UPPER AND LOWER RISKS AND MINIMAX PROCEDURES

R. J. BERAN
UNIVERSITY OF CALIFORNIA, BERKELEY

The essential goal of R. A. Fisher's fiducial argument was to make posterior inferences about unknown parameters without resorting to a prior distribution. Over the past decade, there have been two major attempts at developing a statistical theory that would accomplish this convincingly. One of these efforts has been described in a series of publications by Fraser, the other in papers by Dempster. From the early work [4], [11], [12], [13], which was tied to a fiducial viewpoint, both authors developed statistical theories that were distinct from the fiducial argument, yet achieved the goal of non-Bayesian posterior inference [5], [6], [7], [8], [14], [15], [16].

Despite technical and other differences, the main ideas underlying this later work by Dempster and by Fraser appear to be similar. Fraser's papers, analyzing statistical models that possess a special kind of structure, arrive at "structural probability" distributions for the unknown parameters. Dempster's papers, dealing with less specialized models, derive "upper and lower probabilities" on the parameter space. Disregarding some technicalities, these upper and lower probabilities reduce to structural probabilities for the models considered by Fraser.

To this extent, upper and lower probabilities are a generalization of structural probabilities. However, there appear to be differences in interpretation. Fraser has given a frequency interpretation to structural probabilities in [11], [12] (but not in later work); this interpretation depends upon the special form of the statistical models in his theory, and does not apply to Dempster's theory. Dempster has provided no simple interpretation for upper and lower probabilities; he suggested in [7] that his theory might be "an acceptable idealization of intuitive inferential 'appreciations'." More recently, he has embedded his theory within a generalized Bayesian framework [9], [10]. The justification for the latter is unclear at present (see the discussion to [9]).

Lacking in both the Dempster and Fraser theories are systematic methods for dealing with estimation and hypothesis testing problems (or suitable analogues of such). A method of constructing tests was described by Fraser in [16], but no performance criteria were established. Dempster [5] defined upper and lower risks but did not pursue their application; the statistical meaning of these risks

This research was partially supported by National Science Foundation Grant GP-15283.

is not evident under his interpretation of upper and lower probabilities. Since even simple models suggest a variety of natural estimates and tests, some theory seems necessary as a guide to choice of procedure.

The results presented in this paper proceed in several directions. A statistical interpretation for upper and lower probabilities and risks is described in Section 2; this rationale leads naturally to a minimax criterion for statistical procedures and, in principle, to an alternative to standard decision theory. The desirability of such an alternative stems from well-known awkward features of standard decision theory, such as the possibility that a test of low size and high power may make a decision which is contradicted by the data (see Hacking [17]). A heuristic account of these ideas in a less general context has previously been given by the author in [1].

Section 3 of the paper develops basic mathematical properties of upper and lower probabilities and risks in the light of Choquet's [3] theory of capacities. The results include extensions of properties given by Dempster in [6].

In Section 4, convenient conditions are established for the existence of minimax procedures (as defined in Section 2). An example in a nonparametric setting follows.

2. Statistical background

An experiment is performed, resulting in observation x. It is known that the observed x was generated from a parameter value t and a realized random variable e by the mapping

$$x = \xi(e, t). \tag{2.1}$$

Moreover, t lies in a parameter space T, x lies in an observation space X, and e is realized according to a probability measure P on an elementary space E. Both P and the mapping ξ are known. The problem is to draw inferences concerning t from x and the model.

The following formal assumptions are made: X is a Borel subset of a metric space and is endowed with the σ-algebra $\mathscr{X}$ of all Borel sets. T and E are complete separable metric spaces, endowed with σ-algebras $\mathscr{T}$ and $\mathscr{E}$, respectively. $\mathscr{T}$ consists of all Borel sets in T. P is defined on the Borel sets in E and $\mathscr{E}$ is the completion with respect to P of the σ-algebra of these Borel sets; thus $\mathscr{E}$ contains all analytic sets. The function $\xi: E \times T \to X$ is Borel measurable.

Formally, performing the experiment described above amounts to realizing, through physical operations, a specific triple $(x, t, e) \in X \times T \times E$. Before the experiment is carried out (or the outcome x is noted), the following prospective assertions can be made about the triple to be realized: the chance that $e \in B$, $B \in \mathscr{E}$, is $P(B)$; t is an unspecified element of T; the observable x is related to t and e through (2.1).

Once the experiment has been performed and x has been observed, the particular triple (x, t, e) that was realized can be described more precisely. If

$$T_x(e) = \{t \in T : x = \xi(e, t)\}, \tag{2.2}$$

it is evident that the e realized in the experiment must lie in

$$E_x = \{e \in E : T_x(e) \neq \varnothing\}, \tag{2.3}$$

and whatever that e is, the realized t must belong to the corresponding $T_x(e)$.

Since $E_x = \text{proj}_E[\xi^{-1}(x)]$, E_x is analytic under the assumptions and so lies in $\mathscr{E}$. Let $P[B|E_x]$ denote the conditional probability defined by

$$P[B|E_x] = \frac{P[B \cap E_x]}{P[E_x]}, \qquad B \in \mathscr{E}, \tag{2.4}$$

provided $P[E_x] > 0$. If $P[E_x] = 0$, it may still be possible to condition upon a suitable statistic. In any event, a modification of ξ so as to include round off error incurred in observing x will generally result in $P[E_x] > 0$.

Thus, after the experiment has been performed and x has been observed, the following prospective statements can be made about the realized triple (x, t, e): x is as observed; $e \in X_x$ and the chance that $e \in B$, $B \in \mathscr{E}$, is $P[B|E_x]$; whatever e is, t is an unspecified element of the corresponding set $T_x(e)$; relation (2.1) is necessarily satisfied. This collection of assertions about the triple (x, t, e) will be called the *posterior model* $\mathscr{M}_x$ for the experiment. Both Dempster and Fraser have previously considered reductions of this type, though not in terms of experimental triples.

Since the realized experiment (x, t, e) is described more precisely by the posterior model $\mathscr{M}_x$ than by the original model, it is proposed to evaluate statistical procedures of interest by their average behavior over a hypothetical sequence of independent experiments, each of which is generated under the assumptions of $\mathscr{M}_x$. The aim is to measure how well a statistical procedure performs when applied to hypothetical experimental triples that are as similar as can be arranged to the actual triple (x, t, e).

Let D denote a space of decisions and let $\ell : T \times D \to R^+$ be a nonnegative loss function. Let $\mathscr{R}^+$ denote the σ-algebra of all Borel sets in R^+, and assume that for every $d \in D$, $\ell(\cdot, d)$ is a measurable mapping of $(T, \mathscr{T})$ into $(R^+, \mathscr{R}^+)$. Suppose $d \in D$ is a specific decision whose consequences are to be evaluated relative to the posterior model $\mathscr{M}_x$ under the loss function ℓ.

Let $\{(x, t_i, e_i), i = 1, 2, \cdots\}$ be a sequence of independent hypothetical experiments generated under the posterior model; in other words, $e_1, e_2, \cdots$ are independent random variables, each distributed according to $P[\cdot|E_x]$, t_i is selected arbitrarily from $T_x(e_i)$, x is the observed data. For each i, the equation $x = \xi(e_i, t_i)$ will necessarily be satisfied.

Let the general notation $\text{prop}_n(\pi_i)$ denote the proportion of true propositions among the propositions $\{\pi_i, \pi_2, \cdots, \pi_n\}$. The average loss incurred over the first n hypothetical experiment as a result of taking decision d is $n^{-1} \Sigma_{i=1}^n \ell(t_i, d)$. Since $\ell \geqq 0$,

$$\frac{1}{n} \sum_{i=1}^{n} \ell(t_i, d) = \int_0^\infty \text{prop}_n[\ell(t_i, d) > z]\, dz. \tag{2.5}$$

If $A(z, d) = \{t \in T : \ell(t, d) > z\}$, then $A(z, d) \in \mathscr{T}$ for every $z \in R^+$ and $d \in D$, and $\{\ell(t_i, d) > z\} = \{t_i \in A(z, d)\}$. Therefore,

$$\begin{aligned} (2.6) \qquad \int_0^\infty \text{prop}_n[T_x(e_i) \subset A(z, d), T_x(e_i) \neq \varnothing]\, dz. \\ \leqq \frac{1}{n} \sum_{i=1}^n \ell(t_i, d) \\ \leqq \int_0^\infty \text{prop}_n[T_x(e_i) \cap A(z, d) \neq \varnothing]\, dz. \end{aligned}$$

Now,

$$(2.7) \qquad \int_0^\infty \text{prop}_n[T_x(e_i) \cap A(z, d) \neq \varnothing]\, dz = \frac{1}{n} \sum_{i=1}^n \int_0^\infty I_{C_i}(z)\, dz,$$

where $C_i = \{z \in R^+ : T_x(e_i) \cap A(z, d) \neq \varnothing\}$ and $I_{C_i}(z)$ is the indicator of C_i. Moreover,

$$(2.8) \qquad \int_0^\infty I_{C_i}(z)\, dz = \sup_{t \in T_x(e_i)} \ell(t, d),$$

and for every $d \in D$, the function on the right of (2.8) is a measurable mapping of $(E, \mathscr{E})$ into $(R^+, \mathscr{R}^+)$.

By Fubini's theorem,

$$(2.9) \qquad E \int_0^\infty I_{C_i}(z)\, dz = \int_0^\infty v_x[A(z, d)]\, dz,$$

where for $A \in \mathscr{T}$,

$$(2.10) \qquad v_x(A) = P[e : T_x(e) \cap A \neq \varnothing | E_x]$$

and the expectation is with respect to $P[\cdot | E_x]$.

Since $\{e : T_x(e) \cap A \neq \phi\} = \text{proj}_E[\xi^{-1}(x) \cap E \times A]$, this set is analytic, belongs to $\mathscr{E}$, and therefore $v_x(A)$ is defined. The strong law of large numbers, applied to (2.7), shows that as $n \to \infty$, the upper bound in (2.6) converges with probability one to

$$(2.11) \qquad s_x(\ell, d) = \int_0^\infty v_x[A(z, d)]\, dz.$$

A dual argument shows that the lower bound in (2.6) converges with probability one, as $n \to \infty$, to

$$(2.12) \qquad r_x(\ell, d) = \int_0^\infty u_x[A(z, d)]\, dz,$$

where for $A \in \mathscr{T}$,

$$(2.13) \qquad u_x(A) = P[e : T_x(e) \subset A, T_x(e) \neq \varnothing | E_x].$$

Thus, the lower risk $r_x(\ell, d)$ and the upper risk $s_x(\ell, d)$ measure the smallest and largest long run average loss that could be incurred as a consequence of decision d. The evaluation is made under the posterior model $\mathscr{M}_x$. The relative desirability of various decisions $d \in D$ may be assessed by reference to the corresponding risks. More generally, a decision procedure $\delta: X \to D$ may be compared with other decision procedures by studying the risks as functions on X.

For $\ell \geqq 0$, $s_x(\ell, d)$ and $r_x(\ell, d)$ are equivalent to the upper and lower expectations defined by Dempster in [5], [6]; v_x and u_x are the corresponding upper and lower probabilities defined on $(T, \mathscr{T})$. A frequency interpretation for u_x, v_x is obtained by specializing ℓ in the foregoing; see [1] for the result.

The frequency interpretation for r_x and s_x suggests the following simple optimality criterion.

DEFINITION 2.1. *A decision $d \in D$ is minimax under loss function ℓ and observations x if $s_x(\ell, d) \leqq s_x(\ell, d')$ for every $d' \in D$.*

This definition differs slightly from an earlier one given in [1]. An extension of the definition to decision procedures is

DEFINITION 2.2. *A decision procedure $\delta: X \to D$ is minimax under loss function ℓ if $s_x(\ell, \delta(x)) \leqq s_x(\ell, \delta'(x))$ for every $x \in X$ and every $\delta': X \to D$.*

Finding a minimax decision procedure amounts to finding a minimax decision for each $x \in X$. The existence of minimax decisions is discussed in Section 4.

3. Formal properties

Several basic theorems about u_x, v_x, r_x, s_x are proved in this section. Some of the results have been obtained for finite T by Dempster [6]. Further related results, in different contexts, may also be found in Choquet [3], Huber [18], and Strassen [19]. For notational convenience, the subscript x is dropped throughout the rest of this paper.

Let ϕ be a real valued set function defined on $\mathscr{T}$. For $B, A_1, A_2, \cdots, A_p$ in $\mathscr{T}$, let

$$(3.1) \qquad \Delta_p = \phi(B) - \sum\phi(B \cup A_i) + \sum\phi(B \cup A_i \cup A_j) - \cdots + (-1)^p\phi(B \cup A_1 \cup \cdots \cup A_p),$$

and let

$$(3.2) \qquad \nabla_p = \phi(B) - \sum\phi(B \cap A_i) + \sum\phi(B \cap A_i \cap A_j) - \cdots + (-1)^p\phi(B \cap A_1 \cap \cdots \cap A_p).$$

The sums in (3.1) and (3.2) are taken over all possible distinct combinations of indices, excluding combinations that repeat indices. Following Choquet [3], we say that ϕ is alternating of order p if $\Delta_p \leqq 0$ for arbitrary $B, A_1, \cdots, A_p \in \mathscr{T}$ and is monotone of order p if $\nabla_p \geqq 0$ for arbitrary $B, A_1, \cdots, A_p \in \mathscr{T}$.

PROPOSITION 3.1. *The set function v is alternating of all orders. The set function u is monotone of all orders.*

PROOF (essentially due to Choquet). The probability $P[\cdot|E_x]$ is monotone and alternating of all orders. Now

$$v(A) = P[\psi(A)|E_x], \qquad A \in \mathscr{T}, \tag{3.3}$$

where $\psi(A) = \text{proj}_E[\xi^{-1}(x) \cap E \times A]$. If $A_1, A_2 \in \mathscr{T}$,

$$\psi(A_1 \cup A_2) = \psi(A_1) \cup \psi(A_2), \tag{3.4}$$

therefore v is alternating of all orders. The complete monotonicity of u then follows by property (c) of Proposition 3.2.

PROPOSITION 3.2. *The set functions u and v defined on $\mathscr{T}$ have the following properties:*

(a) $u(\varnothing) = v(\varnothing) = 0$;

(b) $u(T) = v(T) = 1$;

(c) $u(A) + v(\mathscr{C}A) = 1$;

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

(e) *if* $A \subset B$, $u(A) \leqq u(B)$ *and* $v(A) \leqq v(B)$;

(f) $u(A \cup B) + u(A \cap B) \geqq u(A) + u(B)$,
$v(A \cup B) + v(A \cap B) \leqq v(A) + v(B)$;

(g) *if* $A_n \downarrow A$, $u(A_n) \downarrow u(A)$, *while if* $A_n \uparrow A$, $v(A_n) \uparrow v(A)$.

PROOF. Properties (a), (b), (c), (d) are immediate from the definitions of u and v. Property (e) is equivalent to $\nabla_1 \geqq 0$ for u and $\Delta_1 \leqq 0$ for v, while property (f) is implied by $\nabla_2 \geqq 0$ for u and $\Delta_2 \leqq 0$ for v. These inequalities were established in Proposition 3.1. Finally, from (3.3), $v(A_n) = P[\psi(A_n)|E_x]$. If $A_n \uparrow A$,

$$\psi(A_n) \uparrow \bigcup_1^{\infty} \psi(A_n) = \psi\left(\bigcup_1^{\infty} A_n\right) = \psi(A). \tag{3.5}$$

This implies the second half of (g). The first half now follows from (c).

REMARK 3.1. The counterpart of (g) with u and v interchanged does not hold in general.

REMARK 3.2 Properties (c), (d) and the first property in each of (a), (b), (e), (f), (g) imply the remaining properties. All further propositions proved in this section are consequences of Proposition 3.2 alone.

PROPOSITION 3.3. *The following inequalities hold on $\mathscr{T}$. If $A \cap B = \phi$, then*

(a) $u(A) + u(B) \leqq u(A \cup B) \leqq u(A) + v(B)$,

(b) $u(A) + v(B) \leqq v(A \cup B) \leqq v(A) + v(B)$.

PROOF. The lower bound in (a) and the upper bound in (b) follow from (f) of Proposition 3.2. Since $A \cap B = \varnothing$, $B \subset \mathscr{C}A$. Therefore,

$$v(B) + v(\mathscr{C}A \cap \mathscr{C}B) = v(\mathscr{C}A \cap B) + v(\mathscr{C}A \cap \mathscr{C}B) \geqq v(\mathscr{C}A), \tag{3.6}$$

which is equivalent to

$$v(B) + 1 - u(A \cup B) \geqq 1 - u(A). \tag{3.7}$$

The upper bound in (a) holds, consequently. A dual argument establishes the lower bound in (b).

REMARK 3.3. The upper bounds in (a) and (b) are valid without the condition $A \cap B = \varnothing$.

PROPOSITION 3.4. *The following inequalities hold on* $\mathscr{T}$:

(a) $u(A \cup B) + u(A \cap B) \leqq u(A) + v(B)$,

(b) $v(A \cup B) + v(A \cap B) \geqq u(A) + v(B)$.

PROOF. Since $A \cup B = B \cup (A - B)$ and since $A = (A - B) \cup (A \cap B)$, Proposition 3.3 shows that

$$v(A \cup B) \geqq u(A - B) + v(B) \tag{3.8}$$

and

$$u(A) \leqq u(A - B) + v(A \cap B). \tag{3.9}$$

These two inequalities imply (b). Inequality (a) is proved by taking complements in (b).

PROPOSITION 3.5. *If* $A, B, C \in \mathscr{T}$ *and* $B \subset A$, *then*

(a) $v(A \cup C) - v(A) \leqq v(B \cup C) - v(B)$,

(b) $u(A \cap C) - u(A) \leqq u(B \cap C) - u(B)$.

PROOF (Choquet). If $X = B \cup C$ and $Y = A$, then $X \cup Y = A \cup C$ and $X \cap Y = B \cup (A \cap C) \supset B$. Therefore,

$$v(A \cup C) + v(B) \leqq v(X \cup Y) + v(X \cap Y) \leqq v(B \cup C) + v(A) \tag{3.10}$$

by Proposition 3.2. This establishes (a). Inequality (b) is derived by taking complements in (a).

PROPOSITION 3.6. *If the* $\{A_i\}$ *and* $\{B_i\}$ *belong to* $\mathscr{T}$, *and* $B_i \subset A_i$, *then*

(a) $v(\bigcup_1^\infty A_i) - v(\bigcup_1^\infty B_i) \leqq \Sigma_1^\infty [v(A_i) - v(B_i)]$,

(b) $u(\bigcap_1^\infty A_i) - u(\bigcap_1^\infty B_i) \leqq \Sigma_1^\infty [u(A_i) - u(B_i)]$.

PROOF. The result is established by Choquet [3] for finite unions and intersections (through induction and Proposition 3.5). Taking limits and using (g) of Proposition 3.2 completes the proof.

PROPOSITION 3.7. *For every sequence* $\{A_n\}$ *in* $\mathscr{T}$,

(a) $v(\lim_n \inf A_n) \leqq \lim_n \inf v(A_n)$,

(b) $u(\lim_n \sup A_n) \geqq \lim_n \sup u(A_n)$.

PROOF. Since $A_m \supset \inf_{n \geqq m} A_n$ and since $\inf_{n \geqq m} A_n \uparrow \lim_n \inf A_n$ as $m \to \infty$,

$$\liminf_m v(A_m) \geqq \lim_{m\to\infty} v(\inf_{n\geqq m} A_n) = v(\liminf_n A_n), \tag{3.11}$$

using (e) and (g) of Proposition 3.2. A dual argument proves (b).

REMARK 3.4. The roles of u and v cannot, in general, be interchanged in Proposition 3.7.

PROPOSITION 3.8. *Let $\mathscr{S} = \{A \in \mathscr{T} : u(A) = v(A)\}$. Then $\mathscr{S}$ is a σ-algebra and $u = v$ is a probability measure on $(T, \mathscr{S})$.*

PROOF. Clearly $\phi, T \in \mathscr{S}$. If $A \in \mathscr{S}$, then $\mathscr{C}A \in \mathscr{S}$; indeed $u(A) = v(A)$ implies $v(\mathscr{C}A) = u(\mathscr{C}A)$ by (c) of Proposition 3.2. If the $\{A_i\} \in \mathscr{S}$ and are disjoint, then $\bigcup_1^\infty A_i \in \mathscr{S}$; indeed by Proposition 3.3, for any integer $n > 1$,

$$u\left(\bigcup_1^n A_i\right) \geqq \sum_{i=1}^n u(A_i) = \sum_{i=1}^n v(A_i) \geqq v\left(\bigcup_1^n A_i\right). \tag{3.12}$$

Therefore, $u(\bigcup_1^n A_i) = v(\bigcup_1^n A_i)$. Moreover, by Propositions 3.7 and 3.2,

$$\begin{aligned} u\left(\bigcup_1^\infty A_i\right) &= u(\lim_n \bigcup_1^n A_i) \geqq \limsup_n u\left(\bigcup_1^n A_i\right) \\ &= \limsup_n v\left(\bigcup_1^n A_i\right) = v\left(\bigcup_1^\infty A_i\right), \end{aligned} \tag{3.13}$$

so that $u(\bigcup_1^\infty A_i) = v(\bigcup_1^\infty A_i)$. The fact that $u = v$ is a probability on $(T, \mathscr{S})$ follows from Propositions 3.2 and 3.3.

REMARK 3.5. This theorem links upper and lower probabilities to structural probabilities. For Fraser's models, $\mathscr{S}$ contains all Borel sets. In general, however, $\mathscr{S}$ may be trivial.

Let $\mathscr{C}$ denote the vector lattice of measurable functions mapping $(T, \mathscr{T})$ into $(R, \mathscr{R})$, the real line endowed with the σ-algebra of all Borel sets. Let $\mathscr{C}^+ = \{f \in \mathscr{C} : f \geqq 0\}$. The following definitions are abstracted from the upper and lower risks of Section 2.

DEFINITION 3.1. *If $f \in \mathscr{C}^+$, the upper integral $s(f)$ and the lower integral $r(f)$ are defined as*

$$\begin{aligned} s(f) &= \int_0^\infty v[f(t) > z]\, dz, \\ r(f) &= \int_0^\infty u[f(t) > z]\, dz. \end{aligned} \tag{3.14}$$

To extend the definitions to $f \in \mathscr{C}$, let $f^+ = f \vee 0$ and $f^- = -f \vee 0$, so that $f = f^+ - f^-$.

DEFINITION 3.2. *If $f \in \mathscr{C}$, the upper integral $s(f)$ and the lower integral $r(f)$ are defined as*

$$\begin{aligned} s(f) &= s(f^+) - r(f^-), \\ r(f) &= r(f^+) - s(f^-), \end{aligned} \tag{3.15}$$

excluding the indeterminate case $\infty - \infty$.

Definition 3.2 can also be motivated by a frequency interpreration.

PROPOSITION 3.9. *The following assertions hold for functions in* $\mathscr{C}$:

(a) *if* $r(f)$ *and* $s(f)$ *both exist, then* $r(f) \leqq s(f)$;

(b) $s(a + bf) = \begin{cases} a + bs(f) \text{ if } b \geqq 0 \text{ and } s(f) \text{ exists} \\ a + br(f) \text{ if } b \leqq 0 \text{ and } r(f) \text{ exists}; \end{cases}$

(c) $r(a + bf) = \begin{cases} a + br(f) \text{ if } b \geqq 0 \text{ and } r(f) \text{ exists} \\ a + bs(f) \text{ if } b \leqq 0 \text{ and } s(f) \text{ exists}; \end{cases}$

(d) *if* $s(f), s(g)$ *both exist and* $f \leqq g$, *then* $s(f) \leqq s(g)$;

(e) *if* $r(f), r(g)$ *both exist and* $f \leqq g$, *then* $r(f) \leqq r(g)$;

(f) *if the* $\{s(f_n)\}$ *all exist and* $r(f_n^-) < \infty$ *for at least one* n, *then* $f_n \uparrow f$ *implies that* $s(f)$ *exists and* $s(f_n) \uparrow s(f)$;

(g) *if the* $\{r(f_n)\}$ *all exist and* $r(f_n^+) < \infty$ *for at least one* n, *then* $f_n \downarrow f$ *implies that* $r(f)$ *exists and* $r(f_n) \downarrow r(f)$.

PROOF. Assertion (a) holds for $f \in \mathscr{C}^+$ by (d) of Proposition 3.2, and hence as stated. If $f \in \mathscr{C}^+$, $a \geqq 0$, and $b \geqq 0$, a change of variable in Definition 3.1 shows that

$$s(a + bf) = a + bs(f), \qquad r(a + bf) = a + br(f). \tag{3.16}$$

Therefore, if $a \geqq 0$, $b \geqq 0$, $f \in \mathscr{C}$, and $s(f)$ exists,

$$\begin{aligned} s(a + bf) &= s(a + bf^+) - r(bf^-) \\ &= a + bs(f^+) - br(f^-) = a + bs(f). \end{aligned} \tag{3.17}$$

The other cases in assertions (b) and (c) are proved similarly.

For $f, g \in \mathscr{C}^+$, assertions (d) and (e) are immediate from (e) of Proposition 3.2. If $f, g \in \mathscr{C}$, $f \leqq g$, then $f^+ \leqq g^+$, $f^- \geqq g^-$, and assertions (d) and (e) follow as stated.

To prove (f) and (g), note that if the $\{f_n\} \in \mathscr{C}^+$ and $f_n \uparrow f \in \mathscr{C}^+$, then for any $z \in R^+$,

$$\{f_n(t) > z\} \uparrow \bigcup_1^\infty \{f_n(t) > z\} = \{f(t) > z\}, \tag{3.18}$$

consequently $s(f_n) \uparrow s(f)$ by (g) of Proposition 3.2. Similarly, if $\{f_n\} \in \mathscr{C}^+$ and $f_n \downarrow f \in \mathscr{C}^+$, then $r(f_n) \downarrow r(f)$. Now suppose the $\{f_n\} \in \mathscr{C}$ and $f_n \uparrow f$. By the foregoing, $s(f_n^+) \uparrow s(f^+)$ and $r(f_n^-) \downarrow r(f^-)$. Since $r(f_n^-) < \infty$ for at least one n, $r(f) < \infty$; therefore $s(f)$ exists and (f) follows. Assertion (g) is proved analogously.

REMARK 3.6. In general, the roles of r and s cannot be interchanged in (f) and (g) above.

PROPOSITION 3.10. *Let* $f, g \in \mathscr{C}$.

(a) *If either* $s(f^+) + s(g^+) < \infty$ *or* $s(f^-) + r(g^-) < \infty$, *then* $r(f) + s(g) \leqq s(f \vee g) + s(f \wedge g) \leqq s(f) + s(g)$.

(b) *If either* $s(f^-) + s(g^-) < \infty$ *or* $r(f^+) + s(g^+) < \infty$, *then* $r(f) + r(g) \leqq r(f \vee g) + r(f \wedge g) \leqq r(f) + s(g)$.

PROOF. Let $f, g \in \mathscr{C}^+$. Since for any $z \in R^+$,

$$\begin{aligned} \{f(t) \vee g(t) > z\} &= \{f(t) > z\} \cup \{g(t) > z\}, \\ \{f(t) \wedge g(t) > z\} &\doteq \{f(t) > z\} \cap \{g(t) > z\}, \end{aligned} \tag{3.19}$$

inequalities (a) and (b) follow from Propositions 3.2 and 3.4. If $f, g \in \mathscr{C}$, then

$$\begin{aligned} (f \vee g)^+ &= f^+ \vee g^+, \quad (f \wedge g)^+ = f^+ \wedge g^+, \\ (f \vee g)^- &= f^- \vee g^-, \quad (f \wedge g)^- = f^- \wedge g^-, \end{aligned} \tag{3.20}$$

and the proposition follows from the results on $\mathscr{C}^+$.

PROPOSITION 3.11. *Let* $\{f_n\}$ *be a sequence of functions in* $\mathscr{C}$.

(a) *If* $g \in \mathscr{C}$, $r(g^-) < \infty$, *and* $f_n \geqq g$ *for all* n, *then* $s(\lim_n \inf f_n) \leqq \lim_n \inf s(f_n)$.

(b) *If* $g \in \mathscr{C}$, $r(g^+) < \infty$, *and* $f_n \leqq g$ *for all* n, *then* $r(\lim_n \sup f_n) \geqq \lim_n \sup r(f_n)$.

PROOF. In (a), since $\inf_{m \geqq n} f_m \geqq g$, $r([\inf_{m \geqq n} f_m]^-) \leqq r(g^-) < \infty$ for all n, and therefore $s(\inf_{m \geqq n} f_m)$ exists for all n. Similarly, $r(f_m^-) < \infty$ for all m, so that $s(f_m)$ exists for all m. By Proposition 3.9, $s(\lim_n \inf f_n)$ exists and as $n \to \infty$,

$$\inf_{m \geqq n} s(f_m) \geqq s(\inf_{m \geqq n} f_m) \uparrow s(\lim_n \inf f_n), \tag{3.21}$$

which proves (a). A dual argument establishes (b).

Let $\{A_n : A_n \in \mathscr{T}, n = 0, \pm 1, \pm 2, \cdots\}$ be a countable partition of T, and let $\mathscr{A}$ denote the σ-algebra generated by this partition. If $B \in \mathscr{A}$, $B = \cup_I A_i$, where I is countable. Define a set function q on $\mathscr{A}$ as follows:

$$q(A_j) = v\left(\bigcup_{i=j}^{\infty} A_i\right) - v\left(\bigcup_{i=j+1}^{\infty} A_i\right), \qquad j = 0, \pm 1, \cdots. \tag{3.22}$$

More generally, if $B \in \mathscr{A}$, $B = \cup_I A_i$, define $q(B)$ by

$$q(B) = \sum_I q(A_i). \tag{3.23}$$

LEMMA 3.1. *If* $\lim_{n \to \infty} v(\cup_{i=n}^{\infty} A_i) = 0$, *then for every* $B \in \mathscr{A}$, $u(B) \leqq q(B) \leqq v(B)$, *and* q *is a probability measure on* $\mathscr{A}$.

PROOF. To verify that q is a probability on $\mathscr{A}$, note first that q is countably additive by definition. Since v is monotone, $q(A_j) \geqq 0$ for all j and hence $q(B) \geqq 0$ for $B \in \mathscr{A}$. Also

$$(3.24)\qquad q(T) = \sum_{i=-\infty}^{\infty} q(A_i) = \lim_{m,n\to\infty} \sum_{i=-m}^{n} q(A_i)$$
$$= \lim_{m,n\to\infty} \left[v\left(\bigcup_{i=-m}^{\infty} A_i \right) - v\left(\bigcup_{i=n+1}^{\infty} A_i \right) \right] = 1,$$

by Proposition 3.2 and the hypothesis of the lemma.

From Proposition 3.3, applied to (3.22),

$$(3.25)\qquad q(A_j) \leqq v(A_j), \qquad j = 0, \pm 1, \pm 2, \cdots .$$

Let $C = \cup_J A_j$, with J a finite set of natural numbers, and suppose that $q(C) \leqq v(C)$. If k is a natural number smaller than any element of J,

$$(3.26)\qquad q(A_k \cup C) = q(A_k) + q(C)$$
$$\leqq v\left(\bigcup_{i=k}^{\infty} A_i \right) - v\left(\bigcup_{i=k+1}^{\infty} A_i \right) + v(C)$$
$$\leqq v(A_k \cup C),$$

the last inequality coming from (f) of Proposition 3.2. Starting with (3.25) and applying (3.26) a finite number of times shows that for any finite J,

$$(3.27)\qquad q(\bigcup_J A_j) \leqq v(\bigcup_J A_j).$$

Taking limits establishes the inequality $q(B) \leqq v(B)$ for any $B \in \mathscr{A}$. Finally, if $B \in \mathscr{A}$, then $\mathscr{C}B \in \mathscr{A}$ and $q(\mathscr{C}B) \leqq v(\mathscr{C}B)$, hence by Proposition 3.2, $u(B) \leqq q(B)$.

A function $f \in \mathscr{C}$ is elementary if it can be represented in the form

$$(3.28)\qquad f(t) = \sum_{j=-\infty}^{\infty} aI_{A_j}(t),$$

where $\{A_n : A_n \in \mathscr{T}, n = 0, \pm 1, \pm 2, \cdots\}$ is a partition of T when repetitions are excluded, $a_0 = 0$, and $a_{j+1} - a_j \geqq \delta > 0$ for each j and some δ. If all but a finite number of the $\{A_n\}$ equal $\varnothing$, then f is a simple function.

Lemma 3.2. *If $f \in \mathscr{C}$ is elementary, with representation* (3.28), *and if* $|s(f)| < \infty$, *then*

(a) $\lim_{n\to\infty} a_n\, v(\cup_{i=n}^{\infty} A_i) = 0$, $\lim_{n\to-\infty} a_n\, u(\cup_{i=n}^{-\infty} A_i) = 0$,

(b) $s(f) = \Sigma_{j=-\infty}^{\infty} a_j q(A_j)$.

Proof. Under the hypotheses of the lemma,

$$(3.29)\qquad f^+(t) = \sum_{j=1}^{\infty} a_j I_{A_j}(t), \qquad f^-(t) = \sum_{j=-1}^{-\infty} - a_j I_{A_j}(t),$$

and $s(f^+) < \infty$, $r(f^-) < \infty$. From Definition 3.1,

$$(3.30)\qquad s(f^+) = \sum_{j=1}^{\infty} (a_j - a_{j-1})\, v\left(\bigcup_{i=j}^{\infty} A_i \right)$$
$$= \lim_{n\to\infty} \left[\sum_{j=1}^{n} a_j q(A_j) + a_n v\left(\bigcup_{i=n+1}^{\infty} A_i \right) \right].$$

Therefore, since $s(f^+) < \infty$,

$$\lim_{n \to \infty} (a_n - a_{n-1})\, v\left(\bigcup_{i=n}^{\infty} A_i\right) = 0, \tag{3.31}$$

and since $a_n v(\bigcup_{i=n+1}^{\infty} A_i) \geqq 0$,

$$\lim_{n \to \infty} \sum_{j=1}^{n} a_j q(A_j) \leqq s(f^+) < \infty. \tag{3.32}$$

Moreover, because of (3.31), $\lim_{n \to \infty} v(\bigcup_{i=n}^{\infty} A_i) = 0$ and

$$a_n v\left(\bigcup_{i=n}^{\infty} A_i\right) = a_n \sum_{j=n}^{\infty} q(A_j) \leqq \sum_{j=n}^{\infty} a_j q(A_j). \tag{3.33}$$

From these relations follow the first part of (a) and

$$s(f^+) = \sum_{j=1}^{\infty} a_j q(A_j). \tag{3.34}$$

A similar argument on $r(f^-)$ establishes the other part of (a) and

$$r(f^-) = -\sum_{j=-1}^{-\infty} a_j \left[u\left(\bigcup_{i=j}^{-\infty} A_i\right) - u\left(\bigcup_{i=j-1}^{-\infty} A_i\right)\right] = -\sum_{j=-1}^{-\infty} a_j q(A_j), \tag{3.35}$$

the second equality coming from Proposition 3.2. Finally, (b) is a consequence of (3.34) and (3.35).

REMARK 3.7. From Lemma 3.1, the set function q appearing in Lemma 3.2 is a probability. Thus, (b) represents $s(f)$ as an expectation.

PROPOSITION 3.12. *Let $f, g \in \mathscr{C}$ be such that $f + g$ is defined.*

(a) *If either $s(f^+) + s(g^+) < \infty$ or $s(f^-) + r(g^-) < \infty$, then $r(f) + s(g) \leqq s(f + g) \leqq s(f) + s(g)$.*

(b) *If either $s(f^-) + s(g^-) < \infty$ or $r(f^+) + s(g^+) < \infty$, then $r(f) + r(g) \leqq r(f + g) \leqq r(f) + s(g)$.*

PROOF. (i) Let $f, g \in \mathscr{C}$ be elementary functions to which the hypotheses of (a) apply. Then $s(f)$, $s(g)$ and $r(f)$ exist. Assume that $|s(f + g)| < \infty$. The sum $f + g$ may be represented in the form (3.28). If $e(\cdot)$ denotes expectation with respect to q, then by Remark 3.7 and the preceding lemmas,

$$s(f + g) = e(f + g) = e(f) + e(g) \leqq s(f) + s(g). \tag{3.36}$$

(ii) If $f, g \in \mathscr{C}$, each may be approximated from below by a monotone increasing sequence of elementary functions. Under the hypotheses of (a) and if $|s(f + g)| < \infty$, the result of (i) applies to approximating elementary functions. Taking monotone limits establishes

$$s(f + g) \leqq s(f) + s(g). \tag{3.37}$$

Special cases. If $s(f+g) = -\infty$ and the hypotheses of (a) hold, then (3.37) is trivial. If $s(f+g) = \infty$, then $s(f^+ + g^+) \geqq s[(f+g)^+] = \infty$. Since f^+, $g^+ \in \mathscr{C}^+$, each may be approximated from below by a monotone increasing sequence of simple functions, each of which is in $\mathscr{C}^+$ and is bounded. The result in (i) for elementary functions applies; taking monotone limits shows that $s(f^+) + s(g^+) \geqq s(f^+ + g^+) = \infty$. Thus, one of $s(f)$, $s(g)$ is ∞ and (3.37) is valid.

In summary, therefore, if $s(f+g)$ exists and the hypotheses of (a) hold, then (3.37) is valid. Under the same assumptions,

$$s(g) = s(f + g - f) \leqq s(f+g) + s(-f), \tag{3.38}$$

which, by Proposition 3.9, is equivalent to the left inequality in (a).

(iii) Suppose f, $g \in \mathscr{C}$ and the hypotheses of (b) hold, ensuring that $r(f)$, $r(g)$, $s(g)$ exist. Assume also that $r(f+g)$ exists. Since $r(f+g) = -s(-f-g)$, the inequalities of (b) follow from (ii).

(iv) To complete the proof, it is necessary to show that $s(f+g)$, $r(f+g)$ exist under the hypotheses of (a) and (b), respectively. Since $s(f^+ + g^+)$, $r(f^- + g^-)$ exist, it follows from (ii) and (iii) that

$$\begin{aligned} s[(f+g)^+] &\leqq s(f^+ + g^+) \leqq s(f^+) + s(g^+), \\ r[(f+g)^-] &\leqq r(f^- + g^-) \leqq s(f^-) + r(g^-). \end{aligned} \tag{3.39}$$

Thus $s(f+g)$ exists under the hypotheses of (a). A dual argument shows that $r(f+g)$ exists in (b).

COROLLARY 3.1. *Let* $f \in \mathscr{C}$.

(a) *If* $s(f)$ *exists*, $|s(f)| \leqq s(|f|)$.

(b) *If* $r(f)$ *exists*, $|r(f)| \leqq s(|f|)$.

Define sets K, $K^+ \subset R^n$ as follows:

$$\begin{aligned} K &= \{x \in R^n : x_1 \geqq 0, x_2 \geqq 0, \cdots, x_n \geqq 0\}, \\ K^+ &= \{x \in R^n : x_1 > 0, x_2 > 0, \cdots, x_n > 0\}. \end{aligned} \tag{3.40}$$

PROPOSITION 3.13. *Let* $h: K \to R$ *be continuous and concave in* K *and such that* $h(x) > 0$ *and* $h(\lambda x) = \lambda h(x)$ *for every* $x \in K^+$ *and* $\lambda \geqq 0$. *Let* $f_1, f_2, \cdots, f_n \in \mathscr{C}^+$ *be such that* $s(f_i) < \infty$ *for* $1 \leqq i \leqq n$. *Then*

$$s[h(f_1(t), \cdots, f_n(t))] \leqq h[s(f_1), \cdots, s(f_n)]. \tag{3.41}$$

PROOF. By Propositions 3.9 and 3.12, $s(\cdot)$ is an increasing gauge on $\mathscr{C}^+$. The theorem follows from a general result due to Bourbaki (see Berge [2], p. 212).

If $f \in \mathscr{C}$, define $\|f\|_p$ by

$$\begin{aligned} \|f\|_p &= [s(|f|^p)]^{1/p}, \qquad 1 \leqq p < \infty, \\ \|f\|_\infty &= \sup\{z : v(|f(t)| > z) > 0\}. \end{aligned} \tag{3.42}$$

PROPOSITION 3.14. *Let $f, g \in \mathscr{C}$. Then*

(a) $0 \leqq \|f\|_p \leqq \|f\|_q$ *if* $1 \leqq p \leqq q \leqq \infty$,

(b) $v(|f(t)| \neq 0) = 0$ *if and only if* $\|f\|_p =$ *for* $1 \leqq p \leqq \infty$,

(c) $\|af\|_p = a\|f\|_p$ *if* $a \in R^+$ and $1 \leqq p \leqq \infty$,

(d) $\|fg\|_r \leqq \|f\|_p \|g\|_q$ *if* $1 \leqq p, q, r \leqq \infty$ and $r^{-1} = p^{-1} + q^{-1}$,

(e) $\|f + g\|_p \leqq \|f\|_p + \|g\|_p$ *if* $1 \leqq p \leqq \infty$.

PROOF. Assertions (b) and (c) are immediate. Apart from special cases, (d) and (e) follow from Proposition 3.13, or may be proved from the Hölder and Minkowski inequalities along the lines of Proposition 3.12. Assertion (a) is a consequence of (d).

4. Minimax decisions

Conditions for the existence of minimax decisions, as defined in Section 2, are provided by the theorem below. Examples of minimax procedures for a distribution free estimation problem follow.

PROPOSITION 4.1. *Let T, D be compact metric spaces and let $\ell : T \times D \to R$ be continuous. Then*

(a) *$s(\ell, d)$ and $r(\ell, d)$ are uniformly continuous on D,*

(b) *the suprema and infima over D of $r(\ell, d)$ are attained.*

PROOF. Let $m(\cdot\,, \cdot)$ denote the metric on D. Since $T \times D$ is compact metric, $\ell(t, d)$ is uniformly continuous on $T \times D$. Therefore, to every $\varepsilon > 0$ there corresponds an $\eta > 0$ such that

$$m(d, d') < \eta \Rightarrow |\ell(t, d) - \ell(t, d')| < \varepsilon \tag{4.1}$$

for every $t \in T$. Applying Proposition 3.9, parts (b), (c), (d), (e), to the right side of (4.1) establishes

$$|s(\ell, d) - s(\ell, d')| < \varepsilon, \qquad |r(\ell, d) - r(\ell, d')| < \varepsilon, \tag{4.2}$$

hence (a) and (b).

EXAMPLE. An example of the statistical model described in Section 2 is the nonparametric version of the two sample location shift model. If $(x_1, \cdots, x_m)$ are the observations of the first sample and $(y_1, \cdots, y_n)$ are the observations of the second sample, the model can be written in the form

$$\begin{aligned} x_i &= F^{-1}(u_i), && 1 \leqq i \leqq m, \\ y_j &= \mu + F^{-1}(u_{m+j}), && 1 \leqq j \leqq n, \end{aligned} \tag{4.3}$$

where $(u_1, \cdots, u_{m+n})$ are realizations of independent, identically distributed random variables, each uniformly distributed on $[0, 1]$, $F \in \mathscr{F}$, the class of all continuous distribution functions on the real line, $\mu \in \Omega = (-\infty, \infty)$, and (μ, F) is the unknown parameter. Equations (4.3) are of the general form (2.1).

Let $\{d_{i,j} = y_j - x_i, 1 \leqq i \leqq m, 1 \leqq j \leqq n\}$ and let $a_1 < a_2 < \cdots < a_{M-1}$, where $M = mn + 1$, denote the ordered $\{d_{i,j}\}$. Under the original model, the strict ordering will be possible with probability one. Let $\Omega_1 = (-\infty, a_1)$, let $\Omega_i = (a_{i-1}, a_i)$ for $2 \leqq i \leqq M - 1$, and let $\Omega_M = (a_{M-1}, \infty)$. For arbitrary $A \subset \Omega$, define

$$\delta_v(i, A) = \begin{cases} 1 & \text{if } A \cap \Omega_i \neq \varnothing \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

and

$$\delta_u(i, A) = \begin{cases} 1 & \text{if } \Omega_i \subset A \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

Then, as shown in [1], for arbitrary $A \subset \Omega$,

$$\begin{aligned} v(A \times \mathscr{F}) &= \frac{1}{M} \sum_{i=1}^{M} \delta_v(i, A), \\ u(A \times \mathscr{F}) &= \frac{1}{M} \sum_{i=1}^{M} \delta_u(i, A). \end{aligned} \tag{4.6}$$

This collection of upper and lower probabilities determines the upper and lower risks if the loss function does not depend upon F. For example, suppose that it is desired to estimate μ, that $h: R^+ \to R^+$ is strictly monotone increasing, and that the loss function of interest is

$$\ell(\mu, d) = \begin{cases} h(|\mu - d|) & \text{if } |\mu - d| \leqq c \\ h(c) & \text{if } |\mu - d| > c, \end{cases} \tag{4.7}$$

where $c > a_{M-1} - a_1$. Let $B_1 = [a_{M-1} - c, \frac{1}{2}(a_1 + a_2)]$, let $B_i = [\frac{1}{2}(a_{i-1} + a_i), \frac{1}{2}(a_i + a_{i+1})]$ for $2 \leqq i \leqq M - 2$, and let $B_{M-1} = [\frac{1}{2}(a_{M-2} + a_{M-1}), a_1 + c]$. Then for ℓ defined by (4.7),

$$s(\ell, d) = \frac{1}{M}\Big[\sum_{j \neq i} h(|a_j - d|) + 2h(c)\Big] \tag{4.8}$$

if $d \in B_i$ for $1 \leqq i \leqq M - 1$, and

$$\begin{aligned} s(\ell, d) &> s(\ell, a_{M-1} - c) && \text{if } d < a_{M-1} - c, \\ s(\ell, d) &> s(\ell, a_1 + c) && \text{if } d > a_1 + c. \end{aligned} \tag{4.9}$$

Similar expressions may be found for $r(\ell, d)$.

In particular, suppose that $h(x) = x$. Then, if M is even, $s(\ell, d)$ is minimized by any $d \in [\frac{1}{2}(a_{(M/2)-1} + a_{M/2}), \frac{1}{2}(a_{M/2} + a_{(M/2)+1})]$, while if M is odd, the minimizing value is $d = \frac{1}{2}(a_{(M-1)/2} + a_{[(M-1)/2]+1})$. This class of minimax estimates for μ includes the Hodges-Lehmann estimate median $\{a_1, \cdots, a_{M-1}\}$.

If $h(x) = x^2$, the minimax estimate for μ can be described as follows. Let $m_i = M^{-1} \sum_{j \neq i} a_j$. If there exists a k, $1 \leqq k \leqq M - 1$, such that $m_k \in B_k$, $s(\ell, d)$ is minimized by $d = m_k$. Otherwise, there will exist a k, $1 \leqq k \leqq M - 1$

such that $m_k > \frac{1}{2}(a_k + a_{k+1}) > m_{k+1}$; in this event $s(\ell, d)$ is minimized by $d = \frac{1}{2}(a_k + a_{k+1})$. Viewed as functions of $(x_1, \cdots, x_m, y_1, \cdots, y_n)$, these minimax decisions are minimax procedures in the sense of Definition 2.2.

REFERENCES

[1] R. J. BERAN. "On distribution-free statistical inference with upper and lower probabilities," *Ann. Math. Statist.*, Vol. 42 (1971), pp. 157–168.

[2] C. BERGE, *Topological Spaces*, Edinburgh and London, Oliver and Boyd, 1963.

[3] G. CHOQUET, "Theory of capacities," *Ann. Inst. Fourier. (Grenoble)*, Vol. 5 (1953/54), pp. 131–295.

[4] A. P. DEMPSTER, "On direct probabilities," *J. Roy. Statist. Soc. Ser. B.*, Vol. 25 (1963), pp. 102–107.

[5] ———, "New methods for reasoning toward posterior distributions based on sample data," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 355–374.

[6] ———, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 325–339.

[7] ———, "Upper and lower probability inferences based on a sample from a finite univariate population," *Biometrika*, Vol. 54 (1967), pp. 515–528.

[8] ———, "Upper and lower probabilities generated by a random closed interval," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 957–966.

[9] ———, "A generalization of Bayesian inference," *J. Roy. Statist. Soc. Ser. B.*, Vol. 30 (1968), pp. 205–247.

[10] ———, "Upper and lower probability inferences for families of hypotheses with monotone density ratios," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 953–969.

[11] D. A. S. FRASER, "The fiducial method and invariance," *Biometrika*, Vol. 48 (1961), pp. 261–280.

[12] ———, "On fiducial inference," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 661–676.

[13] ———, "On the consistency of the fiducial method," *J. Roy. Statist. Soc. Ser. B.*, Vol. 24 (1962), pp. 425–434.

[14] ———, "Structural probability and a generalization," *Biometrika*, Vol. 53 (1966), pp. 1–9.

[15] ———, "Data transformations and the linear model," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 1456–65.

[16] ———, *The Structure of Inference*, New York, Wiley, 1968.

[17] I. HACKING, *Logic of Statistical Inference*, Cambridge, Cambridge University Press, 1965.

[18] P. J. HUBER, *Théorie de l'Inférence Statistique Robuste*, Montréal, Les Presses de l'Université de Montréal, 1969.

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