The Annals of Statistics
1973, Vol. 1, No. 2, 241-250

ON SOME DIFFICULTIES IN A FREQUENCY THEORY OF INFERENCE

By Donald A. Pierce

Oregon State University

A study of relationships between confidence regions being Bayesian, and the existence of some generalizations of Fisher's notion of relevant subsets. For a betting scheme introduced by Buehler, and for finite parameter space, it is shown that non-Bayesian procedures allow a winning strategy for a statistician's adversary. It is further shown, for finite parameter space, non-Bayesian procedures must admit conditional confidence levels bounded away from the unconditional level, the converse to a theorem of Wallace. For general parameter space these results follow from a procedure not being weak Bayes in a certain sense.

1. Introduction. It is rather widely acknowledged that there are very serious logical difficulties involved in interpreting a single instance of statistical data through procedures whose justification is their performance in repetitions of the experiment. A discussion of some of these difficulties is given by Cox (1958). In particular suppose that a confidence interval procedure has exact confidence level α, but there exists a subset R of the sample space and an $\varepsilon > 0$ such that the conditional confidence level, given R, is less than $\alpha - \varepsilon$ (or greater than $\alpha + \varepsilon$) for all parameter values. Such a subset R was termed by Fisher a *relevant subset*, and his contention was that one should have "confidence" no more than $\alpha - \varepsilon$) (or no less than $\alpha + \varepsilon$) in intervals based on outcomes in R. In the same spirit Hacking (1965) argues that the Neyman-Pearson theory is appropriate for "before trials betting" whereas one should, in interpreting data, be more concerned with "after trials betting." The primary purpose of this paper is to study the existence of relevant subsets. The results are given in terms of confidence regions but similar results can be obtained for conditional risk functions in the decision model setting.

Some of the results given here are related to those of Wallace (1959) where it was shown that (proper) Bayes procedures cannot admit relevant subsets. The results here in Theorems 2 and 3 are in the converse direction; roughly that non-Bayesian procedures must admit a generalization of relevant subsets. For general (continuous as opposed to finite) parameter space, however, the hypothesis must be the stronger one that a procedure is not *weak* Bayes in a certain sense.

Other results here, in Theorems 1 and 3, are in terms of a betting scheme introduced by Buehler (1959), who formulated stronger than the usual frequency desiderata for confidence regions in order to cope with the difficulties alluded to above. The results here show that all non-Bayesian (non-weak-Bayesian for general parameter space) procedures fail to satisfy these desiderata. Bayesian

Received March 2, 1971; revised July 31, 1972.

methods satisfy the additional criteria given by Buehler but generally fail to satisfy the usual frequency requisites. Thus the usual frequency requirements are largely at cross-purposes with the additional desiderata. This tends, from a *frequency* viewpoint, to vitiate criticism of Bayesian methods for not having the usual frequency interpretation.

Results on betting schemes related to those given here were obtained by Cornfield (1969) and by Freedman and Purves (1969). Their results concern not conventional forms of inference, however, but schemes in which the statistician posts betting odds on every subset of the (finite) parameter space. Their results might be (and were in the published discussion of Cornfield's paper) mistakenly interpreted as showing that such a task is too ambitious.

2. Conditional reference sets. Frequency characteristics of statistical procedures are obtained by integrating with respect to each of a family of probability measures $\{P_\theta(\cdot), \theta \in \Omega\}$ on a sample space (X, B). The sample space, or an appropriate subset of it, will be called the reference set. The difficulty is that in most problems the reference set is wholly conceptual. There are distinct reference sets available for interpreting a single instance of data and no effective criteria for selecting from these. It is a well-known strength of Bayesian methods that this choice of reference set is largely irrelevant.

The motivation for this paper is to study the behavior of statistical procedures in conditional reference sets. For any instance of data $x \in X$ and any event $E \in B$ such that $x \in E$, one may consider using the conditional reference set with measures $\{P_\theta(\cdot \mid E), \theta \in \Omega\}$. Some examples are given below to indicate why one might choose to do this.

The decision to study conditional reference sets is not meant to imply that the choice of the original reference set is clear. For example consider a life-testing experiment to gain information about the parameter in a known family of life-distributions, and suppose than in the course of the experiment several units are lost before death due to extraneous reasons. A reference set for interpreting the data can only be obtained by specifying a mechanism for the censoring which will provide a model for censoring in further trials. There will generally be many such models which would agree with the censoring already done, and the reference set will depend crucially on the choice of one. It is even more striking, as pointed out vividly by Pratt (1961), that even if no censoring has occurred the reference set will depend upon what the experimenter *would* have done if, for example, the experiment had lasted beyond the time available. To a certain extent the choice of an original reference set can be studied within the context of conditional reference sets, for any set of distinct reference sets can be thought of as conditional sets within a composite reference set.

It is somewhat well accepted (see, for example Cox (1958)) that if E is an ancillary (similar) event, i.e. $P_\theta(E)$ does not depend upon θ, then the reference set conditional on E should be used. That this principle is inconsistent with the Neyman–Pearson theory is illustrated by Example 1 below. At any rate Basu (1964) has shown that this principle will not always lead to a well-defined reference set.

EXAMPLE 1. Let $x_{(1)} \cdots, x_{(n)}$ be the order statistics from a random sample from the uniform $(\theta, \theta + 1)$ distribution. The uniformly most accurate (inversion of the uniformly most powerful test) α-level lower confidence limit for θ is given by $l = \max\{x_{(1)} - c, x_{(n)} - 1\}$, where $(1 - c)^n = 1 - \alpha$. However, if E is the event that $x_{(n)} - x_{(1)} \geqq 1 - c$ then $P_\theta\{l \leqq \theta \mid E\} \equiv 1$ for all θ. Thus outcomes in E should certainly be interpreted through the conditional reference set. A similar argument suggests conditioning on any value of the ancillary statistic $x_{(n)} - x_{(1)}$.

This example is given because the confidence region given corresponds to a uniformly most powerful test, and hence any procedure obtained by conditioning on the ancillary statistic will be sub-optimal from the Neyman–Pearson viewpoint. Most examples given to illustrate the importance of conditioning on ancillary statistics are less forceful in this contrast because the Neyman–Pearson theory does not provide in them an optimum procedure. In fact Fisher's advocacy of the principle of conditioning on ancillary statistics was largely presented as a means of proceeding when there is no convenient sufficient statistic. For reasons to follow immediately it is not proposed that the reference set conditional on $x_{(n)} - x_{(1)}$ suggested in Example 1 is the appropriate one to use, rather only that its existence makes the use of the unconditional reference set inappropriate.

Although conditioning on ancillary events can be reasonably well-justified on the grounds that they are uniformative, it is difficult to envision a frequency theory which would cope with conditioning on other types of events. The following example given by Brown (1967), following discovery of phenomena of this type by Buehler and Fedderson (1963), is very disconcerting.

EXAMPLE 2. Let x_1, x_2 be a random sample from $N(\mu, \sigma^2)$ and $I(x_1, x_2)$ be the usual Student's interval for μ, at level of confidence one-half. Then

$$P\{\mu \in I(x_1, x_2) \mid |\bar{x}|/|x_1 - x_2| \leqq (1 + 2^{\frac{1}{2}})/2\} > \tfrac{2}{3}$$

for all values of μ, σ^2. There seems to be little reason to single out this conditioning event for interpreting outcomes which satisfy it, but again its existence makes the unconditional reference set seem inappropriate.

The difficulties illustrated in these examples are the subject of Theorems 2 and 3 below, in which it is shown that for any non-Bayesian confidence region procedure there is, in a randomized sense, a conditioning event yielding consequences similar to those above.

The desiderata for confidence regions given by Buehler (1959) relate to a generalization of the notion of conditional reference sets. The statistician is asked to accept bets regarding the success of his confidence regions, at odds appropriate to his level of confidence, with an adversary who gets to select the stakes as a function of the experimental outcome. The existence of conditioning events such as those in the above examples would provide the adversary with a strategy yielding positive expected winnings for all values of the parameter. Buehler's desiderata are that a confidence region procedure allow no such good strategy for the adversary. Theorems 1 and 3 below show that such good strategies for the adversary exist if and only if the statistician's procedure is not Bayes, with certain exceptions relating to weak Bayes strategies.

It is interesting to note that in this betting framework the statistician's objective is to "use all the sample information," in order that the adversary not be able to use such information against him. Likewise the principle of maximizing power is seemingly towards the same general objective. Example 1, shows clearly that the two principles do not coincide.

3. Betting strategies: finite parameter space. The elementary matrix theory of Lemma 1 below will suffice to prove Theorems 1 and 2 for the case that both the parameter space and the sample space are finite. The remaining results follow from generalizations of this basic lemma. Lemma 1 is well known (see, for example Mangasarian (1969) page 31); a proof is given because it is central to this paper.

Write E_k for Euclidean k-space, ϕ for the zero is any such space, and S_k for the simplex $\{\pi \mid \pi \in E_k, \pi = (\pi_1, \cdots, \pi_k), \sum_{i=1}^{k} \pi_i = 1, \pi_i \geqq 0, i = 1, \cdots, k\}$.

LEMMA 1. *Let T be any $m \times n$ matrix. If there is no $\pi \in S_n$ such that $T\pi = \phi$ then there exists an $s \in E_m$ such that $T's > \phi$, i.e. each component of $T's$ is positive.*

PROOF. Let $TS_n = \{y \in E_m \mid y = T\pi, \pi \in S_n\}$. The hypothesis is that $\phi \notin TS_n$. Since TS_n is closed and convex there is a hyperplane strictly separating it from ϕ, that is an $s \in E_m$ such that $s'y > 0$ for all $y \in TS_n$. Therefore $s'T\pi > 0$ for all $\pi \in S_n$ and the result follows by considering the points in S_n having only one nonzero co-ordinate. □

The result necessary for more general sample spaces is a simple generalization. Let (X, B, λ) be a σ-finite measure space. Write $L_1(X, B, \lambda)$ for the space of integrable functions on X, $L_\infty(X, B, \lambda)$ for the space of essentially bounded functions on X, and ϕ for the zero function on X.

LEMMA 2. *Let T be a linear operator on E_n defined by*

$$T\mu = \sum_{i=1}^{n} \xi_i(x)\mu_i, \qquad \mu = (\mu_1, \cdots, \mu_n) \in E_n,$$
$$\xi_i(x) \in L_1(X, B, \lambda), \qquad i = 1, \cdots, n.$$

If there is no $\pi \in S_n$ such that $T\pi = \phi$ a.e. (λ) *then there is an $s \in L_\infty(X, B, \lambda)$ such that*

$$\int s(x)\xi_i(x)\, d\lambda(x) > 0, \qquad i = 1, \cdots, n.$$

PROOF. This result can be established in essentially the same way as Lemma 1. TS_n, T-image of S_n, is a closed convex subset of $L_1(X, B, \lambda)$ and does not contain ϕ. (In the usual manner functions which agree a.e. (λ) are considered equivalent.) Hence there is a separating hyperplane, i.e. a continuous linear functional on $L_1(X, B, \lambda)$ which is strictly positive on TS_n. Since the dual of $L_1(X, B, \lambda)$ is $L_\infty(X, B, \lambda)$ there is thus an $s \in L_\infty(X, B, \lambda)$ such that this linear functional has representation $\int s(x)T\mu(x)\, d\lambda(x)$. The result follows as in Lemma 1 by considering the extreme points of TS_n. □

Let $\Omega = \{\theta_1, \cdots, \theta_n\}$ be any finite parameter space and let the σ-finite measure space (X, B, λ) be the sample space. For each $\theta \in \Omega$ let $f_\theta(x)$ on X be the sampling density with respect to λ. (The symbol x represents the entire sample.)

A confidence region procedure C is a measurable function $C(x, \theta)$ from $X \times \Omega$ to $\{0, 1\}$ with the interpretation

$$(3.1) \qquad C(x, \theta) = 1 \quad \text{if } \theta \text{ is in the confidence region based on } x, \\ = 0 \quad \text{otherwise.}$$

Thus the function

$$(3.2) \qquad \alpha_C(\theta) = \int C(x, \theta) f_\theta(x)\, d\lambda(x)$$

gives the probability that the confidence region includes θ when sampling from $f_\theta(x)$.

A procedure C is α-level Bayes for prior distribution $\pi(\theta)$; $\pi(\theta) \geqq 0$, $\theta \in \Omega$, $\sum_{\theta \in \Omega} \pi(\theta) = 1$; if

$$(3.3) \qquad \textstyle\sum_\theta C(x, \theta)\pi(\theta) f_\theta(x) / \sum_\theta \pi(\theta) f_\theta(x) = \alpha$$

for each $x \in X$ such that the denominator is positive. This relation is more conveniently expressed as

$$(3.4) \qquad \textstyle\sum_\theta [\alpha - C(x, \theta)]\pi(\theta) f_\theta(x) = 0 \qquad \text{for all } x \in X.$$

A procedure C is not α-level Bayes if there is no prior distribution on Ω for which (3.4) holds.

In this paper the interest is upon frequency behavior of confidence region procedures, which clearly will not be changed by redefining C on a set of λ-measure zero. For this reason say that C is essentially α-level Bayes if (3.3), or equivalently (3.4), holds for all x excepting a set of λ-measure zero. If the sample space is finite and λ is taken in the usual way to be counting measure, then there is no distinction between α-level Bayes and essentially α-level Bayes procedures.

The statistician is to select a procedure C in which his "level of confidence" is α, for some $\alpha \in (0, 1)$. It is not necessarily supposed that $\alpha_C(\theta) \equiv \alpha$ or even that $\alpha_C(\theta) \geqq \alpha$ on Ω; it is only supposed that the statistician will accept bets under the following scheme. Knowing C, an adversary selects a strategy $s(x)$, $-\infty < s(x) < \infty$, $x \in X$, determining the stakes for betting. If x is the experimental outcome and θ is the true parameter value, the statistician loses $s(x)\alpha$ if $C(x, \theta) = 0$ and wins $s(x)(1 - \alpha)$ if $C(x, \theta) = 1$. (In Buehler's (1959) scheme $s(\cdot)$ took values only in $\{-1, 0, 1\}$.)

The function

$$\psi^C_{s,\alpha}(\theta) = \int s(x)[\alpha - C(x, \theta)]f_\theta(x)\, d\lambda(x) \tag{3.5}$$

represents the statistician's expected loss as a function of θ. Consideration will be restricted to strategies $s \in L_\infty(X, B, \lambda)$, for which $\psi^C_{s,\alpha}(\theta)$ always exists. Ideally the statistician would like to choose C such that $\psi^C_{s,\alpha}(\theta) \leqq 0$ for all $\theta \in \Omega$ and all s. This is impossible apart from trivial situations since $\psi^C_{s,\alpha}(\theta) \equiv -\psi^C_{-s,\alpha}(\theta)$. The following theorem describes what can be done.

THEOREM 1. *Let Ω be finite, α be any given number in* $(0, 1)$, *and C be any confidence region procedure. There exists an $s \in L_\infty(X, B, \lambda)$ such that $\psi^C_{s,\alpha}(\theta) > 0$ for all $\theta \in \Omega$ if and only if C is not essentially α-level Bayes.*

PROOF. If C is essentially α-level Bayes for prior distribution π, then for each $s \in L_\infty(X, B, \lambda)$,

$$\begin{aligned}\textstyle\sum_\theta \pi(\theta)\psi^C_{s,\alpha}(\theta) &= \textstyle\sum_\theta \pi(\theta) \int s(x)[\alpha - C(x, \theta)]f_\theta(x)\, d\lambda(x) \\ &= \textstyle\int s(x) \sum_\theta [\alpha - C(x, \theta)]\pi(\theta)f_\theta(x)\, d\lambda(x) \\ &= 0\,,\end{aligned}$$

as a result of (3.4). Thus for no s can it be that $\psi^C_{s,\alpha}(\theta) > 0$ for all $\theta \in \Omega$.

Write

$$\xi(x, \theta) = [\alpha - C(x, \theta)]f_\theta(x)\,.$$

If C is not essentially α-level Bayes then there is no prior distribution π such that

$$\textstyle\sum_\theta \xi(x, \theta)\pi(\theta) = 0 \qquad \text{for almost all } x(\lambda)\,.$$

If $X = \{x_1, \cdots, x_m\}$ then Lemma 1, taking the $m \times n$ matrix

$$T = \{\xi(x_i, \theta_j)\}\,,$$

yields an $s = (s_i, \cdots, s_m)$ such that

$$\textstyle\sum_i s_i\xi(x_i, \theta) > 0 \qquad \text{for each } \theta \in \Omega\,.$$

Taking $s(x_i) = s_i$ yields the desired conclusion. If X is not finite it follows from $|\alpha - C(x, \theta)| \leqq 1$ that $\xi(x, \theta) \in L_1(X, B, \lambda)$ for each $\theta \in \Omega$. Lemma 2 yields an $s \in L_\infty(X, B, \lambda)$ such that

$$\textstyle\int s(x)\xi(x, \theta)\, d\lambda(x) > 0 \qquad \text{for all } \theta \in \Omega\,. \;\square$$

Note that if α is taken as a function of x, that is the statistician many vary the odds, then the same method of proof can be used to show that $\alpha(x)$ must agree with Bayes theorem for some prior distribution if one is to preclude strategies yielding positive expected loss for all θ.

The theorem makes no restriction on the function $\alpha_C(\theta)$, but rather asks that the statistician will accept bets at odds $\alpha : 1 - \alpha$. It might seem that the statistician would want to require that $\alpha_C(\theta) \equiv \alpha$ or $\alpha_C(\theta) \geqq \alpha$, and in considering this one should bear in mind that such restrictions are virtually incompatible with C being α-level Bayes. Still considering Ω to be finite, it is not impossible that a procedure satisfy both conditions but there are few problems admitting such procedures. An example is found by taking Ω to be n equally spaced points on a circle, X to be the same points, and the sampling distributions to be a location parameter family. Regions constructed in the usual way from the known distribution of $(x - \theta)$ will be Bayes for the uniform prior distribution. I know of no examples essentially different.

It appears then that in general the statistician can either: (a) Choose C such that $\alpha_C(\theta) \equiv \alpha$ and admit winning strategies for the adversary; or (b) choose a Bayes procedure so that both $\alpha_C(\theta) - \alpha$ and $\psi^C_{s,\alpha}(\theta)$, for every s, have π-average zero.

4. Conditional confidence levels: finite parameter space. The notion of a conditional reference set was introduced in Section 2. A generalization of this notion, introduced by Tukey (1958) and studied by Wallace (1959) and Stein (1961) is based on the concept of a selection. A selection is a λ-measurable function t on X, taking values in the closed unit interval, such that $E_\theta(t) > 0$ for all $\theta \in \Omega$. The interpretation is that $t(x)$ is the probability with which x is "retained" forming a subsequence of trials. A selection taking only values zero and one is called a pure selection.

The function of θ,

$$\alpha_{C,t}(\theta) = \int C(x, \theta) f_\theta(x) t(x)\, d\lambda(x) / \int f_\theta(x) t(x)\, d\lambda(x) , \tag{4.1}$$

represents the theoretical frequency that the confidence region contains θ in the subsequence obtained by the above process. If $\alpha_C(\theta) \equiv \alpha$ and t is a pure selection such that for some $\varepsilon > 0$ either $\alpha_{C,t}(\theta) \geqq \alpha + \varepsilon$ or $\alpha_{C,t}(\theta) \leqq \alpha - \varepsilon$ for all θ, then the set $t^{-1}(1)$ is what was called by Fisher a relevant subset of X. This phenomenon was illustrated in Examples 1 and 2. The introduction of the ε term is of course irrelevant when Ω is finite, but is included here for emphasis and to make the treatment here agree with that of the following section. In the more general setting the ε is important because it seems that a confidence region based in C for $x \in t^{-1}(1)$ should have confidence coefficient no greater than $\alpha - \varepsilon$ if $\alpha_{C,t}(\theta) \leqq \alpha - \varepsilon$ for all $\theta \in \Omega$.

The part of the following theorem relating to behavior of $\alpha_{C,t}(\theta)$ when C is Bayes was given by Wallace (1959), Theorem 1.

THEOREM 2. *Let Ω be finite. If C is essentially α-level Bayes then there is no selection t such that $\alpha_{C,t}(\theta) < \alpha$ for all $\theta \in \Omega$ or such that $\alpha_{C,t}(\theta) > \alpha$ for all $\theta \in \Omega$. If C is such that $\alpha_{C,t}(\theta) \equiv \alpha$ on Ω and C is not essentially α-level Bayes then there are selections t_1 and t_2 and an $\varepsilon > 0$ such that $\alpha_{C,t_1}(\theta) \leqq \alpha - \varepsilon$ and $\alpha_{C,t_2}(\theta) \geqq a + \varepsilon$ for all $\theta \in \Omega$.*

PROOF. If t is a selection then $E_\theta(t) > 0$ for all $\theta \in \Omega$ and

$$\alpha - \alpha_{C,t}(\theta) = [E_\theta(t)]^{-1} \psi^C_{t,\alpha}(\theta) .$$

The consequence of C being essentially Bayes follows as in the proof of Theorem 1.

If C is not essentially α-level Bayes then there is an essentially bounded, and hence a bounded, function s such that $\psi^C_{s,\alpha}(\theta) > 0$ for all $\theta \in \Omega$. That $\alpha_C(\theta) \equiv \alpha$ on Ω implies that

$$\int \xi(x, \theta)\, d\lambda(x) \equiv 0 \,,$$

where $\xi(x, \theta)$ is as defined in the proof of Theorem 1. Thus, taking $t_1(x) = as(x) + b$, where $a > 0$ and b are such that $0 \leqq t_1(x) \leqq 1$ on X,

$$\psi^C_{t_1,\alpha}(\theta) \equiv a\psi^C_{s,\alpha}(\theta) > 0$$

for all $\theta \in \Omega$. Since

$$E_\theta(t_1) \geqq \int t_1(x)\xi(x, \theta)\, d\lambda(x) > 0$$

for all $\theta \in \Omega$, it follows that $\alpha_{C,t_1}(\theta) < \alpha$, i.e. for some $\varepsilon > 0$, $\alpha_{C,t_1}(\theta) \leqq \alpha - \varepsilon$, for all $\theta \in \Omega$. A similar argument yields t_2. □

The selections given by this theorem are in general not pure and it would be of considerable interest to find conditions assuring the existence of pure selections with such properties.

5. General parameter space. Let the parameter space be any measurable space (Ω, A), where A is a σ-field containing all singletons. For any sequence $\{\pi_n\}$ of probability measures on (Ω, A) say that C is a α-level weak Bayes in mean for this sequence if

$$\text{(5.1)} \qquad \lim_{n\to\infty} \int |\int C(x, \theta)\, d\pi_n(\theta \mid x) - \alpha|\, h_{\pi_n}(x)\, d\lambda(x) = 0 \,,$$

where $\pi_n(\cdot \mid x)$ is the posterior probability measure resulting from prior measure π_n, and $h_\pi(x) = \int f_\theta(x)\, d\pi(\theta)$. Note that (5.1) is equivalent to

$$\text{(5.2)} \qquad \lim_{n\to\infty} \int |\int [\alpha - C(x, \theta)] f_\theta(x)\, d\pi_n(\theta)|\, d\lambda(x) = 0$$

Say that C is not α-level Bayes in mean if there is no sequence of probability measures on (Ω, A) for which (5.1) holds.

REMARK. This definition of weak Bayes in mean is apparently novel, but is strongly suggested by the natural approach to extending the preceding theorems. It appears to have some intuitive appeal although it seems somewhat difficult to apply. In particular I have not determined how it relates to a procedure being Bayes with respect to an improper prior distribution. On the other hand it seems worthy of consideration simply because it is precisely the condition which allows the method of proof used here to be extended to the case of a general parameter space. It is presented here with the hope that further study and application of it might be fruitful.

The consequence of C being Bayes in the following theorem was given by Wallace (1959). Consequences of C being weak Bayes (in some sense) are much more complex, and have been studied by Stein (1961), where weak Bayes was taken to mean Bayes with respect to an improper prior distribution. Wallace's (1959) Theorem 2 in this direction is incorrect, as shown by Stein (1961).

THEOREM 3. *If C is essentially α-level Bayes then for no $s \in L_\infty(X, B, \lambda)$, nor for any selection t, can it be that any of the relations $\psi^C_{s,\alpha}(\theta) < 0$, $\psi^C_{s,\alpha}(\theta) > 0$, $\alpha_{C,t}(\theta) < \alpha$, or $\alpha_{C,t}(\theta) > \alpha$ hold for all $\theta \in \Omega$. If C is not α-level weak Bayes in mean then there exists an $s \in L_\infty(X, B, \lambda)$ and an $\varepsilon > 0$ such that $\psi^C_{s,\alpha}(\theta) \geqq \varepsilon$ for all $\theta \in \Omega$. If, moreover, $\alpha_C(\theta) \equiv \alpha$ on Ω then there exist selections t_1 and t_2 and an $\varepsilon > 0$ such that $\alpha_{C,t_1}(\theta) \leqq \alpha - \varepsilon$ and $\alpha_{C,t_2}(\theta) \geqq \alpha + \varepsilon$ for all $\theta \in \Omega$.*

PROOF. If C is α-level Bayes the consequences follow as in the previous theorems, using Fubini's theorem and the fact that $s(x)\,|\alpha - C(x, \theta)|$ is essentially bounded.

Define the transformation T as the integral version of that in Theorem 1, i.e.

$$T\pi = \int [\alpha - C(x, \theta)] f_\theta(x)\, d\pi(x)\,.$$

Write S for the set of all probability measures on (Ω, A) and TS for the image of S. That $TS \subset L_1(X, B, \lambda)$ follows from Fubini's theorem. The hypothesis C is not weak α-level Bayes in mean implies that there is no sequence $\{T\pi_n\} \in TS$ converging to zero in the L_1 norm. Hence $\phi \notin \overline{TS}$, the L_1 closure of TS, and since TS is convex there exists (cf. Wilansky (1969) page 220) a continuous linear functional on $L_1(X, B, \lambda)$ strictly separating $\overline{TS}$ from ϕ. Thus there is an $s \in L_\infty(X, B, \lambda)$ and an $\varepsilon > 0$ such that

$$\int s(x) y(x)\, d\lambda(x) \geqq \varepsilon$$

for all $y \in \overline{TS}$. The proof follows as in the previous theorems. □

Example 2 above illustrates that procedures which are Bayes with respect to an improper prior distribution may admit the selection t_2 of this theorem, and it appears that the algebraic structure, rather than the normality, is the essential condition. Stein (1961) has shown that the Student's interval admits no selection such as t_1.

6. Some conclusions. Consider procedures satisfying the usual condition $\alpha_C(\theta) \geqq \alpha$ on Ω. It seems reasonable to conjecture that the only problems in which there are either α-level Bayes or α-level weak Bayes procedures satisfying this condition are those with the algebraic structure studied by Hora and Buehler [9] (and in these $\alpha_C(\theta) \equiv \alpha$). For practical purposes these are problems of location and/or scale, including multi-sample and multi-variate cases. Theorem 3 would then apply to all other problems and the above discussion suggests that similar results hold in these as well. Thus the statement in the Introduction that standard frequency requirements are virtually incompatible with the desired conditional properties.

A class of problems to which Theorem 3 seems particularly relevant from a practical viewpoint are those in which there is no optimum confidence region procedure obtainable from power considerations. It is common practice to then base confidence regions on the distribution of an insufficient statistic such as a maximum likelihood estimator. In such problems Bayes procedures, possibly based on relatively diffuse prior distributions, are often eschewed because of their failure to have exact frequency interpretation, even though they seem certainly preferable from a sufficiency standpoint and may have reasonable approximate frequency behavior over large regions of the parameter space. The choice should be understood to be between an ad hoc method with $\alpha_C(\theta) \geqq \alpha$ but poor conditional behavior and a Bayesian method with $\alpha_C(\theta)$ having π-expection α and more satisfactory conditional properties.

REFERENCES

[1] Basu, D. (1964). Recovery of ancillary information. *Sankhyā, Ser. A.* **26** 3–16.
[2] Brown, L. (1967). The conditional level of Student's *t* test. *Ann. Math. Statist.* **38** 1068–1071.
[3] Buehler, R. J. (1959). Some validity criteria for statistical inference. *Ann. Math. Statist.* **30** 845–863.
[4] Buehler, R. J. and Feddersen, A. P. (1963). Note on a conditional property of Student's *t*. *Ann. Math. Statist.* **34** 1098–1100.
[5] Cornfield, J. (1969). The Bayesian outlook and its application. *Biometrics* **25** 615–657.
[6] Cox, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.
[7] Freedman, D. A. and Purves, R. A. (1969). Bayes method for bookies. *Ann. Math. Statist.* **40** 1177–1186.
[8] Hacking, I. (1965). *Logic of Statistical Inference*. Cambridge Univ.
[9] Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37** 643–656.
[10] Mangasarian, O. L. (1969). *Nonlinear Programming*. McGraw-Hill, New York.
[11] Pratt, J. W. (1961). Review of *Testing Statistical Hypotheses* by E. L. Lehmann. *J. Amer. Statist. Assn.* **56** 163–167.
[12] Stein, C. (1961). Approximation of prior measures by probability measures. Institute of Mathematical Statistics, Wald Lectures. Unpublished.
[13] Tukey, J. W. (1958). Fiducial Inference. Institute of Mathematical Statistics, Wald Lectures. Unpublished.
[14] Wallace, D. L. (1959). Conditional confidence level properties. *Ann. Math. Statist.* **30** 864–876.
[15] Wilansky, A. (1969). *Functional Analysis*. Blaisdell, New York.

Department of Statistics
Oregon State University
Corvallis, Oregon 97331