, \qquad A \in \mathscr{A}.$$

Since u and v are conjugate to each other:

$$u(A) + v(A^C) = 1,$$

Received November 8, 1971; revised July 17, 1972.

[1] Work done while holding an NSF Senior Foreign Scientist Fellowship at Princeton University.

it suffices to consider one of them, say v. Evidently

$$v(\phi) = 0\,, \qquad v(\Omega) = 1\,, \tag{1}$$

$$A \subset B \Rightarrow v(A) \leqq v(B)\,, \tag{2}$$

$$A_n \uparrow A \Rightarrow v(A_n) \uparrow v(A)\,. \tag{3}$$

If $\mathscr{P}$ is weakly compact, then we have also

$$F_n \downarrow F,\ F_n \quad \text{closed} \Rightarrow v(F_n) \downarrow v(F)\,. \tag{4}$$

More precisely, if $\mathscr{P}$ is weakly closed, then (4) is equivalent to $\mathscr{P}$ being tight (see [1], and Lemmas 2.2, 2.3 below).

We shall call *capacity* any set function v satisfying (1) to (4). Apart from the trivial normalization (1), this agrees with the terminology of Choquet [3], [4]. If, in addition, v satisfies

$$v(A \cup B) + v(A \cap B) \leqq v(A) + v(B)\,, \tag{5}$$

it is called *alternating of order* 2, or short, 2-*alternating.*

The examples near the end of this section will show that not every upper probability is 2-alternating, and that the set

$$\mathscr{P}_v = \{P \in \mathscr{M} \mid P(A) \leqq v(A) \quad \text{for all} \quad A \in \mathscr{A}\}$$

can be strictly larger than the closed convex hull of the set $\mathscr{P}$ determining v.

The set function $u(A) = 1 - v(A^C)$ conjugate to a 2-alternating capacity is called a 2-*monotone capacity;* it satisfies (1), (2) and

$$A_n \downarrow A \Rightarrow u(A_n) \downarrow u(A) \tag{3'}$$

$$G_n \uparrow G,\ G_n \quad \text{open} \Rightarrow u(G_n) \uparrow u(G) \tag{4'}$$

$$u(A \cup B) + u(A \cap B) \geqq u(A) + u(B)\,. \tag{5'}$$

In view of (5) with A^C for B we have

$$u(A) \leqq v(A)$$

with equality for all A iff v is a probability measure.

We recall that every Borel set A is v-capacitable if v satisfies (1) to (4):

$$v(A) = \sup_K v(K) = \inf_G v(G)$$

where K ranges over the compact sets contained in A and G ranges over the open sets containing A (Choquet [3], [4]). The same holds with u in place of v, see below.

Let $\mathscr{C}$ be the space of bounded continuous functions on Ω, with the topology of uniform convergence, and let $\mathscr{C}_+ = \{f \in \mathscr{C} \mid f \geqq 0\}$. For any monotone bounded set function ψ on $\mathscr{A}$, satisfying $\psi(\phi) = 0$, define a functional $\tilde{\psi}$ on $\mathscr{C}_+$ by

$$\tilde{\psi}(f) = \int_0^\infty \psi\{x \mid f(x) > t\}\, dt = \int_0^\infty \psi\{x \mid f(x) \geqq t\}\, dt\,.$$

Any such functional is positive, monotone, positively homogeneous ($\tilde{\varphi}(cf) = c\tilde{\varphi}(f)$ for $c \in R_+$) and continuous; $\tilde{\varphi}$ is subadditive

$$\tilde{\varphi}(f + g) \leqq \tilde{\varphi}(f) + \tilde{\varphi}(g)$$

(or superadditive) iff φ is 2-alternating (or 2-monotone respectively), see [3] page 287 ff. If v is a measure, then $\tilde{v}$ is additive, and

$$\tilde{v}(f) = \int f\, dv\ .$$

Since Borel sets are capacitable, the values of a capacity v (or of its conjugate function u) on $\mathscr{A}$ are uniquely determined through monotonicity by the values of $\tilde{v}$ (or $\tilde{u}$) on $\mathscr{C}_+$.

LEMMA 2.1. *Let v be a capacity. The relation $\tilde{P}(f) = \int f\, dP$ defines a 1-1-correspondence between the probability measures $P \leqq v$ and the positive linear functionals $\tilde{P}$ on $\mathscr{C}$ satisfying $\tilde{P}(1) = 1$ and $\tilde{P}(f) \leqq \tilde{v}(f)$ for $f \in \mathscr{C}_+$.*

PROOF. It suffices to show that $f_n \downarrow 0, f_n \in \mathscr{C}_+$, implies $\tilde{P}(f_n) \downarrow 0$ for every $\tilde{P} \leqq \tilde{v}$ (cf. [10] page 63). We have $v\{f_n \geqq t\} \downarrow 0$ for every $t > 0$ in view of (4), hence $\tilde{P}(f_n) \leqq \tilde{v}(f_n) = \int_0^\infty v\{f_n \geqq t\}\, dt \downarrow 0$ by the monotone convergence theorem. □

LEMMA 2.2. *Let v be a capacity. The set*

$$\mathscr{P}_v = \{P \in \mathscr{M} \mid P \leqq v\} = \{p \in \mathscr{M} \mid P \geqq u\}$$

is tight and weakly closed, and hence compact.

PROOF. $\mathscr{P}_v$ is tight (cf. [1]), iff for every $\varepsilon > 0$, $\delta > 0$ there is a finite union S of spheres of radius δ such that $P(S^C) \leqq \varepsilon$ for all $P \in \mathscr{P}_v$. Let d be a metric in Ω, let $\{x_n\}$ be a countable dense set, let $g_n \in \mathscr{C}_+$ be such that $g_n(x) = 0$ for $d(x_n, x) \leqq \delta/2$, $g_n(x) = 1$ for $d(x_n, x) \geqq \delta$, and put $f_n = \inf\{g_m \mid m \leqq n\}$, $S_n = \{x \mid f_n(x) < 1\}$. Since $f_n \downarrow 0$, we have $\tilde{v}(f_n) \downarrow 0$ as in the proof of the preceding lemma, and tightness now follows from $P(S_n{}^C) \leqq v(S_n{}^C) \leqq \tilde{v}(f_n)$. Lemma 2.1 implies that $\mathscr{P}_v$ is weakly closed. □
Incidentally, a minor extension of the above proof shows that Borel sets are u-capacitable if u satisfies (1), (2), (3′), (4′). (This case does not seem to be covered by Choquet's capacitability theorems [3], [4].)

LEMMA 2.3. *Let v be an upper probability determined by a compact set $\mathscr{P}$. Then* (4) *holds.*

PROOF. Otherwise these there is a sequence $F_n \downarrow F$ and an α such that for all n, $v(F_n) > \alpha > v(F)$. Let $P_n \in \mathscr{P}$ be such that $P_n(F_n) > \alpha$ and $P \in \mathscr{P}$ be a limit point of (P_n). Then, with $P_{n_i} \to P$, $P(F_n) \geqq \limsup_i P_{n_i}(F_n) \geqq \limsup_i P_{n_i}(F_{n_i}) \geqq \alpha$, thus $P(F) \geqq \alpha$, in contradiction to $P(F) \leqq v(F) < \alpha$. □

LEMMA 2.4 *Let v be a 2-alternating capacity. For every upper semicontinuous* *function h on Ω there is a $Q \in \mathscr{P}_v$ such that for all t, $Q\{h > t\} = v\{h > t\}$ and $Q\{h \geqq t\} = v\{h \geqq t\}$. In particular, $\mathscr{P}_v$ is not empty.*

PROOF. We may assume without loss of generality that h is strictly positive and bounded (otherwise replace h by $2 + \arctan h$). There is a sequence $g_n \downarrow h$ with $g_n \in \mathscr{C}_+$; then $\lim \tilde{v}(g_n) = \tilde{v}(h)$ because of (4). Find a linear functional $\tilde{Q}$ separating the open convex set $U = \{f \in \mathscr{C} \mid \tilde{v}(|f|) < \tilde{v}(h)\}$ from the convex set $V = \{g \in \mathscr{C} \mid g \geqq h\}$ with the aid of Hahn–Banach: $\tilde{Q}(f) < \tilde{Q}(g)$ for $f \in U$, $g \in V$. We may normalize $\tilde{Q}$ such that $\inf \{\tilde{Q}(g) \mid g \in V\} = \tilde{v}(h)$; since then $\tilde{v}(|f|) < \tilde{v}(h)$ implies $\tilde{Q}(f) \leqq \tilde{v}(h)$, we must have $\tilde{Q}(f) \leqq \tilde{v}(|f|)$. It follows that $\tilde{Q}$ is induced by a σ-additive (substochastic) measure $Q \leqq v$, and that in particular

$$\tilde{Q}(h) = \int_0^\infty Q\{h > t\}\, dt = \int_0^\infty v\{h > t\}\, dt ,$$

which implies

$$Q\{h > t\} = v\{h > t\}$$

for almost all t. Because of (3), this actually holds for all t; in particular, Q is a probability, since $Q\{h > 0\} = Q(\Omega) = v(\Omega) = 1$. Since

$$v\{h \geqq t\} \leqq \lim_{t_n \uparrow t} v\{h > t_n\} = \lim_{t_n \uparrow t} Q\{h > t_n\} = Q\{h \geqq t\} ,$$

we also have $v\{h \geqq t\} = Q\{h \geqq t\}$. □

LEMMA 2.5. *Let v be a 2-alternating capacity. Then for every $A \in \mathscr{A}$ there is a $Q \in \mathscr{P}_v$ such that $Q(A) = v(A)$. It follows that v coincides with the upper probability determined by $\mathscr{P}_v$.*

PROOF. Let $F_n \subset A$ be an increasing sequence of closed sets such that $\lim v(F_n) = v(A)$. Define an upper semicontinuous function h such that $F_n = \{h \geqq t_n\}$ for some t_n, e.g. $h(x) = n^{-1}$ for $x \in F_n \backslash F_{n-1}$, $h(x) = 0$ for $x \notin \bigcup F_n$. Then apply Lemma 2.4 to find a $Q \in \mathscr{P}_v$ satisfying $Q(F_n) = v(F_n)$. It follows that $Q(A) \geqq \lim Q(F_n) = \lim v(F_n) = v(A)$. □

EXAMPLE 1. Let $\Omega = \{1, 2, 3\}$, $P_0 = (\frac{1}{2}, \frac{1}{2}, 0)$, $P_1 = (\frac{4}{6}, \frac{1}{6}, \frac{1}{6})$ and let v be the upper probability determined by $\mathscr{P} = \{P_0, P_1\}$. Then

$$\mathscr{P}_v = \left\{\left(\frac{3+t}{6}, \frac{3-t-s}{6}, \frac{s}{6}\right) \middle| 0 \leqq s,\ t \leqq 1\right\} ,$$

whereas the convex closure of $\mathscr{P}$ is the proper subset of $\mathscr{P}_v$ determined by $s = t$. In this example v is 2-alternating.

EXAMPLE 2. Let $\Omega = \{1, 2, 3, 4\}$, $P_0 = (\frac{5}{10}, \frac{2}{10}, \frac{2}{10}, \frac{1}{10})$, $P_1 = (\frac{6}{10}, \frac{1}{10}, \frac{1}{10}, \frac{2}{10})$, and let $\mathscr{P} = \{P_0, P_1\}$. Here v is not 2-alternating: let $A = \{1, 2\}$, $B = \{1, 3\}$, then

$$v(A \cup B) + v(A \cap B) = \tfrac{15}{10} > v(A) + v(B) = \tfrac{14}{10} .$$

EXAMPLE 3. Let Ω be compact. Define $v(A) = (1 - \varepsilon)P_0(A) + \varepsilon$ for $A \neq \phi$. Then v is a 2-alternating capacity, and

$$\mathscr{P}_v = \{P \mid P = (1 - \varepsilon)P_0 + \varepsilon H,\ H \in \mathscr{M}\} .$$

EXAMPLE 4. Let Ω be compact. Define $v(A) = \min(P_0(A) + \varepsilon, 1)$ for $A \neq \phi$. Then v is a 2-alternating capacity, and

$$\mathscr{P}_v = \{P \in \mathscr{M} \mid |P(A) - P_0(A)| \leqq \varepsilon \quad \text{for all} \quad A \in \mathscr{A}\} .$$

EXAMPLE 5. Let Ω be compact metric. Define $v(A) = \min(P_0(A^\delta) + \varepsilon, 1)$ for compact sets $A \neq \phi$, where A^δ is the closed δ-neighborhood of the set A, and extend v to $\mathscr{A}$. Then v is a 2-alternating capacity, and

$$\mathscr{P}_v = \{P \in \mathscr{M} \mid P(A) \leqq P_0(A^\delta) + \varepsilon \quad \text{for all} \quad A \in \mathscr{A}\} .$$

Examples 3 to 5 correspond to neighborhoods of P_0 defined by ε-contamination, total variation or Prohorov distance respectively. Further models leading to alternating capacities are described in [5], [6], [11], [12].

3. Bayes tests between capacities. Let v_0, v_1 be two 2-alternating capacities and let u_0, u_1 be their conjugates.

Let A be a critical region for testing between $\mathscr{P}_0 = \{P \in \mathscr{M} \mid P \leqq v_0\}$ and $\mathscr{P}_1 = \{P \in \mathscr{M} \mid P \leqq v_1\}$, that is, reject $\mathscr{P}_0$ if $x \in A$ is observed. Then the upper probability of falsely rejecting $\mathscr{P}_0$ is $v_0(A)$, of falsely accepting $\mathscr{P}_0$ is $v_1(A^C) = 1 - u_1(A)$.

Now assume that $\mathscr{P}_0$ is true with prior probability $t/(1 + t)$, $0 \leqq t \leqq \infty$, then the *upper Bayes risk* of the critical region A is by definition

$$\frac{t}{1+t}\, v_0(A) + \frac{1}{1+t}(1 - u_1(A)) .$$

This is minimized by minimizing the 2-alternating set function

$$w_t(A) = t v_0(A) - u_1(A) .$$

LEMMA 3.1. *For each $t \in [0, \infty]$ there is an $A_t \in \mathscr{A}$ such that*

$$w_t(A_t) = \inf_{A \in \mathscr{A}} w_t(A) .$$

PROOF. Write w instead of w_t and let $c = \inf_A w(A)$. Let $\{A_n\}$ be a sequence of sets such that

$$w(A_n) \leqq c + \varepsilon_n$$

with $\sum \varepsilon_n < \infty$.

We have $w(A_m \cup A_n) + w(A_m \cap A_n) \leqq w(A_m) + w(A_n) \leqq 2c + \varepsilon_m + \varepsilon_n$, which implies

$$w(A_m \cup A_n) \leqq c + \varepsilon_m + \varepsilon_n$$
$$w(A_m \cap A_n) \leqq c + \varepsilon_m + \varepsilon_n .$$

This immediately generalizes to

$$w(\textstyle\bigcup_{n \leqq m \leqq N} A_m) \leqq c + \sum_{n \leqq m \leqq N} \varepsilon_m$$
$$w(\textstyle\bigcap_{n \leqq m \leqq N} A_m) \leqq c + \sum_{n \leqq m \leqq N} \varepsilon_m .$$

Note that w is lower semicontinuous in the sense that for any monotone sequence

B_n, $w(\lim B_n) \leqq \lim w(B_n)$. Thus

$$w(\bigcup_{m\geqq n} A_m) \leqq c + \sum_{m\geqq n} \varepsilon_m ,$$
$$w(\bigcap_{m\geqq n} A_m) \leqq c + \sum_{m\geqq n} \varepsilon_m ,$$

and hence

$$w(\bigcap_n \bigcup_{m\geqq n} A_m) = w(\bigcup_n \bigcap_{m\geqq n} A_m) = c . \qquad \square$$

LEMMA 3.2. *The sets A_t can be chosen to form a decreasing family; more precisely, we may assume either $A_t = \bigcup_{s>t} A_s$ or $A_t = \bigcap_{s<t} A_s$ for all $0 < t < \infty$, whichever is more convenient. In either case, we may put $A_0 = \bigcup_{t>0} A_t$, $A_\infty = \bigcap_{t<\infty} A_t$.*

PROOF. Let $t < s$; since $w_s - w_t = (s-t)v_0$ is a monotone set function, we have

$$w_t(A \cup B) - w_s(A \cup B) \leqq w_t(A) - w_s(A) .$$

We add to this the inequality

$$w_s(A \cup B) + w_s(A \cap B) \leqq w_s(A) + w_s(B)$$

and obtain

$$w_t(A \cup B) + w_s(A \cap B) \leqq w_t(A) + w_s(B) .$$

Now choose sets A_t' minimizing w_t and insert $A = A_t'$, $B = A_s'$, which gives

$$w_t(A_t' \cup A_s') + w_s(A_t' \cap A_s') \leqq w_t(A_t') + w_s(A_s') .$$

This implies at once that

$$w_t(A_t' \cup A_s') = w_t(A_t') = \inf_A w_t(A)$$
$$w_s(A_t' \cap A_s') = w_s(A_s') = \inf_A w_s(A) .$$

Now take any sequence $\{t_n\}$ dense in $(0, \infty)$. Since w_{t_n} is lower semicontinuous (cf. the proof of Lemma 3.1), the result just established implies that also

$$A''_{t_n} = \bigcup \{A'_{t_m} \mid t_m \geqq t_n\}$$

and

$$A'''_{t_n} = \bigcap \{A'_{t_m} \mid t_m \leqq t_n\}$$

minimize w_{t_n}. Since $|w_s(A) - w_t(A)| \leqq |s - t|$, we may define A_t for any finite t by

$$A_t = \bigcup_{t_n > t} A''_{t_n}$$

or by

$$A_t = \bigcap_{t_n < t} A'''_{t_n}$$

Note that $w_t(A_t) \leqq w_t(\phi) = 0$, hence $v_0(A_t) \leqq t^{-1}$; we have in particular $v_0(A_\infty) = 0$, and the result holds also for $t = \infty$. $\square$

Choose (for instance) the first variant of Lemma 3.2 and define

$$\pi(x) = \inf \{t \mid x \notin A_t\}$$

then $x \in A_t$ is equivalent with $\pi(x) > t$.

We call π the *Radon–Nikodym derivative of v_1 with respect to v_0*, for the following reason.

Assume $t < s$, $A_t \supset B \supset A_s$, then $w_t(B) \geqq w_t(A_t)$, $w_s(B) \geqq w_s(A_s)$ implies

$$t[v_0(B) - v_0(A_t)] \geqq u_1(B) - u_1(A_t)$$
$$s[v_0(B) - v_0(A_s)] \geqq u_1(B) - u_1(A_s)$$

or, if the quotients are well-defined,

$$\frac{u_1(B) - u_1(A_s)}{v_0(B) - v_0(A_s)} \leqq s\,,$$
$$\frac{u_1(A_t) - u_1(B)}{v_0(A_t) - v_0(B)} \geqq t\,.$$

If v_0, v_1 are measures, then $u_0 = v_0$, $u_1 = v_1$, and a comparison with the usual proof of the Rand–Nikodym theorem yields that π is the Radon–Nikodym derivative dv_1/dv_0.

We note for future reference that the above inequalities imply in particular

$$t \leqq \frac{u_1(A_t) - u_1(A_s)}{v_0(A_t) - v_0(A_s)} \leqq s\,.$$

REMARK. If v_0 is such that $v_0(A) \geqq \varepsilon_0 > 0$ for $A \neq \phi$ (as in the Examples 3 to 5 of Section 2), then the last part of the proof of Lemma 3.2 shows that $A_t = \phi$ for $t > 1/\varepsilon_0$, hence π is bounded: $\pi(x) \leqq 1/\varepsilon_0$. Similarly, if $v_1(A) \geqq \varepsilon_1 > 0$ for $A \neq \phi$, then $\pi(x) \geqq \varepsilon_1$.

4. The main theorem. The preceding section implies that A_t is a critical region for a minimax (sometimes called maximin) test of level $\alpha = v_0(A_t)$ and guaranteed power $u_1(A_t)$ for testing between $\mathscr{P}_0 = \{P \in \mathscr{M} \mid P \leqq v_0\}$ and $\mathscr{P}_1 = \{P \in \mathscr{M} \mid P \leqq v_1\}$ (cf. [9] page 327).

Actually, a much stronger statement is true.

THEOREM 4.1. *Assume v_0, v_1 are 2-alternating capacities and let π be defined as in Section 3. Then there are probability measures $Q_0 \leqq v_0$, $Q_1 \leqq v_1$ such that for all t*

$$Q_0\{\pi > t\} = v_0\{\pi > t\}$$
$$Q_1\{\pi > t\} = u_1\{\pi > t\}$$

and π is a version of dQ_1/dQ_0.

COROLLARY 4.2. *For any sample size n and any level α, the Neyman–Pearson test of level α, between Q_0 and Q_1, defined by*

$$\begin{aligned} \phi(x_1, \cdots, x_n) &= 1 \quad \textit{for} \quad \textstyle\prod_{i \leqq n} \pi(x_i) > C \\ &= \gamma \qquad\qquad\qquad\quad = C \\ &= 0 \qquad\qquad\qquad\quad < C\,, \end{aligned}$$

where C and γ are chosen such that $E_{Q_0}\phi = \alpha$, is also a minimax test between $\mathscr{P}_0$ and $\mathscr{P}_1$, with the same level α and the same minimum power.

PROOF OF THE COROLLARY. Note that we have also $Q_0\{\pi \geqq t\} = v_0\{\pi \geqq t\}$ and $Q_1\{\pi \geqq t\} = u_1\{\pi \geqq t\}$, cf. the end of the proof of Lemma 2.4 and (3′). Thus, for sample size 1, the corollary essentially coincides with the theorem. For sample size n, assume that the x_i are independent and distributed according to $Q_i' \leqq v_0$, $1 \leqq i \leqq n$. Then $\prod_{i \leqq n} \pi(x_i)$ is stochastically largest, if each factor $\pi(x_i)$ is made stochastically largest (this follows from [9] page 73, Lemma 1), and according to the theorem, this happens if $Q_i' = Q_0$. Similarly, if $Q_i' \geqq u_1$, then $\prod_{i \leqq n} \pi(x_i)$ is made stochastically smallest with $Q_i' = Q_1$. □

PROOF OF THE THEOREM. At first, one might attempt to choose a Q_0 in the manner of Lemma 2.4, with π in place of h, and then to define $dQ_1 = \pi dQ_0$. However, it is easy to see that this does not ensure $Q_1 \geqq u_1$. Loosely speaking, we will thus have to choose Q_0 such that $\pi^{-1}u_1 \leqq Q_0 \leqq v_0$, which is made precise in the following.

Define for $t < s$

$$F(t, s; B) = u_1((B \cap A_t) \cup A_s) - u_1(A_s)$$
$$G(t, s; B) = v_0((B \cap A_t) \cup A_s) - v_0(A_s) .$$

These set functions inherit their monotonicity and continuity properties from u_1 and v_0 respectively; in particular, $F(t, s; \cdot)$ is 2-monotone, and $G(t, s; \cdot)$ is 2-alternating. The inequalities near the end of Section 3 imply

$$F(t, s; B) \leqq sG(t, s; B)$$

and we have

$$G(t, s; B) \leqq v_0(B \cap A_t) - v_0(B \cap A_s) \leqq v_0(B) .$$

Since u_1 is 2-monotone and v_0 is 2-alternating, one has for $t_1 < t_2 < t_3$:

$$F(t_1, t_3; B) \leqq F(t_1, t_2; B) + F(t_2, t_3; B)$$
$$G(t_1, t_3; B) \geqq G(t_1, t_2; B) + G(t_2, t_3; B) .$$

For any finite increasing sequence $J = (t_0, t_1, \cdots, t_n)$ of positive real numbers define

$$u_1^J(B) = \sum_{i=1}^n t_i^{-1} F(t_{i-1}, t_i; B) .$$

We have

$$u_1^J(B) \leqq \sum_{i=1}^n G(t_{i-1}, t_i; B) \leqq G(t_0, t_n; B) \leqq v_0(B) .$$

If the underlying set $|J'|$ of J' contains the underlying set $|J|$ of J, then, evidently

$$u_1^J(B) \leqq u_1^{J'}(B) .$$

We define

$$u_1^*(B) = \sup_J u_1^J(B) ,$$

which is a positive, monotone and 2-monotone set function satisfying $u_1^* \leqq v_0$.

Consider in particular $B = A_t$, then

$$\begin{aligned} G(t_1, t_2; A_t) &= v_0(A_{t_1}) - v_0(A_{t_2}) && \text{for } t \leqq t_1 \\ &= v_0(A_t) - v_0(A_{t_2}) && \text{for } t_1 < t < t_2 \\ &= 0 && \text{for } t \geqq t_2 , \end{aligned}$$

and analogous expressions hold for F. The inequalities near the end of Section 3 imply

$$F(t_1, t_2; A_t) \geqq t_1 G(t_1, t_2; A_t) .$$

Thus, if J is such that $t_{i-1}/t_i \geqq \alpha$ for all i, and $t = t_i$ for some i, then

$$u_1^J(A_t) \geqq \sum_{i=1}^n \frac{t_{i-1}}{t_i} G(t_{i-1}, t_i; A_t) \geqq \alpha \sum_{i=1}^n G(t_{i-1}, t_i; A_t)$$
$$= \alpha[v_0(A_t) - v_0(A_{t_n})] .$$

Since $\alpha < 1$ can be made arbitrarily large, and since $v_0(A_{t_n})$ can be made arbitrarily small, we obtain $u_1^*(A_t) = v_0(A_t)$, and in particular $u_1^*(\Omega) = v_0(A_0)$. If $v_0(A_0) < 1$, we replace u_1^* by $u_1^* + Q \mid A_0^C$, where $Q \leqq v_0$ is a probability satisfying $Q(A_0) = v_0(A_0)$, so that $u_1^*(\Omega) = 1$.

With the aid of Hahn–Banach one finds a linear functional $\tilde{Q}_0$ on $\mathscr{C}$ separating the convex set

$$U = \{f \in \mathscr{C}_+ \mid \tilde{u}_1^*(f) \geqq 1\}$$

from the open convex set

$$V = \{g \in \mathscr{C} \mid \tilde{v}_0(|g|) < 1\} :$$
$$\tilde{Q}_0(g) < \tilde{Q}_0(f) \qquad \text{for } f \in U, g \in V .$$

$\tilde{Q}_0$ is positive and can be normalized: $\tilde{Q}_0(1) = 1$. It is straightforward to see that $\tilde{u}_1^* \leqq \tilde{Q}_0 \leqq \tilde{v}_0$ on $\mathscr{C}_+$, so $\tilde{Q}_0$ defines a probability $Q_0 \leqq v_0$, which satisfies $Q_0(A_t) = v_0(A_t)$ for all t.

Now we define Q_1 by defining separately its restriction to A_∞ and to A_∞^C.

(a) With the aid of Lemma 2.5 find a $Q \geqq u_1$ such that $Q(A_\infty) = u_1(A_\infty)$, put $Q_1 \mid A_\infty = Q \mid A_\infty$.

(b) On A_∞^C put $dQ_1 = \pi(x)\, dQ_0$.

We have, with $J = (t_1, t_2)$,

$$\begin{aligned} Q_1((B \cap A_{t_1}) \cup A_{t_2}) - Q_1(A_{t_2}) &= Q_1(B \cap (A_{t_1} \backslash A_{t_2})) \\ &\geqq t_1 Q_0(B \cap (A_{t_1} \backslash A_{t_2})) \\ &\geqq t_1 u_1^J(B) \\ &= \frac{t_1}{t_2} F(t_1, t_2; B) . \end{aligned}$$

If $J = (t_0, \cdots, t_n)$ satisfies $t_{i-1}/t_i \geqq \alpha$ for all i, we obtain thus

$$\begin{aligned} Q_1((B \cap A_{t_0}) \cup A_{t_n}) - Q_1(A_{t_n}) &\geqq \alpha \textstyle\sum F(t_{i-1}, t_i; B) \\ &\geqq \alpha F(t_0, t_n; B) \\ &= \alpha[u_1((B \cap A_{t_0}) \cup A_{t_n}) - u_1(A_{t_n})] \\ &\geqq \alpha[u_1(B \cap A_{t_0}) - u_1(B \cap A_{t_n})] . \end{aligned}$$

Note that $u_1(B \cap A_0) - u_1(B \cap A_t) \leqq u_1((B \cap A_0) \cup A_t) - u_1(A_t) \leqq t[v_0((B \cap A_0) \cup A_t) - v_0(A_t)] \leqq t$; so, if we go to the limit $\alpha \to 1$, $t_0 \to 0$, $t_n \to \infty$, we obtain

$$Q_1(B \cap A_0) - Q_1(B \cap A_\infty) \geqq u_1(B \cap A_0) - u_1(B \cap A_\infty) .$$

Since $u_1(B \cup A_0) + u_1(B \cap A_0) \geqq u_1(B) + u_1(A_0)$ and $u_1(A_0) = 1$, we have $u_1(B \cap A_0) = u_1(B)$. Hence

$$Q_1(B) \geqq u_1(B) + Q_1(B \cap A_\infty) - u_1(B \cap A_\infty) \geqq u_1(B) .$$

It remains to show that $Q_1(A_t) = u_1(A_t)$. But for $t_1 < t_2$

$$\begin{aligned} Q_1(A_{t_1}) - Q_1(A_{t_2}) &\leqq t_2[Q_0(A_{t_1}) - Q_0(A_{t_2})] \\ &= t_2(v_0(A_{t_1}) - v_0(A_{t_2})) \\ &\leqq \frac{t_2}{t_1}(u_1(A_{t_1}) - u_1(A_{t_2})) . \end{aligned}$$

Thus, if $t = t_0 < t_1 < \cdots < t_n$, with $t_i/t_{i-1} \leqq \alpha$, we obtain

$$Q_1(A_t) - Q_1(A_{t_n}) \leqq \alpha(u_1(A_t) - u_1(A_{t_n})) ,$$

and letting $\alpha \to 1$, $t_n \to \infty$,

$$Q_1(A_t) - Q_1(A_\infty) \leqq u_1(A_t) - u_1(A_\infty) ,$$

hence

$$Q_1(A_t) = u_1(A_t) .$$

This terminates proof of the theorem. □

5. Uniqueness of π.

THEOREM 5.1. *Let π, π' be two versions of dv_1/dv_0, i.e., both $A_t = \{\pi > t\}$ and $A_t' = \{\pi' > t\}$ minimize w_t. Then $v_0\{\pi > t\} = v_0\{\pi' > t\}$ for all t, and $\pi = \pi'$* a.e. $[Q_0 + Q_0']$, *where Q_0, Q_0' are determined by π, π' respectively, as in Section* 4.

PROOF. Since also $\min(\pi, \pi')$ and $\max(\pi, \pi')$ are versions, we may assume $\pi \leqq \pi'$ without loss of generality.

If $v_0\{\pi > t\} = v_0\{\pi' > t\}$ for all t, then $v_0\{\pi > t\} = Q_0\{\pi > t\} \leqq Q_0\{\pi' > t\} \leqq v_0\{\pi' > t\}$, hence $\pi = \pi'$ a.e. Q_0, and similarly for Q_0'.

If not, then there is a t such that $v_0\{\pi > t\} < v_0\{\pi' > t\}$. Let $A_t' = \{\pi' > t\} \supset A_t = \{\pi > t\}$, $D = A_t' \backslash A_t$. Since $w_t(A_t) = w_t(A_t')$, one has

$$\frac{u_1(A_t') - u_1(A_t)}{v_0(A_t') - v_0(A_t)} = t \geqq \frac{Q_1'(A_t') - Q_1'(A_t)}{Q_0'(A_t') - Q_0'(A_t)} = \frac{\int_D \pi'(x)\, dQ_0'}{\int_D dQ_0'} .$$

Since $Q_0'(D) > 0$ and $\pi'(x) > t$ on D, this is impossible. □

6. A characterization of (Q_0, Q_1).

THEOREM 6.1. *Let Φ be any twice continuously differentiable function on* $[0, 1]$, *such that $\Phi'' > 0$. Then the pair $(Q_0, Q_1) \in \mathscr{P}_0 \times \mathscr{P}_1$ satisfies the conclusion of Theorem* 4.1 iff *it minimizes*

$$H(P_0, P_1) = \int \Phi\left(\frac{dP_0}{d(P_0 + P_1)}\right) d(P_0 + P_1)$$

among all $(P_0, P_1) \in \mathscr{P}_0 \times \mathscr{P}_1$.

PROOF. Let $Q_{ji} \in \mathscr{P}_j$ $(i, j = 0, 1)$ and let q_{ji} be their densities respect to, say $\mu = \sum_{ji} Q_{ji}$. Put

$$Q_{jt} = (1 - t)Q_{j0} + tQ_{j1}$$
$$q_{jt} = (1 - t)q_{j0} + tq_{j1} .$$

Then it follows from

$$\frac{d}{dt} H(Q_{0t}, Q_{1t}) = \int \left\{ \Phi\left(\frac{q_{0t}}{q_{0t} + q_{1t}}\right)(q_{01} + q_{11} - q_{00} - q_{10}) + \Phi'\left(\frac{q_{0t}}{q_{0t} + q_{1t}}\right) \frac{q_{01}q_{10} - q_{00}q_{11}}{q_{0t} + q_{1t}} \right\} d\mu ,$$

$$\frac{d^2}{dt^2} H(Q_{0t}, Q_{1t}) = \int \Phi''\left(\frac{q_{0t}}{q_{0t} + q_{1t}}\right) \frac{(q_{01}q_{10} - q_{00}q_{11})^2}{(q_{0t} + q_{1t})^3} d\mu \geqq 0 ,$$

that H is convex, and that (Q_{00}, Q_{10}) minimizes H if and only if for all $(Q_{01}, Q_{11}) \in \mathscr{P}_0 \times \mathscr{P}_1$.

$$\left[\frac{d}{dt} H(Q_{0t}, Q_{1t})\right]_{t=0} \geqq 0 .$$

If we introduce the functions

$$\phi(z) = z\Phi'(z) - \Phi(z) ,$$
$$\psi(z) = (1 - z)\Phi'(z) + \Phi(z) ,$$

which are strictly increasing for $0 < z < 1$, we can write, with $z = q_{00}/(q_{00} + q_{10})$,

$$\left[\frac{d}{dt} H(Q_{0t}, Q_{1t})\right]_{t=0} = \int \phi(z)(q_{10} - q_{11})\, d\mu + \int \psi(z)(q_{01} - q_{00})\, d\mu .$$

Now assume that (Q_{00}, Q_{10}) satisfies Theorem 4.1, then $z = (1 + \pi)^{-1}$, and

(*)
$$\int \phi\left(\frac{1}{1 + \pi}\right)(q_{10} - q_{11})\, d\mu \geqq 0$$
$$\int \psi\left(\frac{1}{1 + \pi}\right)(q_{01} - q_{00})\, d\mu \geqq 0$$

since Q_{00} makes π stochastically largest, Q_{10} stochastically smallest. Hence (Q_{00}, Q_{10}) minimizes H.

Now let (Q_{01}, Q_{11}) be another pair minimizing H; since H is convex, $H(Q_{0t}, Q_{1t})$ must be constant, hence $q_{01}q_{10} - q_{00}q_{11} \doteq 0$ a.e. μ, thus $q_{11}/q_{01} = q_{10}/q_{00} = \pi$ a.e. μ.

Hence, if (Q_{00}, Q_{10}) is any pair minimizing H, then (*) is satisfied for all $(Q_{01}, Q_{11}) \in \mathscr{P}_0 \times \mathscr{P}_1$, and it follows that Q_{00} makes π stochastically largest, Q_{10} stochastically smallest. □

REMARK. If the probability measures in $\mathscr{P}_0$, $\mathscr{P}_1$ are absolutely continuous with respect to fixed measure μ, for instance when Ω is finite, it is possible to prove Theorem 4.1 by first picking a pair minimizing (H for a suitable Φ, H is lower semi-continuous) and then selecting a suitable version of $\pi = dQ_1/dQ_0$.

But for infinite Ω, the more interesting sets $\mathscr{P}_0$, $\mathscr{P}_1$ are not dominated, and then it is not evident whether it is possible to select a suitable π.

7. Necessity of 2-alternating capacities. For this section we assume that Ω is finite, then the continuity assumptions (3), (4) of Section 2 are trivially satisfied.

THEOREM 7.1. *Let $\mathscr{P} \subset \mathscr{M}$ be such that for every $Q_1 \in \mathscr{M}$ there is a $Q_0 \in \mathscr{P}$ with the property that for all $\alpha \in (0, 1)$ the Neyman–Pearson test of level α between Q_0 and Q_1 is a test of the same level between $\mathscr{P}$ and Q_1. Then $\mathscr{P}$ is* (i) *convex and* (ii) *compact,* (iii) $v(A) = \sup\{P(A) \mid P \in \mathscr{P}\}$ *is 2-alternating, and* (iv) $\mathscr{P} = \mathscr{P}_v = \{P \in \mathscr{M} \mid P \leqq v\}$.

PROOF. (i) If $\mathscr{P}$ is not convex, then let $Q_1 = \sum a_i P_i \notin \mathscr{P}$ with $P_i \in \mathscr{P}$, $a_i \geqq 0$, $\sum a_i = 1$. Now let ϕ be any test between $\mathscr{P}$ and Q_1 such that $\sup_{P \in \mathscr{P}} \int \phi \, dP \leqq \alpha$, then $\int \phi \, dQ_1 \leqq \alpha$, but the power of any Neyman–Pearson test of level α between a $P \in \mathscr{P}$ and Q_1 must exceed α. Thus $\mathscr{P}$ is convex.

(ii) A similar argument shows that $\mathscr{P}$ is closed and hence compact.

(iii) We have to verify that v is 2-alternating, or, equivalently, that $\tilde{v}$ is subadditive. For that, it suffices to show that for all functions $h > 0$ there is a $Q \in \mathscr{P}$ such that

$$Q\{h > t\} = \sup_{P \in \mathscr{P}} P\{h > t\} .$$

In fact, then for any f, g there is a Q, such that

$$\begin{aligned} \tilde{v}(f + g) &= \int_0^\infty Q\{f + g > t\} \, dt = \int (f + g) \, dQ \\ &= \int f \, dQ + \int g \, dQ \leqq \tilde{v}(f) + \tilde{v}(g) . \end{aligned}$$

By the assumption of the theorem, for each Q_1 there is a $Q_0 \in \mathscr{P}$ such that

$$Q_0 \left\{ \frac{dQ_1}{dQ_0} > t \right\} = \sup_{P \in \mathscr{P}} P \left\{ \frac{dQ_1}{dQ_0} > t \right\} \qquad \text{for all } t .$$

Now let $h > 0$ be given; the idea is to find a Q_1 such that $h \sim dQ_1/dQ_0$. The construction is as follows.

Let $Q \in \mathscr{P}$ be such that

$$\int \log h(x) \, dQ = \sup_{P \in \mathscr{P}} \int \log h(x) \, dP .$$

Define Q_1 by $dQ_1 = ch \, dQ$, where c is a normalizing constant. Evidently, we have then $dQ_1/dQ = ch$, and

$$\int \log \left(\frac{dQ_1}{dQ} \right) dQ = \sup_{P \in \mathscr{P}} \int \log \left(\frac{dQ_1}{dQ} \right) dP .$$

But this property is equivalent with

$$\int \log \left(\frac{dQ_1}{dQ} \right) dQ = \sup_{P \in \mathscr{P}} \int \log \left(\frac{dQ_1}{dP} \right) dP .$$

To show that both extremal problems have the same solution, put $P_t =$

$(1 - t)Q + tP_1$, then

$$\left[\frac{d}{dt} \int \log\left(\frac{dQ_1}{dP_t}\right) dP_t\right]_{t=0} = \int \log\left(\frac{dQ_1}{dQ}\right) d(P_1 - Q) \leqq 0$$

is the common necessary and sufficient condition for a maximum to occur at Q. A convexity argument similar to one in the proof of Theorem 6.1 shows that the P maximizing $\int \log (dQ_1/dP)\, dP$ is unique; hence we must have $Q = Q_0$.

(iv) To show that $\mathscr{P} = \mathscr{P}_v$, we shall use the fact that every convex compact subset of $\mathbf{R}^n$ is the convex closure of its points of strict convexity (cf. [2]). Thus, it suffices to show that each point of strict convexity of $\mathscr{P}_v$ belongs to $\mathscr{P}$. By definition, if Q is a point of strict convexity of $\mathscr{P}_v$, then there is a f with the property that

$$\int f\, dQ = \sup_{P \in \mathscr{P}_v} \int f\, dP$$

uniquely characterizes Q. Without loss of generality we may assume $f > 0$. But then

$$\int f\, dQ = \tilde{v}(f) = \sup_{P \in \mathscr{P}} \int f\, dP\,.$$

Since Q is unique, we must have $Q \in \mathscr{P}$. □

REFERENCES

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

[2] BOURBAKI, N. (1966). *Espaces vectoriels topologiques* (2e edition). Hermann, Paris, Chapter 2, § 7, Exercice 8b.

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

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

[5] DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38** 325–339.

[6] DEMPSTER, A. P. (1968). A generalization of Bayesian inference. *J. Royal Statist. Soc. Sci. B.* **30** 205–247.

[7] HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758.

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

[9] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[10] NEVEU, J. (1964). *Bases mathématiques du calcul des probabilités.* Masson, Paris.

[11] STRASSEN, V. (1964). Messfehler und Information. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **2** 273–305.

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

ETH—ZURICH
CLAUSIUSTRASSE 55
8006 ZURICH
SWITZERLAND