The Annals of Statistics
1976, Vol. 4, No. 5, 836-865

ADMISSIBILITY OF CONDITIONAL CONFIDENCE PROCEDURES

BY J. KIEFER[1]

Cornell University

A formal structure for conditional confidence (cc) procedures is investigated. Underlying principles are a (conditional) frequentist interpretation of the cc coefficient Γ, and highly variable Γ. The latter allows the stated measure of conclusiveness to reflect how intuitively clear-cut the outcome of the experiment is. The methodology may thus answer some criticisms of the Neyman-Pearson-Wald approach, but is in the spirit of the latter and includes it. Example: X has one of k densities f_ω wrt ν. A nonempty set of decisions $D_\omega \subset D$ is "correct" for state ω. A nonrandomized cc procedure consists of a pair (δ, Z) where δ is a nonrandomized decision function and Z is a conditioning rv. The cc coefficient is $\Gamma_\omega = P_\omega\{\delta^{-1}(D_\omega) \mid Z\}$. If $X = x_0$, we make decision $\delta(x_0)$ with "cc $\Gamma_\omega(x_0)$ of being correct if ω is true"; it is unnecessary, but often a practical convenience (as for un-cc intervals), to have Γ_ω independent of ω. Possible notions of "goodness" are discussed; e.g., $(\bar{\bar{\delta}}, \bar{\bar{Z}})$ at least as good as $(\bar{\delta}, \bar{Z})$ if $P_\omega\{\bar{\bar{\delta}}(X) \in D_\omega$ and $\bar{\bar{\Gamma}}_\omega > t\} \geqq P_\omega\{\bar{\delta}(X) \in D_\omega$ and $\bar{\Gamma}_\omega > t\}$ $\forall\, t, \omega$, and $P_\omega\{\bar{\bar{\Gamma}}_\omega = 0\} \leqq P_\omega\{\bar{\Gamma}_\omega = 0\}$ $\forall\, \omega$. It is proved that cc procedure (δ, Z) is admissible if the non-cc δ is admissible. For 2-hypothesis problems the converse is true; otherwise, "star-shaped" partitions of the likelihood ratio space are needed. Other loss structures are also treated.

0. Introduction. Although there is a large literature of conditioning in statistical inference, there has been no methodical presentation of a frequentist non-Bayesian framework that considers possible criteria of goodness of such procedures, and methods for constructing them, in general statistical settings. The present paper is devoted to decision-theoretic admissibility considerations, in such a framework.

Various illustrations of this approach are contained in [4] and [2]. Its relationship with other work on conditioning, and some discussion of foundations, will be found in [2]. Still, it may be appropriate here to indicate that the present approach was motivated by the observation that many critics of the Neyman-Pearson-Wald (NPW) approach to statistics seem disturbed at the prospect of making a decision that is not accompanied by some data-dependent measure of "conclusiveness" of the experimental outcome. For example, in deciding whether a normal rv X with unit variance has mean -1 or $+1$, where the standard symmetric NP test makes a decision accompanied only by the assessment of error probabilities $\Phi(-1)$, some statisticians may be disturbed by an

Received April 1975; revised February 1976.

[1] Research under NSF Grant MPS 72-04998 A02.

AMS 1970 *subject classifications.* 62C07, 62C15.

Key words and phrases. Conditional inference, conditional confidence coefficients, admissibility, complete classes, canonical procedure.

intuitive feeling that they are much surer of the conclusion when $X = 10$ than when $X = .5$. The authors of various axiomatic studies of statistical foundations (Bayesian, likelihood, fiducial, evidential, etc.) have criticized other aspects of the NP approach as well; but this notion of wanting a measure of conclusiveness that depends on the experimental outcome has appealed also to many practitioners, as is evidenced by the old and continued usage of the methodology that states the level at which a singificance test would "just reject" a null hypothesis (and which is historically a starting point for an extensive theory constructed by Bahadur).

A principle adopted in our developing of a methodology that may satisfy the objection mentioned in the previous paragraph, is that our measure of conclusiveness should have a frequentist (law of large numbers) meaning similar to that emphasized by Neyman for classical NP tests and confidence intervals. As discussed in [2], this implies that, except in rare cases of symmetry, our approach and form of conclusion cannot agree with those of the critics mentioned above.

The methodology we propose is a procedure consisting of a decision rule and a conditioning random variable. The frequentist properties of such a procedure are studied in terms of the conditional probability of a correct decision given the conditioning random variable. This quantity is called the "conditional confidence," extending the usage of "confidence" from its traditional meaning in unconditional interval estimation to the present setting, where it has an analogous frequentist interpretation.

In conformity with our motivating comments, this conditional confidence coefficient should in practice be highly variable as a function of the conditioning random variable; otherwise, we may as well use an unconditional NPW procedure. There is a brief discussion of such variability near the end of Section 2. The basic notions are introduced in Section 1; the ideas are illustrated in examples in Section 1.3. In Section 2 some possible admissibility criteria are defined. The simplest of these regards a procedure as inadmissible if there is another procedure with smaller probability of yielding a zero conditional confidence coefficient, and with stochastically larger coefficient on the set where a correct decision is made.

In Section 3, sufficient conditions for admissibility are obtained. It turns out, perhaps surprisingly, that if a NPW unconditional procedure is admissible in a classical sense, then every conditioning used in conjunction with it yields a procedure admissible in the ordering considered here. Just as in classical decision theory, additional criteria are needed to choose among these; some such criteria are discussed in Section 2 and, more extensively, in [2].

Section 5 contains complete class theorems for the k-hypothesis problem. The result is especially simple when $k = 2$, where the minimal complete class coincides with that obtained in Section 3; for larger k, additional procedures are included. In leading up to these results, Section 4 studies special "canonical" procedures in terms of which closure results are proved. Section 6 discusses some other possible definitions of admissibility.

1. Notation, definitions, examples.

1.1. *Basic ideas and notation.* It is convenient to assume the underlying measure space of possible outcomes of the experiment, $(\mathscr{X}, \mathscr{B}, \nu)$, to be of countable type with compactly generated σ-finite ν with respect to which the possible states of nature $\Omega = \{\omega\}$ have densities f_ω. This implies existence of regular conditional probabilities in the sequel. (See, e.g., [6], pages 193–194.) We write X for the rv representing the experiment with outcome in $\mathscr{X}$. By I_A we denote the characteristic function of the subset A of the domain of this function; in particular, it may be the indicator of the event A.

In the decision space D, we assume that for each ω there is specified a nonempty subset D_ω of decisions that are "correct" when ω is true. We write $\Omega_d = \{\omega : d \in D_\omega\}$. The confidence "flavor" is best exhibited in terms of the simple loss function implied by consideration in these terms; other possible treatments will be described in Section 6.

A (nonrandomized) decision rule $\delta : \mathscr{X} \to D$ is required to have $\delta^{-1}(D_\omega) \in \mathscr{B}$ for each ω. (This may easily be translated in terms of a structure on D.) The apparent neglect of randomization is intended to aid in clarity of exposition. It is not a genuine neglect, since, as will be discussed presently, $\mathscr{X}$ will be regarded as the product of a more primitive sample space $\mathscr{X}'$ and a randomization space.

Any subfield $\mathscr{B}_0$ of $\mathscr{B}$ may be called a *conditioning subfield.* It is convenient to consider only those subfields generated by statistics on $(\mathscr{X}, \mathscr{B})$. If Z is such a statistic, with $Z(\mathscr{X}) = B$,

$$\mathscr{B}_0 = \{S : S \in \mathscr{B}, S = Z^{-1}(A) \text{ for some } A \subset B\}.$$

The relationship between $\mathscr{B}_0$ and the partition of $\mathscr{X}$ induced by Z is discussed in [1]. We denote the resulting (conditioning) partition of $\mathscr{X}$ by $\{C^b, b \in B\}$. We also write, for a given δ and $\mathscr{B}_0$,

$$\begin{aligned} C_d{}^b &= C^b \cap \delta^{-1}(d), \\ C_\omega &= \delta^{-1}(D_\omega) = \{x : \delta(x) \text{ is correct when } \omega \text{ is true}\}, \\ K_\omega{}^b &= C^b \cap C_\omega. \end{aligned} \tag{1.1}$$

The $K_\omega{}^b$ are not necessarily disjoint, but the $C_d{}^b$ $(b \in B, d \in D)$ are, and constitute the *partition* C of $\mathscr{X}$. Note that subscripts always index D or Ω; superscripts index B; symbols such as overbars or left superscripts distinguish different partitions and their operating characteristics.

A *conditional confidence procedure* is a pair $(\delta, \mathscr{B}_0)$ or (δ, Z) or, equivalently, the corresponding partition C. Its associated *conditional confidence function* is the set $\mathbf{\Gamma} = \{\Gamma_\omega, \omega \in \Omega\}$ of conditional probabilities

$$\Gamma_\omega = P_\omega\{C_\omega \mid \mathscr{B}_0\}, \qquad \omega \in \Omega. \tag{1.2}$$

For each ω, this $\mathscr{B}_0$-measurable function on $\mathscr{X}$ can be regarded as a function of Z, and we write $\Gamma_\omega{}^b$ for its value on the set $Z = b$. A *conditional confidence statement* associated with $(\delta, \mathscr{B}_0)$ is the pair (δ, Γ). It is used as follows: if $X = x_0$, and $Z(x_0) = z_0$, we state that "for each ω, we have confidence $\Gamma_\omega(x_0) = \Gamma_\omega{}^{z_0}$ of being correct if ω is true." We make decision $\delta(x_0)$.

Such statements have a conditional frequentist interpretation analogous to that of the NPW setting $\mathscr{B}_0 = \{\mathscr{X}, \emptyset\}$, where Γ_ω is simply the probability of a correct decision when ω is true. In a sequence of n independent experiments, if ω_i is true and X_i is observed in the ith experiment, then the proportion of correct decisions made will be close to $n^{-1} \sum_1^n \Gamma_{\omega_i}(X_i)$ with probability near one when n is large. This is true even if different $(\delta, \mathscr{B}_0)$'s are used in the experiments, but the frequentist meaning is of most intuitive value in cases where the dependence of Γ_ω on ω can be removed, just as in the usual NP treatment of confidence intervals or composite hypotheses with specified minimum power. Thus, if there is a function ψ on $\mathscr{X}$ such that $\Gamma_\omega(x) \geqq \psi(x)$ for all ω and all x, we may make the more succinct statement that "we have conditional confidence at least $\psi(x_0)$ that $\delta(x_0)$ is a correct decision," if $X = x_0$. The frequentist interpretation in terms of the law of large numbers is thereby simplified, and its practical meaning is made clearer if we look only at those experiments in which $\Gamma(X_i) \geqq .9$ (for example): a correct decision is very likely made in about 90% or more of such experiments, if their number is large. Whenever there is a $\mathscr{B}_0$-measurable set A of positive ν measure on which $\Gamma_\omega{}^{Z(x)}$ can be chosen not to depend on ω, we write Γ^b in place of $\Gamma_\omega{}^b$ for b in $Z(A)$.

The class of all conditional confidence procedures C is denoted by $\mathscr{C}$. Additional notation will be introduced as required.

1.2. *More about conditioning partitions.* In Section 2 we shall discuss goodness criteria for partitions. One element of those considerations is the set

$$Q_\omega = \{x : \Gamma_\omega(x) = 0\} . \tag{1.3}$$

This is a (f_ω) maximal $\mathscr{B}_0$-measurable set on which we *never* make a confidence statement correct for ω. A procedure is called nondegenerate if $P_\omega\{Q_\omega\} = 0$ $\forall\, \omega$, and the class of all nondegenerate procedures is denoted by $\mathscr{C}^+$. The intuitive appeal of using a nondegenerate procedure can be seen in the NP setting of testing between two simple hypotheses whose densities are positive throughout $\mathscr{X}$: it amounts there to prohibiting use of the trivial critical regions, $\mathscr{X}$ and $\emptyset$.

We denote by $\mathscr{C}^{B_0}$ all partitions with $B = B_0$, and write $\mathscr{C}^{B+} = \mathscr{C}^B \cap \mathscr{C}^+$. For positive integral L we denote by B^L the finite label set $\{1, 2, \cdots, L\}$ and abbreviate $\mathscr{C}^{B^L}$ by $\mathscr{C}^L$. We have found it convenient not to demand essential minimality of B for a given partition, but instead to permit a procedure in $\mathscr{C}^L$ to have $\nu(C^b) = 0$ for some b; thus, $\mathscr{C}^{L-1}$ can be regarded as essentially a subset of $\mathscr{C}^L$. However, the *nondegenerate partitions* $\mathscr{C}^{L+}$ are disjoint. Between $\mathscr{C}^L$ and $\mathscr{C}^{L+}$ are the *proper partitions* of $\mathscr{C}^L$, those for which each C^b has positive probability for some ω.

Thus, $\mathscr{C}^1$ is the classical NPW case, the Wald framework of unconditional statistical decision theory. It is a useful reference mark against which to compare the various notions we use.

For finite (or denumerably infinite) B, we interpret (1.2) through the simple conditional probability formula: if $P_\omega\{C^b\} > 0$,

$$\Gamma_\omega{}^b = P_\omega\{K_\omega{}^b\}/P_\omega\{C^b\}\,, \tag{1.4}$$

and Q_ω is the union of all C^b's where $P_\omega\{K_\omega{}^b\} = 0$. We shall see that other B's are also of interest, for example, $B =$ a real interval.

1.3. *Examples.* We now illustrate the notions introduced above in terms of a simple 2-*hypothesis setting*, where $\Omega = D = \{1, 2\}$ and $D_i = \{i\}$. To avoid questions of randomization, suppose ν is Lebesgue measure on a real interval $\mathscr{X}$ and that $\nu(\{x : f_2(x)/f_1(x) = c\}) = 0$ for every c. Then Corollary 5.4, together with the comments of Section 4 on elimination of randomization, assert that every admissible procedure (in a sense described in Section 2) is obtained essentially by subdividing the $\{C_1, C_2\}$ of a NP $\mathscr{C}^1$-partition into arbitrary $C_i{}^b$. For an example of computational simplicity, suppose f_i is normal with mean $2i$ and variance 1. The symmetric NP $\mathscr{C}^1$ partition $C_1 = (-\infty, 3)$, $C_2 = [3, \infty)$ has $\Gamma = \Phi(1) = .84$. Using the same C_i with $C^1 = \{x : |x - 3| < 1.5\} = \mathscr{X} - C^2$, we obtain a symmetric $\mathscr{C}^2$ partition with $\Gamma^1 = [\Phi(1) - \Phi(-.5)]/[\Phi(2.5) - \Phi(-.5)] = .78$ and with $\Gamma^2 = \Phi(-.5)/[\Phi(-.5) + \Phi(-2.5)] = .98$. Out of the probability .84 of making a correct decision, a portion .31 is associated with the "more conclusive" C^2 statement for which the conditional confidence coefficient is .98. Asymmetric examples of f_i's and examples in which it is deemed more important to make correct statements under f_1 than f_2, will be given in [2].

For a further 2-hypothesis illustration, suppose $\mathscr{X} =$ real line and that $f_2(x)/f_1(x)$ is nondecreasing, positive and finite. The finest conditioning (continuum B) that makes $\Gamma_1 = \Gamma_2$ w.p. 1 under both f_i can be shown to be determined by letting $\int_{-\infty}^{c_0} [f_1(x)f_2(x)]^{\frac{1}{2}}\, dx = \int_{c_0}^{\infty} [f_1(x)f_2(x)]^{\frac{1}{2}}\, dx$ and $C_1 = (-\infty, c_0)$; and by making $c(x) < c_0$ satisfy, for $x > c_0$, the relation $\int_{-\infty}^{c(x)} [f_1(t)f_2(t)]^{\frac{1}{2}}\, dt = \int_x^\infty [f_1(t)f_2(t)]^{\frac{1}{2}}\, dt$. Then $B = [c_0, \infty)$ and, except for the exceptional value $b = c_0$, we set $C^b = \{c(b), b\}$. This construction is illustrated and discussed in [2].

It will often be a practical convenience in examples like the above, where $\Gamma_\omega{}^b$ is independent of ω, to relabel B in such a way that $\Gamma^b = b$.

1.4. *Randomization.* We now discuss the representation of randomization and its role in the theory. A standard device in $\mathscr{C}^1$ developments is the representation of the randomization device as Lebesgue measure μ^1 on the Borel sets $\mathscr{L}$ of the unit interval I. (We denote Lebesgue measure on I^k by μ^k.) A randomized procedure for a problem with underlying measure space $(\mathscr{X}', \mathscr{B}', \nu')$ can then be regarded as a nonrandomized procedure on $(\mathscr{X}, \mathscr{B}, \nu) = (\mathscr{X}' \times I, \mathscr{B}' \times \mathscr{L}, \nu' \times \mu^1)$, which can be thought of as the background of our model. In particular, the underlying ν is then atomless, and certain sets of operating characteristics considered later are convex. As will be discussed in Section 4, if the original

ν' was atomless, results of Dvoretzky, Wald and Wolfowitz (1951) imply that such an introduction of randomization is unnecessary to achieve these operating characteristics, under assumptions that include $\mathscr{C}^L$ procedures for the k-hypothesis problem treated herein.

In practical terms, the use of randomized procedures is subject to all the criticisms present in $\mathscr{C}^1$ developments, but such criticism is perhaps even more pointed for conditional procedures, where one may find it unappealing to condition on the value of $Z(X)$ but not on the outcome of the randomization scheme. In many discrete case settings, the practitioner will do what he does in $\mathscr{C}^1$ considerations, and sacrifice exact fulfillment of some stated probabilistic objectives in favor of using a nonrandomized procedure. As in the case of $\mathscr{C}^1$, we consider the randomization enlargement from $\mathscr{X}'$ to $\mathscr{X}$ in order to achieve the convexity of some set of operating characteristics.

We also note that, although the set of operating characteristics of procedures in $\mathscr{C}$ or $\mathscr{C}^+$ turns out to be convex (Theorem 4.3), that of procedures in $\mathscr{C}^L$ (or $\mathscr{C}^{L+}$) does not. This can be seen from the fact that a 50–50 random choice between two procedures in $\mathscr{C}^L$ can be viewed in $\mathscr{X}' \times I$ as a $\mathscr{C}^L$ partition of $\mathscr{X}' \times [0, \frac{1}{2})$ together with another partition of $\mathscr{X}' \times [\frac{1}{2}, 1]$, and is thus in $\mathscr{C}^{2L}$ rather than in $\mathscr{C}^L$.

On the other hand, $\mathscr{C}$ and $\mathscr{C}^L$ are "closed" in a natural sense, while $\mathscr{C}^+$ and $\mathscr{C}^{L+}$ are not (Sections 4 and 5).

Further discussion of randomization, and an illustrative example of what is lost by not randomizing in a discrete case, is found in [2].

It has seemed best for the purpose of trying to make the ideas of this paper understandable, not to use the most general notation throughout. Thus, in the previous exposition and in development of the simple sufficient conditions for admissibility of Section 3, we can use nonrandomized (δ, Z), understood to be based on $\mathscr{X} = \mathscr{X}' \times I$ as described above, and thus to be equivalent to use of randomized procedures on $\mathscr{X}'$. Such randomized procedures η are defined in general terms in Section 4, where also the notion of a "canonical procedure" is introduced in order to prove convergence and compactness properties of $\mathscr{C}$. Section 5, where necessary conditions for admissibility are obtained, can then be read without detailed reading of Section 4. Section 5 and the end of Section 3 are written in a special form convenient for the k-hypothesis problem, in terms of the sufficient statistic $\{f_i/\sum_j f_j, 1 \leqq i \leqq k\}$ that takes on values in the $(k-1)$-simplex $\mathscr{S}_{k-1}$. (The role of sufficiency is the same as in $\mathscr{C}^1$ theory and will not be discussed further.) Accordingly, a procedure can be regarded either as a randomized procedure on $\mathscr{S}_{k-1}$ or as a nonrandomized partition of $\mathscr{T}_k = \mathscr{S}_{k-1} \times I$, and we try to use the two points of view in such a manner as to aid understanding: the former when proofs require it, the latter when it gives a simple geometric picture of a procedure, especially when the partition is "almost" a cartesian product of I with a partition of $\mathscr{S}_{k-1}$. In using $\mathscr{T}_k$ we denote partitions by $\hat{C}$.

The classes $\mathscr{C}$, $\mathscr{C}^+$, $\mathscr{C}^B$ are to be viewed as consisting of all randomized η on $\mathscr{X}'$, or equivalently of all partitions of $\mathscr{X}' \times I$ that have the appropriate properties, possibly with $\mathscr{S}_{k-1}$ for $\mathscr{X}'$.

2. Goodness criteria. A natural beginning for optimality considerations is the intuitive notion that, if ω is true, we prefer a procedure that states this truth in a form that reflects strong conclusiveness, to one that asserts it only as a weakly felt conclusion. This comparison is meaningful if the first procedure has at least as much probability of making its assertion as the second; otherwise, a direction of preference is not so clearly natural. We are led, then, to consider the tail probability law

$$G_\omega(t) = P_\omega\{\Gamma_\omega > t;\, \delta(X) \in D_\omega\} \tag{2.1}$$

for $0 \leqq t \leqq 1$. This function is right continuous. It is also defined at $t = 0-$, and we now verify its behavior at 0, for future reference.

LEMMA 2.1. *We have* $G_\omega(0-) = G_\omega(0)$, *and consequently*

$$G_\omega(0) = P_\omega\{\delta(X) \in D_\omega\} = 1 - P_\omega\{Q_\omega\}\,. \tag{2.2}$$

PROOF. We use the fact that Γ_ω is $\mathscr{B}_0$-measurable to compute that the "jump" in question is

$$\begin{aligned} P_\omega\{\delta(X) \in D_\omega,\, \Gamma_\omega = 0\} &= E_\omega\{E_\omega\{I_{\{\delta(X) \in D_\omega\}} I_{\{\Gamma_\omega = 0\}} \mid \mathscr{B}_0\}\} \\ &= E_\omega\{I_{\{\Gamma_\omega = 0\}} E_\omega\{I_{\{\delta(X) \in D_\omega\}} \mid \mathscr{B}_0\}\} \\ &= E_\omega\{I_{\{\Gamma_\omega = 0\}} \Gamma_\omega\} = 0\,. \end{aligned} \tag{2.3}$$

□

A possible notion of admissibility, that quantifies the intuitive notion that led to consideration of (2.1), is to regard $\bar{\bar{C}}$ as better than $\bar{C}$ if

$$\bar{\bar{G}}_\omega(t) \geqq \bar{G}_\omega(t) \quad \forall\, \omega,\, t\,, \tag{2.4}$$

with strict inequality somewhere.

In Section 6 we discuss some of the possible modifications of this criterion. While (2.4), with the adjunction of (2.6) or (2.8) as described hereafter, seems to the author the simplest criterion that conforms with the earlier motivating comments, it is not the aim of this paper to insist on the use of any one such criterion. Rather, these first considerations are intended to typify the formulation and study of reasonable conditional admissibility criteria. Many of the comments of Section 2 apply qualitatively to other criteria of similar character. Some criteria lead to further complications or simplifications. For example, the use of $E_\omega \Gamma_\omega = P_\omega\{C_\omega\}$ in place of (2.1) reduces the admissibility question to a known one of $\mathscr{C}_1$ decision theory, but (Section 6(B)) this criterion may not reflect the aims that motivated use of conditional confidence procedures.

For a procedure in $\mathscr{C}^1$, the function $G_\omega(t)$ has only a single jump, at $t = \Gamma_\omega = P_\omega\{\delta(X) \in D_\omega\}$, of magnitude Γ_ω. Hence, (2.4) is the usual notion of domination in terms of the risk function for zero-one loss function. However, for a procedure outside $\mathscr{C}^1$, (2.4) alone is not satisfactory. To see this, consider the k-hypothesis problem, where $\Omega = \{1, 2, \cdots, k\} = D$ and $D_i = \{i\}$, so that

$C_i^b = K_i^b$. Let us fix the C_i's and consider only partitions $\{C_i^b\}$ with those C_i's. Every such procedure has $G_i(0) = P_i\{C_i\}$, but essentially only the partition $\tilde{C}$ with $\tilde{C}_i^i = C_i$ has $G_i(1-) = P_i\{C_i\}$, and thus it is essentially the only partition with the given C_i that is admissible in the sense of (2.4).

This partition $\tilde{C}$ is highly unappealing intuitively. It always asserts $\Gamma_i = 1$ when decision d_i is reached, but does so in the form of a trivial conditioning, "when ω is true, the conditional probability of asserting it is true, given that it is asserted to be true, is one." It does not come close to achieving an aim that seems essential to the usefulness of conditional confidence procedures: that, for each b, the $\Gamma_\omega{}^b$ be of comparable magnitude (and perhaps even equal). This property is important for two reasons. Firstly, it is useful to be able to regard different values of Z as indexing different levels of conclusiveness; thus, in $\mathscr{C}^2$ we might like to regard decisions made when $X \in C^1$ as being "weakly conclusive" and those made when $X \in C^2$ as being "strongly conclusive." Secondly, from a practical point of view, it is useful (just as it is in the case of $\mathscr{C}^1$ confidence intervals) to be able to state a (conditional) confidence coefficient or useful lower bound, that is independent of ω—the $\Gamma^{Z(x)}$ or $\psi(x)$ of Section 1.

There are two fairly obvious possible modifications to using (2.4) alone. One of these is to impose on the class of procedures being considered a restriction that eliminates procedures like $\tilde{C}$. The restriction to procedures in $\mathscr{C}^+$ achieves this, since $\tilde{C}$ is degenerate. Because of the simplicity of this restriction, it is useful theoretically in admissibility developments. However, in using it one should keep in mind that the difficulty presented by comparison of procedures with $\tilde{C}$ is not completely eliminated by restriction to $\mathscr{C}^+$. This is because, in this k-hypothesis setting, it is easy to modify $\tilde{C}$ very slightly to obtain a procedure $\tilde{\tilde{C}}$ for which $P_i\{\tilde{\tilde{C}}_i{}^j\}$ is positive but small for $i \neq j$; this $\tilde{\tilde{C}}$ is in $\mathscr{C}^+$, and the Lévy distance between $\tilde{G}_\omega$ and $\tilde{\tilde{G}}_\omega$ is small. A procedure C in $\mathscr{C}^+$ with the same C_i's as $\tilde{\tilde{C}}$, and with almost equal $\Gamma_\omega{}^b$'s for each fixed b, will then not be *strictly* worse than $\tilde{\tilde{C}}$ in the sense (2.4), by Corollary 3.2; but C is *close* to being "very much worse" than $\tilde{\tilde{C}}$, in that its G_ω is much worse than $\tilde{G}_\omega$ which is close to $\tilde{\tilde{G}}_\omega$; thus, we have a "subadmissibility" phenomenon. Having found the admissible procedures in $\mathscr{C}^+$, then, further considerations in the form of the next paragraph, or in achieving reasonable closeness of the $\Gamma_\omega{}^b$'s for each fixed b (which rules out $\tilde{\tilde{C}}$), are necessary in choosing satisfactory procedures.

A second possible alteration of (2.4), even without *imposing* the restriction to $\mathscr{C}^+$, involves examination of incorrect assertions: if ω is true, and if we state that it is not true (by making a decision outside D_ω), it is better not to sound so sure about that untruth. This leads to consideration of laws associated with Γ and variants of it, under each possible true ω. Such possibilities will be discussed in Section 6(B), near the end of which the difficulty of using one of the most natural possibilities is discussed. For now, we mention only the simple

$$H_\omega(t) = P_\omega\{\delta(X) \notin D_\omega,\ 1 - \Gamma_\omega > t\}\,, \tag{2.5}$$

motivated by the fact that $1 - \Gamma_\omega$ is the conditional probability, under ω, of making an incorrect decision. We then define $\bar{\bar{C}}$ as *better than* $\bar{C}$ *in the sense* (2.4) $\wedge$ (2.6) if (2.4) is satisfied and also

$$\bar{\bar{H}}_\omega(t) \leqq \bar{H}_\omega(t) \quad \forall\, t, \omega\,, \tag{2.6}$$

with strict inequality somewhere in either (2.4) or (2.6). The procedures $\tilde{C}$ and $\tilde{\tilde{C}}$ perform as poorly in terms of (2.6) as they did well in terms of (2.4), so the anomaly obtained from using (2.4) alone is thereby eliminated. It is somewhat simpler to consider only the value $t = 1-$ in (2.6), yielding

$$P_\omega\{\bar{\bar{\delta}}(X) \notin D_\omega, \bar{\bar{\Gamma}}_\omega = 0\} \leqq P_\omega\{\bar{\delta}(X) \notin D_\omega, \bar{\Gamma}_\omega = 0\} \qquad \text{for all} \quad \omega\,, \tag{2.7}$$

which (by Lemma 2.1) is equivalent to

$$P_\omega\{\bar{\bar{Q}}_\omega\} \leqq P_\omega\{\bar{Q}_\omega\} \qquad \text{for all} \quad \omega\,. \tag{2.8}$$

This leads to another notion of admissibility, obtained by defining $\bar{\bar{C}}$ to be better than $\bar{C}$ if (2.4) and (2.8) hold with strict inequality somewhere. The simplicity of this definition makes it, too, a useful theoretical notion. Although it eliminates the domination of $\tilde{C}$, it cannot eliminate the subadmissibility of $\tilde{\tilde{C}}$, since all procedures $\bar{\bar{C}}$ in $\mathscr{C}^+$ trivially satisfy (2.8) for all $\bar{C}$. Nevertheless, it has proved useful to delimit admissible procedures in the sense of (2.4) $\wedge$ (2.8) and then to consider functions such as those of (2.5) as parts of a supplementary operating characteristic to be studied further along with those of (2.1) in selecting a procedure. As far as *admissibility* is concerned, adjoining something like (2.6) to (2.4) and (2.8) cannot decrease the class of admissible procedures. Since a subset of that class will be seen in Section 3 already to be quite large *without* adjoining (2.6), and since examples studied thus far do not reveal any severe shortcoming of restricting selection of a procedure to those procedures proved in Section 3 to be admissible in the sense of (2.4) $\wedge$ (2.8) (or even to the smaller class of $\mathscr{C}^+$ procedures admissible in the sense of (2.4)), we shall not pursue such weaker admissibility notions further here. If other notions of "loss" are appropriate, this limitation to admissibility in the sense of (2.4) $\wedge$ (2.8) may prove unwise. In Section 5 we characterize (2.4) $\wedge$ (2.6)-admissible procedures partly because of the difficulty of using (2.8) directly as part of the admissibility criterion in a complete class proof that proceeds by maximizing certain linear functionals.

Choosing a conditional confidence procedure from the admissible ones is perhaps more difficult here than in $\mathscr{C}^1$ decision theory, because of the additional properties of interest. Some considerations additional to performance in terms of (1.3), (2.1), (2.5) as described above are: (1) In many problems the Γ_ω^b for each fixed b should be equal or close, for reasons mentioned earlier (although, as in the $\mathscr{C}^1$ NP 2-hypothesis setting, this is not always desirable). (2) The Γ_ω^b should have a large variation in b (possibly reflected through some specific measure of dispersion of G_ω); otherwise, a simpler procedure, with smaller B—perhaps even a $\mathscr{C}^1$ procedure—could as well be used. (3) Usually it is unattractive to a practitioner to use a procedure in which any $\Gamma_\omega{}^b$ is too small; this may, however, motivate enlargement of D to include decisions reflecting indifference for borderline data.

Other aspects enter when the simple loss structure does not accurately represent all the experimenter's potential rewards and penalties. For example, in interval estimation or problems with complex D (see (3) above), some measure of "size" associated with the decision may be relevant. (See Section 6 (C).) The computational complexity of large B might be another concern. Also, since there are so many admissible procedures in $\mathscr{C}^{L+}$, the imposition of some simplifying structure on the form of the procedure may help computationally in the construction and comparison of partitions; for example, in the 2-hypothesis problem one might, in $\mathscr{C}^{L+}$, restrict consideration to procedures with *monotone interval structure*, those in which, if $S_i{}^b$ is the convex closure of the range of $f_2(x)/f_1(x)$ for x in $C_i{}^b$, any two $S_i{}^b$'s have at most one point in common, and $S_1{}^b$ is to the left of $S_1{}^{b'}$ if and only if $S_2{}^b$ is to the right of $S_2{}^{b'}$. (According to Theorem 3.1, some of the intuitively less appealing procedures that violate this form are also admissible.) The operating characteristics of partitions of such a restricted form may not constitute so large a set as those of all $\mathscr{C}^{L+}$ procedures. On the other hand, at least in simple cases (Ω, D finite), there are typically infinitely many procedures with a given operating characteristic (functions of (1.3), (2.1), (2.5)), and it is desirable computationally to be able to achieve any such operating characteristic, at least approximately, with a procedure of simple form.

These aspects of computing and choosing among conditional confidence procedures (in particular, of justifying the restriction of the previous paragraph) are clearly quite involved and will not be treated further here. Illustrations and further discussion will be contained in [2]. Here we remark only that $\mathscr{C}^L$ procedures are more robust for small L.

In the succeeding sections, where various admissibility concepts are used, "admissibility in the sense of (A)" or "(A)-admissibility" will be used to refer to the criterion obtained from the "at least as good as" inequalities (A). We write $\mathbf{G} = \{G_\omega, \omega \in \Omega\}$ and $\mathbf{H} = \{H_\omega, \omega \in \Omega\}$.

3. Sufficient condition for admissibility. For any partition $C = \{C_d{}^b\}$ in $\mathscr{C}$, there is an associated $\mathscr{C}^1$ partition $|C| = \{|C_d|\}$ defined by $|C_d| = \bigcup_b C_d{}^b = \{x : \delta(x) = d\}$. We call $|C|$ the *decision partition* or *underlying* $\mathscr{C}^1$ *partition* of C. For procedures $\mathscr{C}^1$, admissibility in the sense of (2.4) $\wedge$ (2.8) (or (2.4) alone) is the standard decision-theoretic admissibility for risk function $P_\omega\{\delta(X) \notin D_\omega\}$. In this section we shall prove

THEOREM 3.1. *If $|C|$ is admissible among $\mathscr{C}^1$ procedures, then C is admissible in $\mathscr{C}$ in the sense of* (2.4) $\wedge$ (2.8) (*and hence in the sense of* (2.4) $\wedge$ (2.6)).

Since all $\mathscr{C}^+$ procedures have $P_\omega\{Q_\omega\} = 0$, we have

COROLLARY 3.2. *If $C \in \mathscr{C}^+$ and $|C|$ is admissible in $\mathscr{C}^1$, then C is admissible in $\mathscr{C}^+$ in the sense of* (2.4).

By restricting competitors in the proof of Theorem 3.1, we obtain

THEOREM 3.3. *If $\mathscr{C}$, $\mathscr{C}^+$ are replaced by $\mathscr{C}^B$, $\mathscr{C}^{B+}$ for any fixed B, the statements of Theorem* 3.1 *and Corollary* 3.2 *remain true.*

Since (Section 1.4) $\mathscr{X}$ can be thought of as containing a randomization component I, the admissibility asserted above and in Corollary 3.4 allows also for competitors that randomize among conditionings on the original $\mathscr{X}'$.

Note that it has not been necessary in the above statements to make regularity assumptions on D and Ω (in particular, on their finiteness, so the conclusions apply to such settings as interval estimation), because we do not touch on the detailed characterization of $\mathscr{C}^1$-admissible procedures; under various assumptions there are well-known results in terms of Bayes procedures and finer structure (e.g., [7]). In contrast, in Section 5 we treat the k-hypothesis problem in order to obtain explicit *necessary* conditions for admissibility. The results of the present section would be even more obvious if the operating characteristic for admissibility were $\{P_i\{C_i^b\}\}$; it is not that, but the proof in terms of (2.1) is not much more complicated. The results here establish the unfortunately large collection of admissible procedures mentioned in Section 2. Section 5 indicates that there are in general even more admissible procedures than those characterized in the present section. However, the admittedly limited calculation of examples to date indicates no great practical advantage to be found in using procedures outside those characterized above.

PROOF OF THEOREM 3.1. We begin with a simple computation (3.2) that holds for arbitrary partitions C. Let B_ω be the σ-field induced on $\mathscr{X}$ by Γ_ω. Define the measures $K_\omega{}^*$ and $G_\omega{}^*$ on $(\mathscr{X}, \mathscr{B}_\omega)$ by

$$(3.1) \qquad K_\omega{}^*(A) = P_\omega\{A\}\,, \qquad G_\omega{}^*(A) = P_\omega\{A; \delta(X) \in D_\omega\}\,,$$

and let $K_\omega{}^\#$ and $G_\omega{}^\#$ be corresponding Borel measures on $[0, 1]$ induced by Γ_ω; that is $G_\omega{}^\#(L) = G_\omega{}^*(\Gamma_\omega{}^{-1}(L))$, and similarly for $K_\omega{}^\#$. A straightforward conditional expectation calculation shows that $dG_\omega/dK_\omega{}^* = \Gamma_\omega$ a.e. $(K_\omega{}^*)$ and (since $-\int_L dG_\omega(t) = G_\omega{}^\#(L)$) consequently that $(dG_\omega{}^\#/dK_\omega{}^\#)(t) = t$ a.e. $(K_\omega{}^\#)$. Moreover, $K_\omega{}^\# \ll G_\omega{}^\#$ on $(0, 1]$, where $(dK_\omega{}^\#/dG_\omega{}^\#)(t) = t^{-1}$. Thus,

$$(3.2) \qquad -\int_{0+}^1 t^{-1}\, dG_\omega(t) = \int_{0+}^1 t^{-1} G_\omega{}^\#(dt) = \int_{0+}^1 K_\omega{}^\#(dt) = 1 - P_\omega\{Q_\omega\}\,.$$

We also recall that, if $\bar{F}$ and $\bar{\bar{F}}$ are two df's on $(0, \infty)$, then

$$(3.3) \qquad \bar{F}(t) \geqq \bar{\bar{F}}(t) \quad \forall\, t$$

implies, for every strictly increasing function h integrable under both $\bar{F}$ and $\bar{\bar{F}}$,

$$(3.4) \qquad \int_{0+}^\infty h(t)\, d\bar{F}(t) \leqq \int_{0+}^\infty h(t)\, d\bar{\bar{F}}(t)\,,$$

and strict inequality for some t in (3.3) implies strict inequality in (3.4).

Now suppose $\bar{C}$ is a partition whose underlying $\mathscr{C}^1$ partition is $\mathscr{C}^1$-admissible, but which is not itself admissible in the sense (2.4) $\wedge$ (2.8); let $\bar{\bar{C}}$ be better. Thus, (2.4) and (2.8) hold, with strict inequality somewhere. At $t = 0$, (2.4) becomes (by Lemma 2.1) $P_\omega\{\bar{\bar{\delta}}(X) \in D_\omega\} \geqq P_\omega\{\bar{\delta}(X) \in D_\omega\}\ \forall\ \omega$, and equality must hold because the underlying $\mathscr{C}^1$ partition of $\bar{C}$ is $\mathscr{C}^1$-admissible. Thus, $\bar{G}_\omega(0) = \bar{\bar{G}}_\omega(0)\ \forall\ \omega$. For each ω for which $\bar{G}_\omega(0) > 0$ (and it is easily seen that we need consider only such ω), we consequently have from (3.1) that

$$\begin{aligned}\bar{F}_\omega(t) &\equiv 1 - \bar{G}_\omega(t)/\bar{G}_\omega(0)\,, \\ \bar{\bar{F}}_\omega(t) &\equiv 1 - \bar{\bar{G}}_\omega(t)/\bar{G}_\omega(0)\end{aligned} \tag{3.5}$$

are both df's on (0, 1]. By (2.4) we have (3.3) and thus (3.4) for these df's. For $h(t) = -t^{-1}$, we obtain from (3.2)

$$P_\omega\{\bar{Q}_\omega\} \leqq P_\omega\{\bar{\bar{Q}}_\omega\} \quad \forall\ \omega\,, \tag{3.6}$$

and since this h is strictly increasing we conclude that strict inequality in (2.4) for any ω (for some t) implies strict inequality in (3.6). Since we cannot have both (3.6) and (2.8) with a strict inequality somewhere, this contradicts existence of the assumed $\bar{\bar{C}}$. □

Before turning to preliminaries used in establishing necessary conditions for admissibility, we describe some of the above results in other terms that will be useful later. For that purpose, we need only treat the case where Ω and D are finite, say $\Omega = \{1, 2, \cdots, k\}$ and $D = \{1, 2, \cdots, r\}$. In terms of the discussion of Section 1, we think of procedures as partitions of $\mathscr{X} = \mathscr{X}' \times I$. A familiar description of $\mathscr{C}^1$-Bayes procedures is given in terms of the probability k-vectors, whose range is the $(k-1)$-simplex $\mathscr{S}_{k-1} = \{(s_1, \cdots, s_k) : \sum_1^k s_i = 1;\ s_i \geqq 0\ \forall\ i\}$. In order to work in terms of nonrandomized procedures (as discussed in Section 1), we replace $\mathscr{S}_{k-1}$ by $\mathscr{T}_k = \mathscr{S}_{k-1} \times I$. Write $f^\# = (f_1^\#, \cdots, f_k^\#)$ for the mapping of $\mathscr{X}'$ into $\mathscr{S}_{k-1}$ defined (a.e. $\sum_1^k f_i\, d\nu$) by $f_\omega^\#(x) = f_\omega(x)/\sum_1^k f_i(x)$, and let

$$\hat{\nu}(A) = \textstyle\int_{f^{\#-1}(A)} \sum_1^k f_i(x)\nu(dx)\,. \tag{3.7}$$

The sufficient statistic $S = f^\#(X)$ then has density $\hat{f}_i(s) = s_i$ on $\mathscr{S}_{k-1}$ with respect to $\hat{\nu}$, when i is true. It does not matter that this reduction from $\mathscr{X}'$ to $\mathscr{S}_{k-1}$ may introduce atoms, since we still have μ^1 on I (the last component of $\mathscr{T}_k$) to represent "randomization," as described in Section 1. In place of partitions of $\mathscr{X}$ we then consider partitions $\hat{C}$ of $\mathscr{T}_k$. Every (nonrandomized) $\mathscr{C}^1$-Bayes procedure on $\mathscr{T}_k$ in these terms is equivalent to a partition $|\hat{C}|$ of $\mathscr{T}_k$ which, for some probability vector $p = \{p_i\}$, satisfies

$$\hat{C}_j \subset \{(s, u) : \textstyle\sum_{i \in \Omega_j} p_i s_i = \max_h \sum_{i \in \Omega_h} p_i s_i\} \quad \forall\ j\,. \tag{3.8}$$

An important feature is the well-known convexity of the sets on the right side of (3.8). The $\mathscr{C}^1$-Bayes procedures may include some inadmissible procedures, but that cannot occur if all f_i have the same support; and Bayes procedures relative to strictly positive p are admissible in any case. Characterization of admissible $\mathscr{C}^1$ Bayes procedures in general can be found in [7]. To shorten further discussion, we paraphrase Corollary 3.2 and Theorem 3.3 in the simplest case, in terms of partitions of $\mathscr{T}_k$. We hereafter denote by cl A, int A, and cc A, the closure, interior, and convex closure of a subset A of $\mathscr{S}_{k-1}$ or $\mathscr{T}_k$ in the relative Euclidean topology of either of these. By $s^{(i)}$ we denote the ith vertex of $\mathscr{S}_{k-1}$, where $s_i^{(i)} = 1$. The *border* of $\mathscr{S}_{k-1}$ is the subset $\{s: \text{some } s_i = 0\}$.

COROLLARY 3.4. *For the k-hypothesis problem where all f_i have the same domain of positivity, a partition $\hat{C}$ of $\mathscr{T}_k$ in $\mathscr{C}^+$ is admissible in $\mathscr{C}^+$ in the sense of* (2.4), *if for some strictly positive p the decision partition $|\hat{C}|$ satisfies*

$$(3.9) \qquad \begin{array}{ll} \text{(i)} & \textit{for each } i, \ \text{int } \hat{C}_i \ \textit{contains} \ s^{(i)} \times I; \\ \text{(ii)} & \text{cc } \hat{C}_i \cap \text{cc } \hat{C}_j \subset \{(s, u): p_i s_i = p_j s_j\}, \qquad \textit{for} \ i \neq j. \end{array}$$

The conclusion holds if $\mathscr{C}^+$ is replaced by $\mathscr{C}^{B+}$ for fixed B.

We hereafter say a partition $\hat{C}$ of $\mathscr{T}_k$ has *structure W* if its $|\hat{C}|$ satisfies (3.9). Under the assumptions of Corollary 3.4, the only procedures admissible in $\mathscr{C}^+$ for the k-sample problem that are covered in Corollary 3.2 are those of structure W. The procedures $\hat{C}$ with structure W have $|\hat{C}|$ in $\mathscr{C}^{1+}$, but of course require further conditions on the subpartition of $\hat{C}_j$ into the $\hat{C}_j^b$, in order that $\hat{C}$ be in $\mathscr{C}^+$.

When the f_i do not have common domain of positivity, it is possible to extend Corollary 3.4 to include some procedures that do not satisfy (3.9)(i), if $\nu > 0$ on the border of $\mathscr{S}_{k-1}$, e.g., if $\nu(s^{(i)}) > 0$ for some i.

4. Canonical procedures, randomization and topology of $\mathscr{C}$. Throughout this section Ω and D are finite, although Lemma 4.1 will be seen at once to hold more generally and the other results can be proved under compactness assumptions on D and (except for the result taken from [3]) on Ω. We begin by treating a difficulty related to but beyond the phenomena that motivate the notion of regular convergence of $\mathscr{C}^1$ decision functions. In the latter, in such a simple setting as the two (simple)-hypothesis problem with $\nu = \mu^1$ on $\mathscr{X}' = [0, 1)$, a sequence of nonrandomized decision functions need not converge a.e. to a nonrandomized decision function (or anything else). For example, define the measure ξ_n on $\mathscr{X}' = [0, 1)$ by its density

$$(4.1) \qquad \begin{aligned} d\xi_n(x)/dx &= 1 \qquad \text{if} \ 2i/n \leqq x < (2i+1)/n \ \text{ for any integer } i, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

Then, if ${}^n\bar{\eta}$ is the $\mathscr{C}^1$ decision function whose probability of making d_1 when $X = x$ is ${}^n\bar{\eta}(d_1 \mid x) = d\xi_n(x)/dx$, so that ${}^n\bar{\eta}$ is nonrandomized, the sequence of ${}^n\bar{\eta}$ converges almost nowhere pointwise, but converges "regularly" to $\bar{\eta}^*(d_1 \mid x) \equiv \frac{1}{2}$. (Formal definitions will be given below.) Now, with $B = \mathscr{X}'$ and $Z(x) = x$, consider the nonrandomized conditional confidence procedure ${}^n\eta$ for which the probability of making decision d_i and conditioning with label $b = Z(x)$ when

$X = x$ is given by

$$(4.2)\qquad {}^n\eta(d_1, Z(x) \mid x) = 1 - {}^n\eta(d_2, Z(x) \mid x) = d\xi_n(x)/dx\,.$$

(${}^n\eta(d_i, Z(x) \mid x)$ is simply the characteristic function of ${}^nC_{d_i}^{Z(x)}$. The degeneracy of these ${}^nC$'s is inessential and only for the purpose of simplifying the illustration; $\mathscr{C}^+$ examples are easily given.) It might seem natural to consider these ${}^n\eta$ to converge "regularly," if at all, to $\eta^*(d_i, Z(x) \mid x) \equiv \frac{1}{2}\ \forall\, i$. But if $f_1(x) \equiv 1$, the use of a fair coin to choose between d_1 and d_2 for each x would make $P_i\{\text{make } d_i \mid X = x\} = \frac{1}{2}\ \forall\, x$ whereas ${}^n\Gamma_1(x) = P_1\{{}^n\delta(x) = d_1 \mid X = x\}$ is only 0 or 1. We would thus not have the desired convergence of ${}^nG_1$. The difficulty of course stems from the fact that $\xi_n \to \xi^*$ weakly does not imply $d\xi_n/dx \to d\xi^*/dx$ anywhere, and this last, of no consequence in $\mathscr{C}^1$ theory, matters here because of our use of ${}^n\Gamma_i$ and ${}^n\mathbf{G}$ as measures of conditional confidence performance. A way out of the difficulty (which is more obvious in the case of finite B) is directed by the comments in Section 1 alluding to conditioning on the outcome of a randomization. If we let $B^* = [0, 1) \times M$ where $M = \{1, 2\}$, and replace $\mathscr{X}'$ by $\mathscr{X}^* = \mathscr{X}' \times M$ with ν replaced by $\nu^* = \nu \times \rho$ with $\rho(1) = \rho(2) = \frac{1}{2}$ (the randomization), we obtain ${}^*\Gamma_i = 0$ or 1 with the desired probabilities. Here we replaced B by two copies of it in a manner that might not be so obvious to imitate in general. We now introduce a structure that makes such a construction mechanical (and which yields a different but simpler solution to the difficulty in the above example).

For fixed B with σ-field $\mathscr{B}_0'$, let $J = D \times B$ with σ-field $\mathscr{J} = 2^D \times \mathscr{B}_0'$. It may help motivative the definition that follows to note that a nonrandomized (δ, Z) can be thought of as a mapping from $\mathscr{X}'$ to J; in this interpretation, $\Gamma_\omega = P_\omega\{(\delta, Z) \in (D_\omega \times B) \mid \mathscr{D}_0 \times \mathscr{B}_0'\}$ where $\mathscr{D}_0$ is the trivial σ-field on D. A conditional confidence procedure, expressed in terms that allow for randomization, is a function η on $\mathscr{J} \times \mathscr{X}'$ such that $\eta(A' \mid x)$ is a probability measure $(\mathscr{J})$ in A' for each x and is measurable $(\mathscr{B})$ for each A'. If η is used and $X = x$, and if the randomization according to $\eta(\cdot \mid x)$ produces outcome (d, b), we make decision d and regard the conditioning label as b. Define $P_{\omega,\eta}\{A'\} = E_\omega \eta(A' \mid X)$. The conditional probability $P_{\omega,\eta}\{D_\omega \times B \mid \mathscr{D}_0 \times \mathscr{B}_0'\}$ is well defined and essentially unique $(P_{\omega,\eta}, \forall\, \omega)$. We define

$$(4.3)\qquad \gamma_\omega = P_{\omega,\eta}\{D_\omega \times B \mid \mathscr{D}_0 \times \mathscr{B}_0'\}\,,$$

so that γ_ω is a $\mathscr{D}_0 \times \mathscr{B}_0'$-measurable function on $D \times B$, which we can thus view as a $\mathscr{B}_0'$-measurable function on B. We denote its value at b by $\gamma_\omega{}^b$. If one begins with a nonrandomized δ with its Γ_ω, the γ_ω of the corresponding η is seen to satisfy $\Gamma_\omega(x) = \gamma_\omega{}^{Z(x)}$.

A *canonical conditional confidence procedure* is one that satisfies (with labels distinguished by *)

$$(4.4)\qquad \begin{aligned} &B^* = I^k = \{b^*\} = \{(b_1^*, b_2^*, \cdots, b_k^*)\}\,,\\ &\gamma_\omega{}^{b^*} = b_\omega^* \qquad \text{w.p. 1 under } \omega\,, \quad \forall\, \omega\,, \end{aligned}$$

with $\mathscr{B}_0^{*\prime}$ the Borel sets. (This formalizes the practically convenient relabeling mentioned after the examples of Section 1, in which we regard B as the set of attainable conditional confidence values; but now we relabel whether or not Γ_ω^b is independent of ω.)

Note that, for a canonical η^*, the conditional probability γ_ω is measurable relative to the cylinder sets over Borel sets in the ωth coordinate.

The set of operating characteristics $\mathbf{G}$ of all η^* is easily seen to be convex.

We now prove

LEMMA 4.1. *For every* η, *there is a canonical* η^* *with the same operating characteristic* $\mathbf{G}$.

PROOF. For an arbitrary η, define $g_\eta: B \to L^k$ by $g_\eta(b) = ({}^\eta\gamma_1^b, \cdots, {}^\eta\gamma_k^b)$. Let η^* be the (canonical) procedure given by

$$\eta^*(\Delta \times A \mid x) = \eta(\Delta \times g_\eta^{-1}(A) \mid x) \tag{4.5}$$

for $\Delta \subset D$ and $A \in \mathscr{B}_0^{*\prime}$ in I^k, and then extended from such cylinder sets $\Delta \times A$. To see that η^* is a canonical procedure, we note that, for $A \in \mathscr{B}_0^{*\prime}$, from (4.5),

$$\begin{aligned} \textstyle\int_A b_\omega^* P_{\eta^*,\omega}\{D \times db^*\} &= \textstyle\int_{g_\eta^{-1}(A)} P_{\eta,\omega}\{D_\omega \times B \mid \mathscr{D}_0 \times g_\eta^{-1}(\mathscr{B}_0^{*\prime})\} P_{\eta,\omega}\{D \times db\} \\ &= P_{\eta,\omega}\{D_\omega \times g_\eta^{-1}(A)\} = P_{\eta^*,\omega}\{D_\omega \times A\} \\ &= \textstyle\int_A P_{\eta^*,\omega}\{D_\omega \times B \mid \mathscr{D}_0 \times \mathscr{B}_0^{*\prime}\} P_{\eta^*,\omega}\{D \times db^*\}, \end{aligned} \tag{4.6}$$

so that the first integrand is indeed a version of the last. Note that the possible lumping of several b's with the same vector $\{{}^\eta\gamma_\omega^b, \omega \in \Omega\}$ into a single $g_\eta(b)$ is exhibited in the second integrand of (4.6), where $g_\eta^{-1}(\mathscr{B}_0^{*\prime}) \subset {}^\eta\mathscr{B}_0'$.

Since ${}^\eta\gamma_\omega^b = {}^{\eta^*}\gamma_\omega^{g(b)}$ and $P_{\eta,\omega}\{D_\omega \times g^{-1}(A)\} = P_{\eta^*,\omega}\{D_\omega \times A\}$ by construction, putting $A = \{b^* : b_\omega^* > t\}$ we obtain ${}^\eta G_\omega = {}^{\eta^*}G_\omega \;\forall\, \omega$. □

We call the η^* of Lemma 4.1 a canonical procedure coresponding to η.

In any convergence or compactness proofs concerning $\{G_\omega\}$ we may hereafter limit consideration to canonical procedures.

On the space of operating characteristics $\mathbf{G}$ we use the usual notion of *weak convergence* of a sequence ${}^N\mathbf{G}$ to $\bar{\mathbf{G}}$, as meaning $\int_0^1 c\, d\,{}^N G_\omega \to \int_0^1 c\, d\bar{G}_\omega$ for each ω and continuous c. This is equivalent to convergence in the sense of the Lévy metric of weak convergence; noting that G_ω is nonincreasing, we define $\|\bar{G}_\omega - \bar{\bar{G}}_\omega\|$ here to be the maximum distance along the lines of slope 1 between the graphs (including vertical segments at jumps) of $\bar{G}_\omega$ and $\bar{\bar{G}}_\omega$, and the Lévy metric is $\|\bar{\mathbf{G}} - \bar{\bar{\mathbf{G}}}\| = \sum_1^k \|\bar{G}_\omega - \bar{\bar{G}}_\omega\|$. It will suffice to consider sequences in studying compactness. On the space of canonical procedures η^*, we adopt Wald's $\mathscr{C}^1$ notion of *regular convergence of* $\{{}^N\eta^*\}$ *to* $\bar{\eta}^*$, defining it in the present context (by substituting J for the usual D) to mean

$$\textstyle\int_{\mathscr{X}} g(x)\, \{\int_J c(j)\, {}^N\eta^*(dj \mid x)\}\nu(dx) \to \int_{\mathscr{X}} g(x) \{\int_J c(j) \bar{\eta}^*(dj \mid x)\}\nu(dx) \tag{4.7}$$

for every continuous real function c on J and every g in $\mathscr{L}_1(\mathscr{X}, \mathscr{B}, \nu)$. It is well known ([7] or [5]) that the set of procedures $\{\eta\}$ is sequentially compact in this topology.

The purpose of reduction to canonical procedures is given in the following:

LEMMA 4.2. *If a sequence* ${}^N\eta^*$ *of canonical procedures converges regularly to a procedure* $\bar{\eta}$, *then* $\bar{\eta}$ *is canonical.*

PROOF. By definition, η^* is canonical if and only if $b_\omega = P_\omega\{D_\omega \times B \mid \mathscr{D}_0 \times \mathscr{B}_0'\}$; that is,

$$\int_{\mathscr{X}} \int_J I_A(b) b_\omega \eta^*(dj \mid x) f_\omega(x)\nu(dx) = \int_{\mathscr{X}} \eta^*(D_\omega \times A \mid x) f_\omega(x)\nu(dx) = \int_{\mathscr{X}} \int_J I_{D_\omega}(d) I_A(b) \eta^*(dj \mid x) f_\omega(x)\nu(dx) \tag{4.8}$$

for each ω and every A in $\mathscr{B}_0^{*\prime}$. Since $\mathscr{B}_0^{*\prime}$ is the Borel field on $B^* = I^k$, (4.8) is equivalent to the equation obtained from the two extreme members of (4.8) if I_A is replaced there by an arbitrary continuous function c on B^*. Since both $c(b)b_\omega$ and $I_{D_\omega}(d)c(b)$ represent continuous functions on J, the equation with these two functions as integrands is preserved under weak convergence of a sequence of η^*'s to a limiting $\bar{\eta}^*$, so the latter is canonical. □

From Lemma 4.2 and the previously stated compactness of the set of all η in the sense of regular convergence, we conclude that the set $\{\eta^*\}$ of canonical procedures is compact in that sense. We now prove the rest of

THEOREM 4.3. *The set of all canonical* η^* *is compact in the sense of regular convergence. The set of operating characteristics* ${}^{\eta}\mathbf{G}$ *of the family of all procedures* η, *or of all canonical* η^*, *is convex and is compact in the sense of weak convergence. These results, except for convexity, remain valid if* η *and* η^* *are restricted to* $\mathscr{C}^L$.

PROOF. We observe that, for all continuous c on I,

$$\int_0^1 c(t)\, d\,{}^{\eta^*}G_\omega(t) = \int_{\mathscr{X}} \int_J I_{D_\omega}(d) c(b_\omega) \eta^*(dj \mid x) f_\omega(x)\nu(dx)\,. \tag{4.9}$$

Since $I_{D_\omega}(d)c(b_\omega)$ represents a continuous function on J, we infer the continuity of the map $\eta^* \to {}^{\eta^*}\mathbf{G}$ from the set of canonical procedures in the topology of regular convergence to the set of operating characteristics in the topology of weak convergence. By compactness of the sub-df's on [0, 1], for any sequence $\{{}^N\mathbf{G}\}$ corresponding to canonical $\{{}^N\eta^*\}$ there is a subsequence for which ${}^{N'}\eta \to \bar{\eta}^*$ regularly, with $\bar{\eta}^*$ canonical, and by the previous sentence, ${}^{N'}\mathbf{G} \to {}^{\bar{\eta}^*}\mathbf{G}$ weakly. Finally, regular convergence of a sequence of canonical procedures in $\mathscr{C}^L$ to a limit clearly implies that the limit is in $\mathscr{C}^L$. □

We now enlarge the operating characteristic. Consideration of a sequence of procedures ${}^N\eta$ like the $\tilde{\tilde{C}}$ of Section 2, converging regularly to $\tilde{C}$ there, shows that $P_\omega\{{}^NQ_\omega\}$ does not converge to $P_\omega\{{}^{\tilde{C}}Q_\omega\}$. However, everything we have developed for G_ω can be carried out for H_ω, now defined by

$$H_\omega(t) = E_\omega \eta([D - D_\omega] \times \{b\colon 1 - \gamma_\omega{}^b > t\} \mid X)\,, \tag{4.10}$$

with $\|\bar{H}_\omega - \bar{\bar{H}}_\omega\|$ being obtained as the maximum distance along a line of slope

-1. Lemma 4.2 requires no change, and the straightforward changes in the remainder of our development yield

THEOREM 4.4. *The statements of Lemma* 4.1 *and Theorem* 4.3 *remain valid with* $\mathbf{G}$ *replaced by* $(\mathbf{G}, \mathbf{H})$.

Theorem 5.1 implies such results as completeness of the admissible procedures (in the sense of (2.4) $\wedge$ (2.6)).

Another decision-theoretic consideration (not used in the sequel) is that of "elimination of randomization." The development of pages 5–8 of Dvoretzky, Wald and Wolfowitz (1951) shows that, for our compact metric $D \times B$ (= their D), since Ω is finite, if ν is atomless every randomized η can be replaced by a nonrandomized ${}^*\eta$ with the same $\mathbf{G}, \mathbf{H}$. It is only necessary to interpret their δ^l as an ${}^l\eta$ in our scheme, so that the displays on page 7 give $P_{{}^l\eta,\omega}\{d \times {}^l\beta\} = P_{\eta,\omega}\{d \times {}^l\beta\}$ for each set ${}^l\beta$ in an increasingly fine net of partitions of B. The weak convergence of $\{{}^l\eta\}$ to a nonrandomized ${}^*\eta$ and the remainder of their proof yield $P_{{}^*\eta,\omega}\{d \times A\} = P_{\eta,\omega}\{d \times A\}$ for all d and $A \in \mathscr{B}_0'$, as desired.

The remaining result we need in the next section is one of approximating $\mathscr{C}$ by $\mathscr{C}^L$ in the sense of $\mathbf{G}$ and $\mathbf{H}$. Recall that any η may be regarded as a partition of $\mathscr{X}' \times I$.

LEMMA 4.5. *For every* $\varepsilon > 0$ *there is a finite* $L(\varepsilon)$ *such that, for every* η *in* $\mathscr{C}$, *there is an* $\bar{\eta}$ *in* $\mathscr{C}^{L(\varepsilon)}$ *with* $||{}^\eta\mathbf{G} - {}^{\bar{\eta}}\mathbf{G}|| + ||{}^\eta\mathbf{H} - {}^{\bar{\eta}}\mathbf{H}|| < \varepsilon$.

PROOF. Partition I^k into a finite number of Borel sets $\{U^r; r = 1, 2, \cdots, L(\varepsilon)\}$, each with $\max\{|b_\omega^* - \bar{b}_\omega| : b^*, \bar{b}^* \in U^r, \omega \in \Omega\} < \varepsilon/6k$. Given η, let η^* be the correponding canonical procedure. Define $\bar{\eta}$ (in $\mathscr{C}^{L(\varepsilon)}$) by $\bar{\eta}(d \times r \,|\, x) = \eta^*(d \times U^r \,|\, x)$. Let b^{*r} be any point in U_r. On U_r we have $b_\omega^{*r} - \varepsilon/6k < {}^{\eta^*}\gamma_\omega < b_\omega^{*r} + \varepsilon/6k$ w.p. 1. Integrating over U_r with respect to $P_{\eta^*,\omega}\{D \times db^*\}$, we obtain

$$(4.11) \qquad (b_\omega^{*r} - \varepsilon/6k)P_{\eta^*,\omega}\{D \times U_r\} < P_{\eta^*,\omega}\{D_\omega \times U_r\} < (b_\omega^{*r} + \varepsilon/6k)P_{\eta^*,\omega}\{D \times U_r\}\,.$$

By construction of $\bar{\eta}$, (4.11) is the same as

$$(4.12) \qquad (b_\omega^{*r} - \varepsilon/6k)P_{\bar{\eta},\omega}\{D \times r\} < P_{\bar{\eta},\omega}\{D_\omega \times r\} < (b_\omega^{*r} + \varepsilon/6k)P_{\bar{\eta},\omega}\{D \times r\}\,,$$

so that $|{}^{\bar{\eta}}\gamma_\omega^r - b_\omega^{*r}| < \varepsilon/6k$. The range of values of the rv ${}^{\eta^*}\gamma_\omega$ on the set U_r is thus entirely within $\varepsilon/3k$ of the value of ${}^{\bar{\eta}}\gamma_\omega$ at r, and these sets of values, coupled with any d, have the same probability $P_{\eta^*,\omega}\{d \times U^r\} = P_{\bar{\eta},\omega}\{d \times r\}$. It follows that $||{}^{\eta^*}G_\omega - {}^{\bar{\eta}}G_\omega|| < \varepsilon/2k$ and $||{}^{\eta^*}H_\omega - {}^{\bar{\eta}}H_\omega|| < \varepsilon/2k$. □

5. Necessary condition for admissibility in the k-hypothesis problem. Unfortunately, there seems to be many more admissible partitions than the simple ones obtained in Section 3. Equally unfortunately, the essentially complete class obtained in the present section appears to be far from minimal, except in the case $k = 2$ (Corollary 5.3). This will be discussed at the end of the section. Our aim is to characterize admissibility in terms of underlying $\mathscr{C}^1$-partitions.

Since the results thus far obtained seem far from definitive, we shall give them here only for the k-sample problem. A corresponding statement for general D can be obtained by the same methods, but is even less satisfactory; some particular multidecision problems are treated in [4].

For the sake of easy comparison with the results of Section 3, we state our results in language of Corollary 3.4. In the k-hypothesis problem, the decision partition $|\hat{C}|$ of a partition $\hat{C}$ of $\mathscr{T}_k$ will be said to be *star-shaped* if

(5.1)
(i) $s^{(i)} \times I \in \hat{C}_i$ for each i;
(ii) for each i and line $L = \{s: \sum_j c_{ij}(s_j - s_j^{(i)}) = 0\}$ through $s^{(i)}$, there is a closed segment L' of L such that $(\text{int } L') \times I \subset (L \times I) \cap \hat{C}_i \subset L' \times I$,

where int L' refers to the topology of R^1. The partition will be called *polyhedral* if each int $\hat{C}_i$ is the interior of a finite union of closed nondegenerate (irregular) k-simplices. Thus, each star-shaped decision partition of $\mathscr{T}_k$ is obtained by taking the cartesian product with I of a star-shaped partition of $\mathscr{S}_{k-1}$ (defined by removing I everywhere in (5.1)), possibly modifying it by introducing randomization on common boundaries of the $\hat{C}_i$. A decision partition satisfying (5.1) and also (3.9)(i) will be termed *positive star-shaped.* This last property is possessed by structure W partitions, but the projections on $\mathscr{S}_{k-1}$ of the cl $\hat{C}_i$ of a positive star-shaped partition are not even necessarily convex. The possibility of replacing star-shapedness by a more special structure will be discussed at the end of this section.

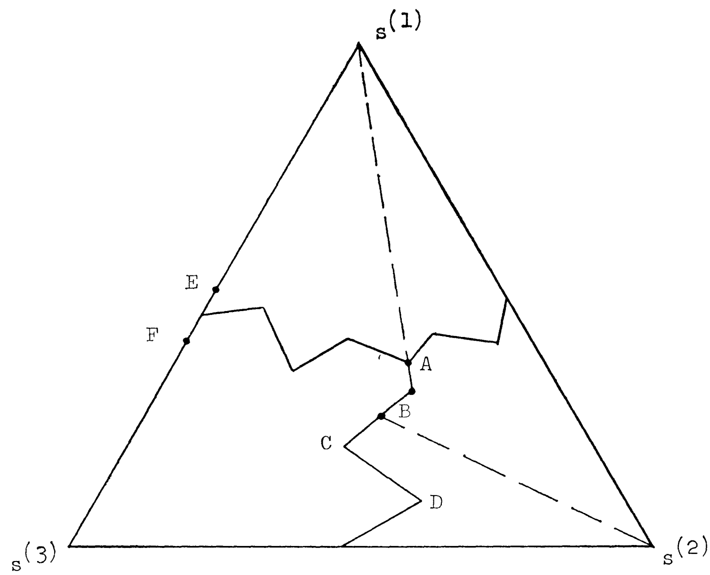

FIG. 1. A positive polyhedral almost star-shaped partition, shown on $\mathscr{S}_2$ when $k = 3$. Solid lines border stars; dashed lines are portions of lines L of the definition. C is the vertex of star-point BCD of the star around $s^{2)}$. The exceptions to star-sharpedness, of Theorem 5.4, can occur on segments such as AB and EF.

In discussing phenomena that can occur for polyhedral star-shaped regions, we shall find it convenient to call the projection of cl $\hat{C}_i$ onto $\mathscr{S}_{k-1}$ the *star around* $s^{(i)}$. Of course, this may even be convex. However, in cases where this star contains long narrow protrusions into the remainder of $\mathscr{S}_{k-1}$ (a very qualitative notion), it will be useful to call these *star-points*. We will be more precise when $k = 3$: a triangular star-point of the star about $s^{(i)}$ is any triangle of that star two sides of which are on the boundary of the star. The angle between them is at the *vertex* ot the star point. Some of these definitions are illustrated in Figure 1.

We also require a slight extension of these notions. The motivation is that, if a sequence of procedures with polyhedral star-shaped (positive or not) decision partitions, with bounded number of sides on the polyhedra, converges regularly to a limiting procedure, that limit clearly has the following *property* A: it is equivalent ($\hat{\nu} \times \mu^1$) to a procedure whose decision partition differs from (5.1) on at most a finite number of sets of the form $H \times I$ where each H is a $(k-2)$-dimensional hyperplane through some $s^{(i)}$. We call any decision partition of $\mathscr{T}_k$ (positive or not, polyhedral or not) *almost star-shaped* if it has property A. In the motivating polyhedral case, that part H' of such an H on which (5.1) is violated typically arises as the limit of a union of sharp star-points of a star around a given $s^{(i)}$, which union is in a narrowing dihedral angle (about $s^{(i)}$) approaching H. If $\hat{\nu}$ is absolutely continuous with respect to μ^{k-1} on $\mathscr{S}_{k-1}$, every almost star-shaped region is equivalent ($\hat{\nu} \times \mu^1$) to a star-shaped region.

It is to be noted that the theorems and corollaries of this section characterize essentially complete classes in terms of decision partitions. The class is obtained by superimposing arbitrary conditioning partitions $\{\hat{C}^b\}$ (subject to restriction to $\mathscr{C}^+$ or $\mathscr{C}^L$ if appropriate) on the decision partitions $|\hat{C}|$ with the stated property of star-shapedness.

We shall sometimes save space by writing "star-shaped $\hat{C}$" for "$\hat{C}$ with star-shaped $|\hat{C}|$," etc.

For $\mathscr{C}^{L+}$, our parallel to Corollary 3.4 is

THEOREM 5.2. *For the k-sample problem where all f_i have the same domain of positivity, an essentially complete class in $\mathscr{C}^{L+}$ in the sense of* (2.4) *consists of all procedures $\hat{C}$ in $\mathscr{C}^{L+}$ with positive almost polyhedral star-shaped decision partitions $|\hat{C}|$ of $\mathscr{T}_k$.*

As in Corollary 3.4, almost star-shaped decision partitions $|\hat{C}|$ are in $\mathscr{C}^{1+}$ but require further conditions to insure $\hat{C}$ in $\mathscr{C}^+$. The reason for asserting "essential completeness" rather than "completeness" is that the reduction from $\mathscr{X}$ to $\mathscr{T}^k$ may eliminate some procedures equivalent in the sense of (2.4).

If the f_i do not have the same support, an admissible partition in $\mathscr{C}^{L+}$ can have $\hat{C}_i = s^{(i)} \times I$ for some i, and thus not satisfy (3.9)(i); Theorem 5.2 can be restated for this case. For $\mathscr{C}^+$ (general B) we are unable, even when all f_i have the same support, to verify (3.9)(i) for admissible $\hat{C}$ with the present proof, and can only improve slightly on the result for $\mathscr{C}$. We now state the latter (without the assumption of Theorem 5.2):

THEOREM 5.1. *For the k-sample problem, an essentially complete class in* $\mathscr{C}$ *in the sense of* (2.4) $\wedge$ (2.6), *and hence in the sense of* (2.4) $\wedge$ (2.8), *consists of the procedures* $\hat{C}$ *for which* $|\hat{C}|$ *is in the regular closure of the polyhedral positive star-shaped partitions of* $\mathscr{T}_k$.

The shortcoming of this result over that of Theorem 5.2 will be discussed in Remark 1 at the end of this section: more extensive randomization may be required in some of the procedures of the Theorem 5.1 essentially complete class, than in star-shaped procedures.

In the 2-hypothesis problem, the star-shapedness of (5.1), as well as almost star-shapedness, reduces to division of $\mathscr{S}_1$ into two intervals with randomization on the boundary of the two intervals, but never any randomization at an $s^{(i)}$ if that is the boundary. These are exactly the $\mathscr{C}^1$-admissible partitions for $|\hat{C}|$, and Theorem 5.1 combined with the results of Section 3 thus yields the first sentence of the following:

COROLLARY 5.3. *For the* 2-*hypothesis problem, an essentially complete class of admissible procedures in* $\mathscr{C}$ *in the sense* (2.4) $\wedge$ (2.6) *or* (2.4) $\wedge$ (2.8) *consists of all partitions in* $\mathscr{C}$ *whose underlying* $\mathscr{C}^1$-*partitions have the interval structure* (5.1). *This conclusion remains true if* $\mathscr{C}$ *is replaced everywhere by* $\mathscr{C}^L$, *or by* $\mathscr{C}^+$ *or* $\mathscr{C}^{L+}$ *with* (2.4) $\wedge$ (2.8) *replaced by* (2.4).

The last sentence of this corollary is obvious for $\mathscr{C}^+$, and an examination of that part of the proof of Theorem 5.1 that pertains to $\mathscr{C}^L$ shows the validity of the remainder when $k = 2$.

When $k = 3$, we cannot argue for the analogue of Corollary 5.3, but we can somewhat sharpen Theorem 5.1:

COROLLARY 5.4. *For the* 3-*hypothesis problem, an essentially complete class in* $\mathscr{C}$ *in the sense of* (2.4) $\wedge$ (2.6), *and hence in the sense of* (2.4) $\wedge$ (2.8), *consists of all procedures* $\hat{C}$ *for which* $|\hat{C}|$ *is an almost star-shaped partition of* $\mathscr{T}_3$, *and for which the exceptions to star-shapedness lie on at most the three lines* $\{s_i = 0\}$ *and one additional line through a single* $s^{(i)}$, *part of which borders the stars of the other two* $s^{(j)}$.

This will be proved below.

PROOF OF THEOREM 5.1. We shall work entirely in terms of partitions of $\mathscr{T}_k$ (which are equivalent to η's on $\mathscr{S}_{k-1}$ as discussed earlier), and therefore simplify notation by dropping carats. Let $\bar{C}$ be a given partition of $\mathscr{T}_k$. Our steps in finding a $\bar{\bar{C}}$ that is at least as good as $\bar{C}$ in the sense of (2.4) $\wedge$ (2.6), and that satisfies (5.1), are these: (a) For each integer $n > 0$, there is by Lemma 4.5 a finite L_n and a procedure ${}^n\bar{C}$ in $\mathscr{C}^{L_n}$ such that $||\bar{G}_\omega - {}^n\bar{G}_\omega||$ and $||\bar{H}_\omega - {}^n\bar{H}_\omega||$ are $< n^{-1}$ for all ω, where the norm is that of Section 4. (b) For each fixed n, there is a partition ${}^n\tilde{C}$ in $\mathscr{C}^{L_n}$ that is at least as good as ${}^n\bar{C}$ in the sense of

(2.4) $\wedge$ (2.6), and that is a regular limit, as $m \to \infty$, of a sequence $\{^{(n,m)}C\}$ of procedures with positive polyhedral star-shaped decision partitions. (c) We may assume the $^{(n,m)}C$ are canonical, since that is merely a relabeling (with possible lumping) of B and yields the same operating characteristics. For each fixed n, a subsequence $^{(n,m')}C$ converges regularly and its operating characteristics $\mathbf{G}$, $\mathbf{H}$ converge weakly by Theorem 4.4. A diagonalization of (n, m') produces a subsequence $\{^N\bar{\bar{C}}\}$ of the $^{(n,m')}C$'s that, by Theorem 4.4, converges regularly to a partition $\bar{\bar{C}}$ of $\mathscr{T}_k$, with $^N\bar{\bar{\mathbf{G}}}$, $^N\bar{\bar{\mathbf{H}}}$ converging weakly to $\bar{\bar{\mathbf{G}}}$, $\bar{\bar{\mathbf{H}}}$. The conclusion of Theorem 5.1 follows from this last together with (a) and (b), and it remains to prove the latter.

For fixed L, we shall show that the admissible procedures in $\mathscr{C}^L$ are regular limits of procedures with positive polyhedral star-shaped decision partitions. Since the operating characteristics of procedures in $\mathscr{C}^L$ are compact (Theorem 4.4), this yields (b). Suppose that, for two partitions $\tilde{C}$ and $\tilde{\tilde{C}}$ of $\mathscr{T}_k$ in $\mathscr{C}^L$ we have

$$P_i\{\tilde{\tilde{C}}_i^{\,b}\} \geqq P_i\{\tilde{C}_i^{\,b}\} \quad \forall\, b, i\,, \qquad P_i\{\tilde{\tilde{C}}^b - \tilde{\tilde{C}}_i^{\,b}\} \leqq P_i\{\tilde{C}^b - \tilde{C}_i^{\,b}\} \quad \forall\, b, i\,. \tag{5.2}$$

Then, whenever $P_i\{\tilde{C}_i^{\,b}\}$ is positive, so is $P_i\{\tilde{\tilde{C}}_i^{\,b}\}$. Since $1 - (\Gamma_i^{\,b})^{-1} = P_i\{C^b - C_i^{\,b}\}/P_i\{C_i^{\,b}\}$, (5.2) implies that $\tilde{\tilde{\Gamma}}_i^{\,b} \geqq \tilde{\Gamma}_i^{\,b}$. Hence, $\tilde{\tilde{C}}^b$ contributes at least as large an atom of probability as does $\tilde{C}^b$ to the sub-df $G_i(1-t)$, and that atom corresponds to at least as large a value of Γ_i. Thus, (5.2) implies $\tilde{\tilde{G}}_i(t) \geqq \tilde{G}_i(t)\ \forall\, i, t$, and similarly $\tilde{\tilde{H}}_i(t) \leqq \tilde{H}_i(t)\ \forall\, i, t$. Moreover, strict inequality holds somewhere in one of these last two sets of inequalities if it holds somewhere in (5.2). We conclude that (2.4) $\wedge$ (2.6)-admissibility in $\mathscr{C}^L$ implies (5.2)-admissibility in $\mathscr{C}^L$, and it remains to show that (5.2)-admissible partitions in $\mathscr{C}^L$ have the structure stated at the outset of this paragraph.

For C a $\mathscr{C}^L$-partition of $\mathscr{T}_k$, let r_c be the point in R^{2kL} with coordinates $P_i\{C_i^{\,b}\}$, $-P_i\{C^b - C_i^{\,b}\}$ for $i \in \Omega$, $b \in B_L$. These points can be analyzed using standard decision theoretic results. The set of all such r_c's is convex and compact, and ([7], Theorem 3.19) the set of (5.2)-admissible elements is contained in the closure of the set of all "anti-Bayes" elements with respect to strictly positive prior vectors; such an anti-Bayes element with respect to a $2kL$-vector $\{a_i^{\,b}, \bar{a}_i^{\,b}\}$ of positive coordinates is a partition in $\mathscr{C}^L$ that *maximizes*

$$\textstyle\sum_{i,b} [a_i^{\,b} P_i\{C_i^{\,b}\} - \bar{a}_i^{\,b} P_i\{C^b - C_i^{\,b}\}]\,. \tag{5.3}$$

The representation of the partition C of $\mathscr{T}_k$ as a randomized procedure η on $\mathscr{S}_{k-1}$ is

$$\eta(i, b \mid s) = \textstyle\int_0^1 I_{C_i^{\,b}}(s, u)\, du \tag{5.4}$$

as the (randomization) probability, given $S = s$, that $(\delta, Z) = (i, b)$; of course $\sum_{i,b} \eta_c(i, b \mid s) = 1$. Write

$$H_i^{\,b}(s) = a_i^{\,b} s_i - \textstyle\sum_{j \neq i} \bar{a}_j^{\,b} s_j\,. \tag{5.5}$$

We can then rewrite (5.3) as

$$(5.6)\qquad \sum_{i,b} \int_{\mathscr{S}_{k-1}} [a_i^b \eta(i, b \mid s) - \bar{a}_i^b \sum_{j\neq i} \eta(j, b \mid s)] \hat{f}_i(s)\hat{\nu}(ds) = \sum_{i,b} \int_{\mathscr{S}_{k-1}} H_i^b(s)\eta(i, b \mid s)\hat{\nu}(ds) .$$

From (5.5) we see that, for fixed (i_0, b_0) and $i \neq i_0$, the hyperplane $\{s: H_{i_0}^{b_0}(s) = H_i^b(s)\}$ divides $\mathscr{S}_{k-1}$ into two regions with $s^{(i_0)}$ and $s^{(i)}$ in their interiors. Hence, the set

$$(5.7)\qquad M_{i_0}^{b_0} = \{s: H_{i_0}^{b_0}(s) \geqq H_i^b(s) \text{ for all } (i, b) \text{ with } i \neq i_0\}$$

is convex, polyhedral, and contains $s^{(i_0)}$ in its interior. We conclude that $M_{i_0} = \bigcup_b M_{i_0}^b$ is the i_0-star of a positive polyhedral star-shaped partition of $\mathscr{S}_{k-1}$, and upon replacing the weak inequality by strict inequality in (5.7) one obtains sets whose union over b_0 is int M_{i_0}. A standard argument on (5.6) shows that, a.e. $\hat{\nu}$, every maximizer of (5.3) satisfies, for each i_0,

$$(5.8)\qquad \begin{aligned} \eta(i_0, B_L \mid s) &= 1 \quad \text{if} \quad s \in \operatorname{int} M_{i_0} , \\ &= 0 \quad \text{if} \quad s \notin M_{i_0} . \end{aligned}$$

Thus, every such η is equivalent to a positive polyhedral star-shaped partition of $\mathscr{T}_k$. □

PROOF OF THEOREM 5.2. Although the set $\{r_C: C \in \mathscr{C}^{L+}\}$ is not closed, a trivial geometric observation shows that, for each $\bar{C}$ in $\mathscr{C}^{L+}$, there is an anti-Bayes partition $\bar{\bar{C}}$ in $\mathscr{C}^{L+}$ (relative to $\{a_i^b, \bar{a}_i^b\}$ of nonnegative components, some but not all of which may now be zero) that is at least as good as $\bar{C}$ in the sense of (5.2) and hence (2.4). Moreover, the development of the previous paragraphs shows that these $\bar{\bar{C}}$ are all limits of sequences procedures with positive polyhedral star-shaped decision partitions having a bounded number of faces, and hence they have decision partitions that are almost polyhedral star-shaped. It remains to show that these anti-Bayes partitions $\bar{\bar{C}}$ in $\mathscr{C}^L$ have $|\bar{\bar{C}}|$ *positive* star-shaped; we must verify (3.9)(i) for them.

Fix i_0. There are three cases to consider. (i) If there is a b_0 for which $a_{i_0}^{b_0} > 0$, then clearly $s^{(i_0)} \in \operatorname{int} M_{i_0}^{b_0}$. (ii) If there is a b' for which $\bar{a}_{i_0}^{b_0'} > 0$, then for every b and $i \neq i_0$ the set M_i^b excludes $\{s: H_i^b(s) < H_{i_0}^{b'}(s)\}$, and the latter clearly contains a neighourhood of $s^{(i_0)}$. (iii) If no $a_{i_0}^{b_0}$ or $\bar{a}_{i_0}^{b_0'}$ is positive, then (since some a_i^b or $\bar{a}_i^b$ is positive) one computes easily that M_{i_0} contains no points of $\{s: s_i > 0\ \forall\, i\} = \mathscr{S}_{k-1}^*$ (say). But $\hat{\nu}(\mathscr{S}_{k-1} - \mathscr{S}_{k-1}^*) = 0$ if all f_i have the same support, so $\hat{\nu}(M_{i_0}) = 0$ in this last case. Thus, either there is an i_0 for which (iii) holds, in which case the resulting anti-Bayes partition is not in $\mathscr{C}^{L+}$, or else for each i_0 $(1 \leqq i_0 \leqq k)$ either (i) or (ii) holds, in which case the resulting anti-Bayes partition is in $\mathscr{C}^{L+}$. □

PROOF OF COROLLARY 5.4. Fix $\varepsilon > 0$. Let C have a $|C|$ that is a positive star-shaped partition of $\mathscr{T}_3$, and let V be any triangular star-point of the star about $s^{(1)}$, with vertex P. Suppose cl V is surrounded by C_2, in a neighborhood of P. The angle θ of cl V facing $s^{(1)}$ at P, because of (5.1) with $i = 1$, has the property that it contains part of the segment $s^{(1)}P$ near P; but it is also such that its side closest to $s^{(3)}$, when extended beyond P away from $s^{(1)}$, intersects the segment $s^{(1)}s^{(2)}$, because of (5.1) with $i = 2$. If $P \in N_\varepsilon = \{s\colon \text{all } s_i \geqq \varepsilon\}$, elementary trigonometry shows that $\theta \geqq 2\tan^{-1}(2\varepsilon/3^{\frac{1}{2}}) \geqq c_0\varepsilon$, where $c_0 > 0$. It follows (essentially from compactness of graphs satisfying the same Lipschitz condition) that, for each sequence ${}^n C$ of procedures with positive polyhedral star-shaped decision partitions of $\mathscr{S}_2$ and each $\varepsilon > 0$, there is a subsequence that converges regularly on N_ε to a procedure with star-shaped decision partition on N_ε, except for at most a single exceptional H in the definition of "almost star-shaped": for each n, it is easy to see that at most one ${}^n C_{i_0}$ can have a star-point V each of whose two sides is bordered near the vertex by a *different* one of the other two C_i. This single V alone can be in N_ε and have an angle $\theta < c_0\varepsilon$, and in the limit a subsequence of such V's can collapse toward an H. A subsequence $\varepsilon_m \to 0$ and diagonalization completes the proof, in view of Theorem 5.1. □

Remarks on the necessary conditions for admissibility.

1. For $k = 2$ or 3, elementary geometry has shown (Corollaries 5.3 and 5.4) that taking the regular closure in Theorem 5.1 does not introduce much randomization over that present in Theorem 5.2 for $\mathscr{C}^{L+}$. We have been unable to verify the corresponding fact for $k = 4$, although it is natural to conjecture that such randomization should not be necessary. The difficulty is that, viewing the boundary of a polyhedral star $\hat{C}_{i_0}$ as the graph of distance from $s^{(i_0)}$ (as a function of angular position), we might have a sequence of such graphs with increasing number of increasingly frequent oscillations (sharp star-points) whose magnitudes stay bounded away from zero. Such a sequence ${}^n\hat{C}$ could converge regularly to a procedure ${}^\infty\hat{C}$ for which $|{}^\infty\hat{C}|$ randomizes (i.e., for which $0 < \eta(i \times B \mid s) < 1$ on a set of positive μ^{k-1} measure in $\mathscr{S}_{k-1}$, similar to what occurs in $\mathscr{C}^1$ decision theory for sequences like (4.1)). When $k = 3$, the star-shapedness for $i \neq i_0$ limited the existence of such sharp points, for positive polyhedral procedures, and ${}^\infty\hat{C}$ could only differ from star-shapedness on one line segment H (in the definition of almost star-shapedness). Moreover, that H was associated with the limit of a sequence of points of ${}^n\hat{C}_{i_0}$ that had each of the other two ${}^n\hat{C}_i$'s on one of its sides. When $k = 4$ the failure of this last property seems to make possible the existence of a large number of sharp points of nonvanishing magnitude in ${}^n\hat{C}_{i_0}$. To see what can happen, we now give an example, when $k = 4$, of a ${}^n\hat{C}_1$ that has a single large sharp point a large portion of which is completely surrounded by ${}^n\hat{C}_2$. It is not clear at this writing that one cannot have such an increasing number of points that ${}^\infty\hat{C}$ has the undesirable extent of randomization noted above. If so, we still do not know whether such an eventuality must be part of a complete class for general B, or whether the difficulty lies in the reduction to using (5.2) as discussed in Remark 2 below, so that such ${}^\infty\hat{C}$ could be eliminated by a different admissibility proot.

For this example, suppose $L = 3$. The star about $s^{(1)}$ will be our concern. It will be convenient notationally to label the elements of B as $\{1, 3, 4\}$. To simplify the arithmetic, we make most of the 24 coefficients zero; adding a small positive value to all of them does not change the substance of the example. We now construct ${}^rC$ for r large. Let ε be small; for definiteness, take $r\varepsilon = \frac{1}{2}$. Let

$$a_1^1 = a_2^4 = a_2^3 = 1 ,$$
$$\bar{a}_1^1 = \bar{a}_2^4 = \bar{a}_2^3 = r^2 ,$$
$$\bar{a}_3^1 = \bar{a}_4^1 = r(\tfrac{1}{2} - \varepsilon) ,$$
$$\bar{a}_3^4 = \bar{a}_4^3 = r ,$$

and let all other a_i^b and $\bar{a}_i^b$ be zero. A simple geometric picture may be obtained as follows: Define, in $\mathscr{S}_3$, the lines

$$L' = \{s: s_1 + s_2 = 1\} ,$$
$$L'' = \{s: s_3 + s_4 = 1\} ,$$
$$L = \{s: s_1 = s_2, s_3 = s_4\} ,$$
$$L^* = \{s: s_3 = s_4, s_2 - s_1 = 2r\varepsilon s_3\} ,$$

the planes

$$\Pi^1 = \{s: s_1 = s_2\} ,$$
$$\Pi^4 = \{s: H_1^1(s) = H_2^4(s)\} = \{s: (s_2 - s_1) = r[(\tfrac{1}{2} + \varepsilon)s_3 - (\tfrac{1}{2} - \varepsilon)s_4]\} ,$$
$$\Pi^3 = \{s: H_1^1(s) = H_2^3(s)\} = \{s: (s_1 - s_2) = r[(\tfrac{1}{2} - \varepsilon)s_3 - (\tfrac{1}{2} + \varepsilon)s_4]\} ,$$

and the points

$$p' = (\tfrac{1}{2}, \tfrac{1}{2}, 0, 0) ,$$
$$p'' = (0, 0, \tfrac{1}{2}, \tfrac{1}{2}) ,$$
$$p^* = (\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}) ,$$
$$p_3 = (0, 0, \tfrac{1}{2} - \varepsilon, \tfrac{1}{2} + \varepsilon) ,$$
$$p_4 = (0, 0, \tfrac{1}{2} + \varepsilon, \tfrac{1}{2} - \varepsilon) .$$

Take L' to be vertical with $s^{(2)}$ at its top. L' is bisected by Π^1, which contains L and L'', with $L \cap L' = p'$ and $L \cap L'' = p''$. The points p_2 and p_3 are very close to p'', and hence the lines $\Pi^4 \cap \Pi^1$ and $\Pi^3 \cap \Pi^1$ are close to L. On the other hand, the line $L^* = \Pi^4 \cap \Pi^3$ is not everywhere close to L; e.g., the point on L^* above p^* is $((1 - r\varepsilon)/4, (1 + r\varepsilon)/4, \frac{1}{4}, \frac{1}{4})$. Hence, the region V above Π^1, under Π^4 and Π^3, and with $\frac{1}{4} \leqq s_3 + s_4 \leqq \frac{3}{4}$, is a very thin wedge of length $\frac{1}{2}$ and thickness $< 2\varepsilon$, whose height is $r\varepsilon/4 = \frac{1}{8}$ at p^*.

The coefficients a_i^b and $\bar{a}_i^b$ have been chosen so that $\operatorname{int} V \subset {}^rC_1$ but so that the region just on the other side of Π^4 and Π^3 is in ${}^rC_2$, with the part bordering Π^b in ${}^rC_2^b$. To see this, one computes that $H_1^1 - H_2^b$ has the right behavior at Π^b (for $b = 4, 3$), and that $H_2^4(s) - H_2^3(s) \geqq 0 \Leftrightarrow s_4 \geqq s_3$. All that remains is then to verify that $H_i - H_i^b > 0$ on a neighbourhood of V for all (i, b) other than (2, 3) and (2, 4). This is the case because the dominating term in $H_1^1 - H_1^b$, of order r^2, is $\bar{a}_1^1 s_1$ in the case $b = 1$ and $\bar{a}_1^b s_2$ otherwise.

This completes the example.

2. Even in the restricted context of Theorem 5.2, except when $k = 2$ it is generally false that the star-shaped partitions are all admissible, even if they are positive and polyhedral. The main cause of this is the use of (5.2), in terms of which it is more difficult for a given $\tilde{C}$ to be improved upon by a $\tilde{\tilde{C}}$ than in the original sense of (2.4) $\wedge$ (2.6), because the improvement must be *on each* $\tilde{C}^b$. In terms of G_ω or H_ω, we can have domination without it occuring for each b. For example, one can imagine $L = 2$ and $\tilde{G}_1$ with jumps of .3, .3 at .5, .7 while $\tilde{\tilde{G}}_1$ has jumps of .2, .4 at .6, .8; then $\tilde{G}_1 \leqq \tilde{\tilde{G}}_1$, but (5.2) cannot be satisfied for any labeling of C_i^b's. (It is easily seen that inadmissibility of $\tilde{C}$ in the sense of (5.2) does permit any relabeling of $\tilde{\tilde{C}}$ by permutation of B while fixing the labels for $\tilde{C}$.) The difficulty is also reflected in the fact that the operating characteristics (**G**, **H**) of $\mathscr{C}^L$ partitions are not a convex set, unlike the set of vectors r_C.

3. We now mention briefly some possible improvements on the necessary conditions of this section. There are special cases where small support of $\hat{\nu}$ can yield superficially better descriptions of admissible $\hat{C}$ that cannot hold in general. More interesting improvements may be sought in the characterization of $|\hat{C}|$ or in the breakup into $\{\hat{C}^b\}$. For example, the development of (5.3)—(5.7) for $\mathscr{C}^L$ yields C_i^b with convex interiors provided that the a_i^b and $\bar{a}_i^b$ are so different that $H_{i_0}^{b'}(s) = H_{i_0}^{b''}(s)$ only on a hyperplane of $\mathscr{S}_{k-1}$. This is always the case if the star around $s^{(i_0)}$ has L star-points. If, instead, $a_{i_0}^{b'} = a_{i_0}^{b''}$ and $\bar{a}_j^{b'} = \bar{a}_j^{b''}$ for $j \neq i_0$, then any measurable breakup of $M_{i_0}^{b'} - \bigcup_{b \neq b', b''} M_{i_0}^b$ can be used for $C_{i_0}^{b'}, C_{i_0}^{b''}$; they need not be convex. Such a degeneracy in the $\bar{a}_j^b$ and $a_{i_0}^b$ is reflected in M_{i_0} being a union of fewer than L convex sets. When C_{i_0} has fewer than L star-points (or analogous points on the boundary of $\mathscr{S}_{k-1}$), our development yields little about the breakup into $\{C_{i_0}^b\}$ outside of a neighborhood of the boundary of C_{i_0}. In short, the greatest possible departures from convexity (most points on stars) of the $\hat{C}_i$ can occur only with a breakup into the simplest $\hat{C}_i^b$ (those with convex closure).

4. It is unclear to what extent the star-shapedness can be reduced in Theorem 5.2 (aiming toward less departure from convexity), but simple examples when $k = 3$ and $L = 2$ show that it cannot be eliminated for the sense (5.2). (For example, if $a_3^1 = a_3^2 = \bar{a}_1^1 = \bar{a}_2^2 = 1$ and all other coefficients are ε, small and positive, then C_3 differs little from the set $\{s: s_3 > \min(s_1, s_2)\}$.) A different proof, taking account of Remark 2, would apparently be needed for significant reduction. One possibility that the author has not been able to implement is the maximization of $\sum_i \int h_i(t) d[-G_i(t)]$ (together with a linear combination reflecting (2.6) or (2.8)) for increasing functions other than the $h_i(t) = -c_i t^{-1}$ used in the sufficient condition of Section 3. For another possibility, we refer the reader to an "exchange" argument in the case $k = 2$ that appears in [2], and in which points of a $C_1^{b'}$ can be exchanged with those of larger s_1 in a $C_2^{b''}$ to

improve both G_i; this argument fails to yield convexity for $k > 2$, and thus far has yielded only minor information bounding the slope of the boundary of the stars as a function of position in $\mathscr{S}_{k-1}$.

The use of (2.6) rather than (2.8) alone in the proof, even when (2.4) $\wedge$ (2.8) is the criterion of admissibility, reflects the difficulty of treating the $P_\omega\{Q_\omega\}$ as part of a vector like r_c.

5. One *can* introduce a more detailed operating characteristic, and accompanying notion of admissibility, relative to which the star-shaped anti-Bayes procedures of the present section become admissible except for modifications on the border of $\mathscr{S}_{k-1}$ analogous to those of [8] (so that, for example, one must have $\hat{C}_i \cap \{(s, u) : s_i = 0\} = \emptyset$). Such a more detailed operating characteristic for $\mathscr{C}^L$ is given by the equivalence class $r_c{}^*$ of all vectors obtained from r_c by permutations (relabeling) of B^L; the partition $\bar{C}$ is at least as good as $\bar{\bar{C}}$ in this sense if there is a permutation σ for which

$$(5.9) \qquad P_i\{\bar{C}_i{}^b\} \geqq P_i\{\bar{\bar{C}}_i{}^{\sigma b}\} \qquad \text{and} \qquad P_i\{\bar{C}^b - \bar{C}_i{}^b\} \leqq P_i\{\bar{\bar{C}}^{\sigma b} - \bar{\bar{C}}_i{}^{\sigma b}\} \quad \forall\, b, i\,.$$

Equivalently, one can adjoin to (5.9) the condition $\bar{\Gamma}_i{}^b \geqq \bar{\bar{\Gamma}}_i{}^{\sigma b}$ $\forall\, b, i$, since the latter follows from (5.9). Then the star-shaped partitions that are anti-Bayes relative to strictly positive vectors (as in our development (5.3)—(5.8)) are admissible; and so are the star-shaped partitions obtained as limits of these, except for the indicated restrictions on the border of $\mathscr{S}_{k-1}$. Remark 1 is however still relevant.

This criterion seems intuitively too "fine," and yields too many admissible procedures to be taken as justification for blindly using an arbitrary (positive) anti-Bayes procedure. The G_i obtained by "lumping" of $\{\Gamma_i{}^b, P_i\{C_i{}^b\}, b \in B\}$ reflect the notion of goodness in a preferred degree of detail, and represent the conditional confidence aim better than does $r_c{}^*$.

6. Since the results for $k \geqq 4$ are more informative for $\mathscr{C}^L$ than for $\mathscr{C}$, the use of finite approximations of general partitions ((a) of proof of Theorem 5.1) was expeditious. However, anti-Bayes procedures for general B can also be computed directly, the role of r_c being induced by the probability measures on $B \times D$ determined by rectangles, i.e., by $P_i\{Z \in A, \delta = d\}$ for $A \in \mathscr{B}_0$ and $d \in D$. As in the case of $\mathscr{C}^L$, this yields a weaker notion of admissibility than (2.4) $\wedge$ (2.6). Thus far we have not been able to reduce the huge complete class for $\mathscr{C}$ through this approach.

7. The remarks of Section 4 on elimination of randomization can be used to simplify the results of Theorem 5.2 and Corollaries 5.3 and 5.4. However, when applied to Theorem 5.1 they yield nonrandomized partitions whose structure is quite unclear.

6. Other loss and risk structures. We mention briefly here only four of the many possible modifications to using (2.4) $\wedge$ (2.6) with the implied 0–1 loss structure from decisions in D_ω and its complement when ω is true.

(A) In Section 3, if we replace the conditional probability $\Gamma_i{}^b$ by a *conditional* *gain* $W_i^b = \sum_{j \in D_i} w_{ij} P_i\{K_j^b\}$ where the given numbers w_{ij} are positive for $j \in D_i$, we see that the proofs of that section are easily modified, *provided* w_{ij} *is independent of* j *in* D_i. In the k-hypothesis setting of Section 5, consideration of $w_{ii}\Gamma_i^b$ in place of Γ_i^b does not alter the conclusions. If performance is measured instead in terms of conditional expected *loss* $\sum_{j \notin D_i} m_{ij} P_i\{C_j^b\}$ where $m_{ij} > 0$ for $j \notin D_i$, the conclusion of Section 3 remains valid if the m_{ij} are equal for $j \notin D_i$; the development of Section 5 proceeds with replacement of r_c by the $2kL$-vector $\{\sum_{j \neq i} m_{ij} P_i\{C_j^b\}, -P_i\{C^b\}\}$, and the changes thereafter are obvious.

(B) For brevity, the next modification will be described only for the k-hypothesis problem and in terms of the original 0–1 losses. A more detailed picture of conditional probabilities of *incorrect* decisions than is present in the H_ω of (2.5) can be given for $j \neq i$, in terms of the functions

$$H_{ij}(t) = P_i\{\delta = d_j; \Gamma_{ij} > t\}, \tag{6.1}$$

where

$$\Gamma_{ij} = P_i\{C_j \mid \mathscr{B}_0\}. \tag{6.2}$$

The Γ_{ij} have an obvious frequentist interpretation, and in practice if $Z(X) = b_0$ and $\delta(x) = d_{j_0}$ the vector $\{\Gamma_{ij_0}^{b_0}, i \neq j_0\}$ of confidences associated with the states i for which d_{j_0} is incorrect, could be stated in addition to $\Gamma_{j_0}^b$. Also define, in analogy with (1.3),

$$Q_{ij} = \{x : \Gamma_{ij}(x) = 0\}. \tag{6.3}$$

Intuitively, we prefer the $P_i\{Q_{ij}\}$ to be large rather than small. Thus we are led to the possibility of defining $\bar{\bar{C}}$ to be at least as good as $\bar{C}$ if, for all unequal i and j,

$$\bar{\bar{H}}_{ij}(t) \leqq H_{ij}(t) \quad \forall\, t \tag{6.4}$$

and

$$P_i\{\bar{\bar{Q}}_{ij}\} \geqq P_i\{\bar{Q}_{ij}\}. \tag{6.5}$$

(It is unnecessary to adjoin (2.4) $\wedge$ (2.8), but the conclusion below will be the same with that addition.)

In the development of Section 3, we now replace classical $\mathscr{C}^1$-admissibility of $|C|$ by admissibility in the sense of the classical vector risk $\{P_i\{C_j\}, i \neq j\}$, in terms of which $\bar{\bar{C}}$ is at least as good as $\bar{C}$ if $P_i\{\bar{\bar{C}}_j\} \leqq P_i\{\bar{C}_j\}$ for all unequal i, j. (Alternatively, this sense of admissibility may be replaced by one that implies it, that there is some set of losses $m_{ij} > 0$ for which $|C|$ is admissible for the risk function $r(i) = \sum_{j \neq i} m_{ij} P_i\{C_j\}$.) In analogy with (3.1)—(3.2), we obtain $P_i\{C_j; \Gamma_{ij} = 0\} = 0$ and $-\int_{0+}^{\infty} t^{-1} d[-G_{ij}(t)] = P_i\{Q_{ij}\} - 1$. Parallel to the development of the paragraph following (3.4), we have (6.4) at $t = 0$ yielding $P_i\{\bar{\bar{C}}_j\} \leqq P_i\{\bar{C}_j\}$ for all $i \neq j$; if $|\bar{C}|$ is classically vector-admissible, these must be equalities. The remainder of the proof parallels (3.5) and (3.6), leading to the conclusion that $|\bar{C}|$ vector-admissible implies $\bar{C}$ admissible in the sense induced by (6.4) $\wedge$ (6.5).

The sense (6.4) $\wedge$ (6.5) seems somewhat less ad hoc than that of the $r_c{}^*$ of Section 5 (Remark 5), but it would be more natural still to work with the law of Γ_j under state $i \neq j$, in place of that of Γ_{ij}. A useful admissibility result has not been obtained for the resulting criterion because of the failure thus far to find a substitute for (3.2).

The development of Section 5 for the criterion (6.4) is straightforward.

(C) Another consideration is a measure of "size" of a decision, analogous to length of a confidence interval. For example, the making of a decision d for which Ω_d contains many elements should intuitively be penalized in terms of some measure of the size of Ω_d. Let $e_j > 0$ be the "size" penalty incurred when d_j is made. The question of whether conditional or unconditional expectation of e is more suitable to consider, with illustrations, will be discussed in [2], [4]. For the moment, we treat unconditional expectation, and define $\bar{\bar{C}}$ to be at least as good as $\bar{C}$ if (2.4), (2.8), and

$$\sum_j e_j P_\omega\{\bar{\bar{\delta}} = d_j\} \leqq \sum_j e_j P_\omega\{\bar{\delta} = d_j\} \quad \forall\, \omega \tag{6.6}$$

hold. The corresponding $\mathscr{C}^1$ notion for $|C|$ is (6.6) together with $P_\omega\{\bar{\bar{\delta}} \in D_\omega\} \geqq P_\omega\{\bar{\delta} \in D_\omega\}\ \forall\, \omega$. If $|\bar{C}|$ is $\mathscr{C}^1$-admissible but $\bar{\bar{C}}$ is better than $\bar{C}$ in the sense of (2.4) $\wedge$ (2.8) $\wedge$ (6.6), either (i) G_ω, $P_\omega\{Q_\omega\}$ are the same for both procedures $\forall\, \omega$, in which case the $\mathscr{C}^1$-admissibility of $|\bar{C}|$ is contradicted by the strict inequality holding somewhere in (6.6); or else (ii) equality holds in (6.6), in which case the proof of Section 3 again leads to a contradiction. Thus, Theorems 3.1 and 3.3 and Corollary 3.2 remain valid for this notion of admissibility.

In Theorem 3.4, a somewhat larger class is now obtained than would have been obtained in Section 3 for the general (not necessarily k-hypothesis) decision problem. The (2.4) $\wedge$ (2.8)-admissible procedures are still admissible, but we also obtain the "anti-Bayes" partitions which, for any set of positive values $\{a_\omega, \bar{a}_\omega\}$, maximize

$$\sum_\omega a_\omega P_\omega\{\delta \in D_\omega\} - \sum_{\omega, j} \bar{a}_\omega P_\omega\{d_j\} e_j\,. \tag{6.7}$$

The standard argument then yields $\delta(x)$ to be a value j maximizing (a.e. ν)

$$\sum_{\omega \in \Omega_j} a_\omega f_\omega(x) - e_j \sum_\omega \bar{a}_\omega f_\omega(x)\,, \tag{6.8}$$

which is easily translated into the language of Theorem 3.4. The interior of each region $\delta^{-1}(d_j)$ is convex and polyhedral, but is of a more complex form than that obtained for (2.4) $\wedge$ (2.8), where all $\bar{a}_\omega = 0$ in (6.8).

The developments analogous to those of Section 5 proceed in a similar manner for (2.4) $\wedge$ (2.6) $\wedge$ (6.6) to those for (2.4) $\wedge$ (2.6).

A similar analysis applies for admissibility in terms of such operating characteristics as (analogous to G_ω) the law of $\Gamma_\omega(x) - e_{\delta(x)}$ on $\delta^{-1}(D_\omega)$ and of $-e_{\delta(x)}$ on the complement of $\delta^{-1}(D_\omega)$, when ω is true.

Illustrations of these approaches for ranking and selection problems are contained in [4].

It would be interesting to treat losses that are general functions of the Γ_ω. As indicated in Remark 4 of Section 5, we do not know how to do this.

(D) It has been suggested by Larry Brown that our admissibility in terms of the measure G_ω of performance of Γ_ω on $\delta^{-1}(D_\omega)$ is no more natural than that in terms of the df Π_ω of Γ_ω on all of $\mathscr{X}$. The present paper's considerations are motivated by the possible attractiveness to the practitioner of knowing that Γ_ω has a tendency to be large on the set where a correct decision is made, in contrast with the questionable appeal of its being large on the complementary set where the decision is incorrect. This can be thought of as a rudimentary loss consideration that attempts to have the same "shape" as that of a loss function L on $\Omega \times D \times I^\Omega$ that reflects the combined effect of decision and conditional confidence on loss, to yield a traditional unconditional risk computed as $E_\omega L(\omega, \delta(X), \Gamma(X))$ (which, as remarked in (c) above, we are unable to handle). On the other hand, Brown cites possible practical goals of meeting specified confidence coefficient standards, that could lead to his formulation. In any event, the author does not have very strong preference for any particular admissibility scheme to the exclusion of others that seem reasonable and lead to the identification of sensible conditional confidence procedures.

Brown's admissibility considerations are related to ours by $G_\omega(t) = \int_{\tau>t} \tau \, d\Pi_\omega(\tau)$ and $H_\omega(t) = \int_{\tau<1-t} (1 - \tau) \, d\Pi_\omega(\tau)$. These relations imply that admissibility as defined by replacing G by $-\Pi$ in (2.4) ($\bar{\bar{\Gamma}}_\omega$ stochastically larger than $\bar{\Gamma}_\omega$) implies (2.4)—(2.6)-admissibility, but there are additional admissible procedures in Brown's sense. His considerations relate to increasing convex functions c to be integrated with respect to Π_ω, the counterpart of our increasing functions integrated with respect to G_ω. He treats such problems as finding, for a given Z, the δ that maximizes a function of the quantities $E_\omega c(\min_\omega \Gamma_\omega)$, motivated by the idea that $\min_\omega \Gamma_\omega$ is all one can claim as guaranteed confidence for all ω. (This formalizes the spirit of the remarks in the paragraph below (1.2).) He obtains results on procedures with monotone structure, described near the end of Section 2. These results will appear in a paper by Brown.

Acknowledgment. The author is grateful to Larry Brown, both for the material of Section 6(D), and also for extensive comments that were used in rewriting this paper.

REFERENCES

[1] BAHADUR, R. R. and LEHMANN, E. L. (1955). Two comments on "Sufficiency and statistical decision function." *Ann. Math. Statist.* **26** 139–142.

[2] BROWNIE, C. and KIEFER, J. (1975). Conditional confidence statements. To appear.

[3] DVORETZKY, A., WALD, A. and WOLFOWITZ, J. (1951). Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. *Ann. Math. Statist.* **22** 1–21.

[4] KIEFER, J. (1975). Conditional confidence approach in multidecision problems. *Proc. Fourth Dayton Multivariate Conference*. To appear.

[5] LE CAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

[6] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probabilities.* Holden-Day, San Francisco.

[7] WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.

[8] WALD, A. and WOLFOWITZ, J. (1951). Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 149–158, Univ. of California Press.

DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853