The Annals of Statistics
1977, Vol. 5, No. 5, 881–891

A CONDITIONAL CONFIDENCE PRINCIPLE

BY JAMES V. BONDAR

Canadian Radio-Television and Telecommunications Commission

A conditional confidence property is examined in the context of invariant statistical models, for set estimators of equivariant functions of the parameter. Set estimators deduced from likelihood considerations are then identical to Bayes credible sets induced from a right invariant prior. It is shown that amenability of the group ensures that these intervals satisfy a betting interpretation of confidence sets due essentially to Buehler and Tukey. As a corollary, a level α Bayes set estimator is of level at-most-α as a Neyman–Pearson confidence interval if the group is amenable.

0. Introduction. Stone (1976) has given examples (see the counterexample of Section 2 of the present paper) of problems in which the Bayes credible interval of level α based on a right invariant prior turns out to be a confidence interval (in the Neyman–Pearson sense) of rather less than level α for all parameter values; in one particularly amusing example, the Bayes probability of the interval covering the parameter is 25% for all x, while the confidence coefficient is at least 75% for all θ. Such bad behavior (related to Stone's concept of "strong inconsistency" of decision rules) is shown to be impossible in a class of statistical problems as a corollary of work on conditional confidence levels which we now describe.

Following on a discussion of conditional confidence properties by Fisher, Buehler (1959) considered a statistician, Peter, who after an experiment is performed, calculates a level α set estimate for an unknown parameter θ; it may or may not be a confidence set. Paul, knowing the observation x, bets with Peter at α to $1 - \alpha$ odds that this set fails to cover the true value of θ if x lies in some given subset T of the sample space $\mathscr{X}$. Let the conditional expectation of Peter's winnings given the subset T be called $e_T(\theta)$; T is called a *relevant subset* for the set estimator if there exists $\varepsilon > 0$ such that either $e_T(\theta) < -\varepsilon$ for all θ, or $e_T(\theta) > +\varepsilon$ for all θ. Buehler (1959) and Tukey (unpublished Wald lectures 1959) independently proposed what may be loosely called a consistency principle (as opposed to an optimality criterion) for level α set estimates: use no set estimator for which there is a relevant subset.

Buehler's principle is of interest in connection with certain difficulties of non-Bayesian, nondecision-theoretic inference, which often appear in the literature in connection with conditioning on ancillary statistics. See Cox (1958), and especially Birnbaum (1969). In particular the principle is strong enough to eliminate the conditionality "paradox" of Cox (Buehler (1959), Example 3.2):

Received December 1974; revised February 1977.

AMS 1970 *subject classifications.* Primary 62A05; Secondary 62A15.

Key words and phrases. Conditional confidence, relevant subset, Bayes set estimator.

a fair coin is tossed; if heads occur, one observation is taken from an independent $N(\mu, 1)$ distribution; if tails, a $N(\mu, 100)$ observation is made. The criterion of shortest expected length leads us to discard the "obvious" 40% interval in favor of one of the form: θ lies in an interval of zero length if tails occur; θ lies in an interval $X \pm 1.28$ (an 80% interval) if heads. This procedure seems not to be "sensible," but unless Neyman–Pearson inferential criteria are supplemented by some other principle (such as that of Tukey and Buehler), it seems impossible for a Neyman–Pearsonian to avoid this paradox.

Regrettably, the relevant subset principle is *too* strong: the example of Buehler and Feddersen ((1963), also Brown (1967)) shows that this principle eliminates the usual interval for the normal mean based on Student's t, which we all know to be a Good Interval ("everyone is born with the normal and t distributions inside him," Miller (1966)), and this has in some circles been taken to mean that the relevant subset principle is hopelessly faulty (this conclusion is drawn in, e.g., Durbin (1969), page 647). Stronger forms of the principle, considered by Tukey and by Wallace (1959) are subject to the same difficulty.

There are two ways around this difficulty. The first is due to G. K. Robinson who points out that while it is true that one would think poorly of the estimator if $e_T(\theta)$ were less than $-\varepsilon$ for all θ (T is then called a *negatively biased subset*), nevertheless $e_T(\theta) > \varepsilon$ is *not* upsetting, since this shows that the estimator is "doing at least as well as its nominal level in that part of the sample space." Such behavior on the part of an estimator is commendable, and yet the relevant subset principle rejects estimators which have this property! Consequently, Robinson proposes the consistency principle: use no set estimator for which there are negatively biased subsets. Robinson (1976) proves that this principle does not reject Student's interval for the normal mean, thus rediscovering another of Stein's unpublished results (Stein (1961)), and also shows that the Behrens–Fisher interval is not rejected. This (which we shall call *Robinson's principle* for convenience, even though it is not the only one of Robinson's contributions to this field) is thus not too strong and yet it is strong enough to eliminate Cox's paradox, at least in the form above.

The present paper considers a second approach to Buehler–Feddersen's objection to the relevant subset principle. It has been pointed out obliquely by Stone (1972, page 236) that a relevant subset T may have little probability for "most" values of θ, in which case x is most likely to fall in T^c which may not be a relevant subset at all, indeed e_{T^c} may be close to zero. This suggests that we replace conditional by unconditional expectation to arrive at the following consistency principle III: use no interval estimator for which there is a subset T of $\mathscr{X}$ and an $\varepsilon > 0$ such that Peter's expected gain is less than $-\varepsilon$ for all θ in Ω. This principle is weaker than both Robinson's and the relevant subset principles, but is strong enough to eliminate the Cox paradox, at least in the form given above. Moreover, if we set $T = \mathscr{X}$, our principle is just the condition that the strong inconsistency of Stone's example does not occur. The theorem of this paper shows that the class of Bayes credible sets based on right invariant priors will conform to principle III (in both the one-sided form given above and also the corresponding two-sided form) if the model is invariant under an amenable group. A forthcoming paper, Robinson (1975), defines and discusses a class of related but more complex principles; the chief difference between these and our principle III is that the amount bet by Paul need not be zero or one; a confidence interval rejected by our principle III has a *super-relevant set* in Robinson's nomenclature, and will also be rejected by the principles of Robinson (1975).

For one of the above principles to provide a satisfactory resolution of the difficulties of statistical inference mentioned earlier, it must at least give us reasonable answers to the question of whether or not to condition on an ancillary (and if so, which one). The principle will have to compete with ad hoc rules of the form "*always* condition on ancillaries" (more general rules of this sort are suggested in Barndorff-Nielsen (1973)), and must succeed in a wide class of statistical problems. Only time and further study can tell if any of the above approaches will lead us to our goal.

1. The theorem. In the following, G denotes a locally compact group of transformations with the identity element e, acting on a Euclidean space $\mathscr{X}$. Let the model be *invariant* under G, i.e., the data X is a random vector taking values in $\mathscr{X}$ and having the probability distribution $P(\cdot;\theta)$, $\theta \in \Omega$; if X has the distribution $P(\cdot;\theta)$ then gX has the distribution $P(\cdot; g\theta)$ for some $\theta' = g\theta \in \Omega$. We shall later assume that G is *transitive* on the parameter space Ω (i.e., for any $\theta_1, \theta_2 \in \Omega$ there is at least one $g \in G$ such that $g\theta_1 = \theta_2$), and that G is *free* on $\mathscr{X}$ (i.e., for any $x_1, x_2 \in \mathscr{X}$ there is at most one $g \in G$ such that $gx_1 = x_2$).

We shall say a second countable group is *amenable* if there is a sequence $\{P_n\}$ of probability measures defined on G such that for every measurable $B \subset G$ and $g \in G$, $|P_n(Bg) - P(B)| \to 0$ as $n \to \infty$. This is well known as the condition of the Hunt–Stein theorem, and is equivalent to the usual definition of amenability which is to be found in the literature of pure mathematics (see Bondar and Milnes (1976)). It is also known (see Bondar and Milnes (1976), Section 2, or Stone and von Randow (1968)) that the class of amenable groups includes every compact group, translation groups, the triangular matrix groups, the location-scale group given by $(t, s)x = sx + t$, and the group of the Behrens–Fisher problem. However, the group of nonsingular $m \times m$ matrices ($m \geqq 2$) is not amenable, which rules out some multivariate applications.

If $f(\cdot;\theta)$ is the density of $P(\cdot;\theta)$ w.r.t. some σ-finite measure, then the Bayes posterior for θ given x with the measure m as prior gives probability

$$K \int_\Theta f(x;\theta)\, dm(\theta) \tag{1}$$

to the (measurable) set $\Theta \in \Omega$; provided the norming constant K is nonzero ($1/K = \int_\Omega f(x;\theta)\, dm(\theta)$). We shall call (1) "the m-Bayes posterior distribution"—if $0 < K < \infty$. For any $x \in \mathscr{X}$, let $C(x;\cdot)$ be the (measurable) indicator function of a set estimate for θ given x. If C is an exact level α m-Bayes set estimator (i.e., the estimating set given x contains α of the posterior probability (1)), then (1) yields

$$\int_\Omega (C(x;\theta) - \alpha) f(x;\theta)\, dm(\theta) = 0\,. \tag{2}$$

In Buehler's betting interpretation of set estimators, as explained in the introduction, the (unconditional) expectation of Peter's gain is

$$\psi_T(\theta) = \int_T (C(x;\theta) - \alpha)\, dP(x;\theta)\,. \tag{3}$$

We shall show in our theorem that if m is a right-invariant prior for θ, if G is amenable and transitive on Ω, and if our model satisfies the regularity condition R (see below), then there is no set T in $\mathscr{X}$ for which $\psi_T(\theta)$ is bounded uniformly above zero for all θ (or below zero for all θ).

REGULARITY CONDITION R. G is a locally compact second countable complete metric topological group, and the multiplication law $(g, x) \to gx$ is measurable on $G \times \mathscr{X}$. Also, the action of G on $\mathscr{X}$ is a *free product* action, i.e., $\mathscr{X}$ is isomorphic (where *isomorphic* means existence of a measurable one-to-one map commuting with the action of G) to the cartesian product $G \times A$ of G with some second countable locally compact space A, such that $g(h, a) = (gh, a)$ for all g, h in G and a in A; also that each $P(\cdot;\theta)$ is dominated by $\mu \times \alpha$ where μ is a left invariant measure on G and α is a σ-finite measure on A such that, for any integrable F,

$$\int_{\mathscr{X}} F(h, a)\, d\mu(h)\, d\alpha(a) = \int_A \int_G F(gh, a)\, d\mu(g)\, d\alpha(a)\,. \tag{4}$$

Here and henceforth, the isomorphism between $\mathscr{X}$ and $G \times A$ is written $x \to (h, a)$; $d\alpha(a)$ will be written as da. The function $f(x;\theta) = f(h, a;\theta)$ shall denote the density at x of $P(\cdot;\theta)$ with respect to $\mu \times \alpha$.

Although this condition seems very specialized and restrictive when formally stated, most of the continuous models one encounters in parametric statistics satisfy regularity condition R if the group acts freely on $\mathscr{X}$. For example, the group actions of classical multivariate analysis are product actions (see Wijsman (1967) and (1972), Theorem 7.1); a more general and abstract result of this sort is Bondar (1976).

We now assume that G is transitive on Ω. It will be useful to reparametrize the distributions on $\mathscr{X}$ as follows: choose some fixed θ_0 in Ω; if $g\theta_0 = \theta$, we shall identify the group element g with the parameter value θ (at least one such g exists by transitivity of G). In particular, θ_0 will be associated with e in G. The same $P(\cdot;\theta)$ may be parametrized by more than one g; it also means that after reparametrization, our new Ω is G, and each θ is a group element, permitting us to speak, for example, of θ^{-1} and θx.

We shall write ν for the right-invariant measure on G corresponding to μ, and Δ for the modulus of G. See Nachbin (1976), Chapter 2. If we use ν as our prior m, use the reparametrized Ω, and write x as (h, a) then the posterior

(1) becomes

(1a) $$k(a)\Delta(h) \int_\Theta f(h, a; \theta)\, d\nu(\theta)$$

where $1/k(a) = \Delta(h) \int_G f(h, a; \theta)\, d\nu(\theta)$ does not depend on h. It is shown in, e.g., Dawid, Stone and Zidek (1973) that if G has a free product action on $\mathscr{X}$, and f is the density of x w.r.t. $\mu \times \alpha$, then

(5) $$f(h, a; \theta) = f(gh, a; g\theta) \qquad \text{almost all} \quad a\,, \quad h\,; \quad \text{all} \quad g\,.$$

From this it follows that $0 < k(a) < \infty$ for almost all x (since $1 = P(\mathscr{X}; \theta) = \int_A \int_G f(h, a; e)\, d\mu(h)\, da = \int\int f(e, a; h^{-1})\, d\mu(h)\, da = \int\int f(e, a; \theta)\, d\nu(\theta)\, da = \int_A k^{-1}(a)\, da$. Thus $k^{-1}(a) < \infty$ a.e.); hence the posterior distribution (1) is well defined. We note that (1a) is identical to D. A. S. Fraser's structural distribution for θ given x (Fraser (1968), page 64), hence our results also apply to structurally derived set estimators. When G is the translation or translation-scale group, these set estimators are associated with the name of E. J. G. Pitman.

For any subsets B and C of G, we define $BC = \{bc; b \in B, c \in C\}$ and $B[C] = \bigcap_{c \in C} Bc^{-1} = \{g; g \in G \text{ and } gC \subset B\}$. We say G has the property H_1 if there is an infinite sequence $\{G_n\}$ of closed subsets G_n of G with $\nu(G_n) < \infty$ such that $\nu(G_n[C])/\nu(G_n) \to 1$ as $n \to \infty$ for all compact $C \subset G$. It follows from these definitions that for all compact C and $D \neq \emptyset$ in G, $(G_n[C])[D] \supset G_n[DC]$, $\nu(G_n[C])/\nu(G_n[D]) \to 1$ and also $\{\nu(G_n) - \nu(G_n[C])\}/\nu(G_n[D]) \to 0$. It is known that H_1 is equivalent to amenability (Bondar and Milnes, Theorem 1; the proof is essentially in Emerson and Greenleaf (1967)).

We now have the machinery to state and prove our

THEOREM. *Let $(\mathscr{X}, G, \Omega)$ be an invariant statistical model under the group G, satisfying the regularity condition* R, *and for which G is transitive on Ω and is amenable; if a set estimator is exact α-level Bayes* w.r.t. *right Haar measure*, **then** *for any $\varepsilon > 0$ there is no set T in $\mathscr{X}$ such that $\psi_T(\theta) > \varepsilon$ for all $\theta \in \Omega$, and no T such that $\psi_T(\theta) < -\varepsilon$ for all θ.*

PROOF. First we note that there is an increasing sequence of compact $B_i \subset G$ and an increasing sequence of compact $K_i \subset A$ such that $\mathscr{X} = \bigcup_{n=1}^\infty B_i \times K_i$ (a proof is obtained by observing that $\mathscr{X}$ is a union of an increasing sequence of compact sets C_i, and that the projections $\pi_G : \mathscr{X} \to G$ and $\pi_A : \mathscr{X} \to A$ are continuous, hence $\pi_G(C_i) \times \pi_A(C_i)$ will suffice for our $B_i \times K_i$). Therefore for any $\delta > 0$ there are compact $B \subset G$ and $K \subset A$ such that $P(B \times K; \theta = e) > 1 - \delta$, and $e \in B$. Next, consider

$$\begin{aligned}\Delta(h) \int_K \int_{hB^{-1}} f(h, a; \theta)\, d\nu(\theta)\, da \\ &= \Delta(h) \int_K \int_{hB^{-1}} f(\theta^{-1}h, a; e)\, d\nu(\theta)\, da \quad \text{by expression (5)}\,.\end{aligned}$$

Set $y = \theta^{-1}h$, and since $d\nu(hy^{-1}) = \Delta(h)^{-1}\, d\mu(y)$, this becomes

$$\begin{aligned}&= \Delta(h) \int_K \int_B f(y, a; e)\, d\nu(hy^{-1})\, da \\ &= \int_K \int_B f(y, a; e)\, d\mu(y)\, da > 1 - \delta\,.\end{aligned}$$

Hence

$$\Delta(h) \int_K \int_{hB^{-1}} f(h, a; \theta)\, d\nu(\theta)\, da > 1 - \delta . \tag{6}$$

Integrating (2) w.r.t. da,

$$0 = \int_K \int_G (C(h, a; \theta) - \alpha) f(h, a; \theta)\, d\nu(\theta)\, da . \tag{7}$$

Now,

$$\begin{aligned} &|\textstyle\int_K \int_{hB^{-1}} (C(h, a; \theta) - \alpha) f(h, a; \theta)\, d\nu(\theta)\, da| \\ &\qquad = |\textstyle\int_K \int_G (C - \alpha) f\, d\nu(\theta)\, da - \int_K \int_{(hB^{-1})^c} (C - \alpha) f\, d\nu(\theta)\, da| . \end{aligned}$$

By (7), the first term of this expression is zero, so the expression is

$$\begin{aligned} &\leqq |\textstyle\int_K \int_{(hB^{-1})^c} (C - \alpha) f\, d\nu(\theta)\, da| \\ &\leqq \textstyle\int_K \int_{(hB^{-1})^c} f(\theta^{-1}h, a; e)\, d\nu(\theta)\, da . \end{aligned}$$

After the transformation $g = \theta^{-1}h$ and using the fact that $d\nu(hg^{-1}) = \Delta(h)^{-1}\, d\mu(g)$, this becomes

$$\begin{aligned} &= \Delta(h)^{-1} \textstyle\int_K \int_{((hB^{-1})^c)^{-1}h} f(g, a; e)\, d\mu(g)\, da \\ &= \Delta(h)^{-1} P(((hB^{-1})^c)^{-1}h \times K; e) \\ &= \Delta(h)^{-1} P(B^c \times K; e) \\ &\leqq \Delta(h)^{-1} P((B \times K)^c; e) \\ &\leqq \delta/\Delta(h) , \end{aligned} \tag{8}$$

the latter step by (6).

Using property H_1 we choose a sequence $\{G_n\}$ of closed subsets of G such that $\nu(G_n) < \infty$ and $\nu(G_n[D])/\nu(G_n) \to 1$ for all compact $D \subset G$.

Define $A_0 = \{(h, a; \theta) : \theta \in G, a \in K, h \in \theta B\}$,

$$\begin{aligned} C_n &= \{(h, a; \theta) : \theta \in G_n[B]\} , \\ A_{n1} &= \{(h, a; \theta) : (h, a; \theta) \in A_0, h \in G_n[B^{-1}B]\} , \\ A_{n2} &= A_0 \cap C_n - A_{n1} \quad \text{and} \quad A_{n3} = C_n - A_0 . \end{aligned}$$

Now consider

$$\begin{aligned} &\frac{1}{\nu(G_n[B])} |\textstyle\int_{G_n[B]} \phi_T(\theta)\, d\nu(\theta)| \\ &\qquad = \frac{1}{\nu(G_n[B])} |\textstyle\int_{C_n} (C(h, a; \theta) - \alpha) f(h, a; \theta) t(h, a)\, d\mu(h)\, da\, d\nu(\theta)| \end{aligned} \tag{9}$$

where t is the characteristic function of T. Now $A_{n1} \subset C_n$ (since $B^{-1}B \supset B$ implies $G_n[B^{-1}B] \subset G_n[B]$), hence $C_n = A_{n1} \cup A_{n2} \cup A_{n3}$. Thus expression (9) is

$$\leqq \frac{1}{\nu(G_n[B])} \{|\textstyle\int_{A_{n1}} (C - \alpha) f\, d\mu(h)\, da\, d\nu(\theta)| + \int_{A_{n2} \cup A_{n3}} f(h, a; \theta)\, d\mu(h)\, da\, d\nu(\theta)\} .$$

The integral over A_{n1} is of a bounded function over a set of finite measure, so Fubini's theorem applies; the second integral is of a nonnegative function, so

Fubini's theorem again holds. If λ is the product measure with element $d\lambda(h, a; \theta) = d\mu(h)\, da\, d\nu(\theta)$, the above bound for expression (9) becomes

$$\frac{1}{\nu(G_n[B])}\{|\textstyle\int_{A_{n1}}(C-\alpha)f\,d\lambda| + \int_{A_{n2}\cup A_{n3}} f\,d\lambda\}. \tag{10}$$

By the remarks after the definition of property H_1, $G_n[B^{-1}B] \subset (G_n[B])[B^{-1}]$, so that the section through A_{n1} of points with a given h-value $h = h_0$ is $(h_0 B^{-1}) \times K$. Therefore

$$\frac{1}{\nu(G_n[B])}|\textstyle\int_{A_{n1}}(C-\alpha)f\,d\lambda|$$

equals (Fubini again):

$$\frac{1}{\nu(G_n[B])}|\textstyle\int_{G_n[B^{-1}B]}\int_K\int_{hB^{-1}}(C-\alpha)f\,d\nu(\theta)\,da\,d\mu(h)|$$

$$\leqq \frac{1}{\nu(G_n[B])}\textstyle\int_{G_n[B^{-1}B]}\Delta(h)|\int_K\int_{hB^{-1}}(C-\alpha)f\,d\nu(\theta)\,da|\,d\nu(h).$$

By (8), this is

$$\leqq \frac{1}{\nu(G_n[B])}\textstyle\int_{G_n[B^{-1}B]}\delta\,d\nu(h)$$

$$= \frac{\nu(G_n[B^{-1}B])}{\nu(G_n[B])}\delta \to \delta \qquad \text{as} \quad n\to\infty \tag{11}$$

(the last step by a remark after the definition of H_1).

Next,

$$\frac{1}{\nu(G_n[B])}\textstyle\int_{A_{n2}} f(h, a; \theta)\,d\lambda(h, a; \theta)$$

$$\leqq \frac{1}{\nu(G_n[B])}\textstyle\int_{G_n-G_n[B^{-1}B]}\int_A\int_G f(h, a; \theta)\,d\nu(\theta)\,da\,d\mu(h).$$

The inner double integral (over A and G) can be written

$$\textstyle\int_A\int_G f(\theta^{-1}h, a; e)\,d\nu(\theta)\,da\,;$$

after the transformation $g = \theta^{-1}h$, this double integral becomes $\Delta(h)^{-1}$, hence our integral over A_{n2} is

$$\leqq \frac{1}{\nu(G_n[B])}\textstyle\int_{G_n-G_n[B^{-1}B]}\Delta(h)^{-1}\,d\mu(h)$$

$$= \frac{\nu(G_n - G_n[B^{-1}B])}{\nu(G_n[B])} \to 0 \tag{12}$$

(the last step by another remark after the definition of H_1).

Lastly,

$$\frac{1}{\nu(G_n[B])}\textstyle\int_{A_{n3}} f(h, a; \theta)\,d\lambda(h, a; \theta) \tag{13}$$

$$\leqq \frac{1}{\nu(G_n[B])}\textstyle\int_{G_n[B]}\iint_{(\theta B\times K)^c} f(h, a; \theta)\,d\mu(h)\,da\,d\nu(\theta).$$

Now the inner double integral of this last expression is $1 - P(\theta B \times K; \theta) = 1 - P(B \times K; e)$ which, by definition of B and K, is $\leqq \delta$. Hence, expression (13) is

$$\leqq \frac{1}{\nu(G_n[B])} \int_{G_n[B]} \delta \, d\nu(\theta) = \delta .$$

Taking this last inequality together with (10), (11), and (12), we find that expression (9) is less than or equal to a quantity which converges to the limit 2δ as $n \to \infty$, and δ may be made arbitrarily small. But if the theorem were false, there would be an $\varepsilon > 0$ such that $\psi_T > \varepsilon$ for all θ, or $\psi_T < -\varepsilon$, all θ, hence (9) would be greater than ε for all n. This establishes the theorem by contradicting its negation. □

2. Consequences and examples. Given a set estimator with indicator function $C(x; \theta) = C(h, a; \theta)$, we define $\beta(\theta)$ as the probability (in the usual Neyman–Pearson sense) that the set covers θ when the latter is the true value of the parameter. Define $\gamma(x)$ as the probability under the Bayes posterior derived from a right invariant prior, of all the θ values which are included in the confidence set when x is observed, that is to say $\gamma(x) = \Pr(\{C(x, \theta) = 1\} \mid x)$ where Pr is the Bayes posterior probability. Now we can set $T = \mathscr{X}$ in the theorem and get the

COROLLARY. *Under the assumptions of the theorem, if $C(\cdot; \cdot)$ is any set estimator for θ of ν-Bayes level exactly α (in the sense that $\gamma(x) = \alpha$ for all x), then*

$$\inf_{\theta \in \Omega} \beta(\theta) \leqq \alpha \leqq \sup_{\theta \in \Omega} \beta(\theta) .$$

In an important class of cases it has long been known that equality of α and β holds:

PROPOSITION (Stein (1965)). *Under the assumptions of the theorem, if $C(h, a; \theta) = C(gh, a; g\theta)$ for all h, g in G, a in A, then*:

(1) *C is the indicator function of a ν-Bayes set estimator with γ independent of x.*

(2) *This set estimator is also a Neyman–Pearson confidence set whose β is independent of θ.*

(3) $\beta = \gamma$.

The condition in the statement of the proposition may be restated: if C_x is the confidence set given x, then $gC_x = C_{gx}$ for all x and g. This equivariance of C_x occurs in many simple problems; in particular, Bayesians using right invariant priors, and Fraserian structural inference types will want to use a set containing that part of Ω which has the highest density for θ given x (the HPD estimator), and since this density is itself equivariant, the HPD estimator for θ will also have this property.

The proposition tells us that the level of the estimator when considered as a ν-Bayes credible set is equal to its level as a structural set estimator (in the sense of D. A. S. Fraser) which in turn equals the confidence level of the estimator considered as a Neyman–Pearson confidence set—a happy circumstance causing these three schools of inference to agree in their assessment of the set estimator (formally at least; they may differ in their interpretation of β and γ). In other situations (such as the Behrens–Fisher problem (Example 2 below)) the function of the parameter being estimated is not equivariant, but is torn apart by the group action; there is then no nontrivial equivariant C. It is in these situations that the Neyman–Pearson and ν-Bayes schools of thought may differ in their assessment of a set estimator. The above corollary may be viewed as a limitation on the degree of conflict between these assessments—if G is amenable.

EXAMPLE 1. $x_1, \cdots, x_n$ are i.i.d. $N(\mu, \sigma)$; the usual confidence interval $\bar{x} \pm t_{\alpha/2}s/n^{\frac{1}{2}}$ for μ satisfies the conditions of the proposition where G is the location-scale group. Thus by the proposition, this is both a ν-Bayes credible set and a Neyman–Pearson confidence interval of level α. By the theorem, it satisfies principle III. As mentioned in the introduction, it also satisfies Robinson's principle, but not the relevant subset principle.

EXAMPLE 2 (Behrens–Fisher problem). $x_{i1}, \cdots, x_{in_i}$ are independent $N(\mu_i, \sigma_i)$ $(i = 1, 2)$. G is the set of all g such that $gx_{ij} = s_i x_{ij} + t_i$ where $s_i > 0$ and $-\infty < t_i < \infty$. The parameter function $\mu_1 - \mu_2$ is not equivariant, and no nontrivial interval estimator based upon $\bar{x}_1 - \bar{x}_2$ can be equivariant. The Behrens–Fisher interval is well known to be a ν-Bayes interval and thus satisfies principle III by the theorem. Robinson (1976) showed it also satisfies his principle. The statement in the corollary may be proved directly in this case: $\beta(\theta)$ equals α when $\sigma_1 = 0$, $\sigma_2 = 1$, hence $\inf_\beta \beta(\theta) \leqq \alpha \leqq \sup_\theta \beta(\theta)$ directly without using the corollary. Extensive numerical calculation by G. K. Robinson (Savage (1976), page 470, footnote) establishes the stronger result that $\sup \beta(\theta) = \alpha$ to five decimal places for a wide range of n_i, σ_i, and α; i.e., the Behrens–Fisher interval is a Neyman–Pearson confidence interval. The Welch–Aspin interval for $\mu_1 - \mu_2$ was shown by Fisher to have a negatively biased relevant subset, namely $\{0 < a < s_1/s_2 < b < \infty\}$. However, using this set as our T does not violate principle III (since $|\psi_T(\theta)| < P(T; \theta)$; by making σ_1/σ_2 arbitrarily large, $P(T; \theta)$ can be made arbitrarily close to zero); if it could also be proved that no other set T did so, then we would have an example of an estimator rejected by Robinson's principle but accepted by principle III.

EXAMPLE 3. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be i.i.d. observations from a bivariate $N(\mathbf{0}, \sigma_1, \sigma_2, \rho)$ distribution. The problem is invariant under the group of nonsingular triangular matrices (Fraser (1964), page 846; in Fraser's language, G is the "progression group"). When a structural distribution for $(\sigma_1, \sigma_2, \rho)$ is found (or equivalently when a Bayes posterior is induced from right Haar measure on G) and the marginal distribution for ρ is calculated from this joint distribution, one gets precisely the fiducial distribution of Fisher for ρ given the $\mathbf{x}$'s (Fraser (1964), page 853). This means that fiducial (= structural = Bayes-with-right-invariant-prior) intervals for ρ satisfy principle III.

That some ergodic condition such as amenability is necessary for the theorem is shown by the following:

COUNTEREXAMPLE (Stone (1976), Example A). Let $G = \mathscr{X} = \Omega$ be the free group on two generators a and b, acting on itself by left multiplication. Then $\mu = \nu$ is counting measure ($\mu(g) \equiv 1$). Let $P(x; e)$ be $\frac{1}{4}$ if $x = a, b, a^{-1}$ or b^{-1}, and be zero otherwise. Let $d(x) = xa^{-1}$ if $x = a^n b^m \cdots b^q a^r$ with $r > 0$, and $d(x) = xa$ if x has this form with $r < 0$. Let $d(x) = xb^{-1}$ if $x = a^n b^m \cdots a^q b^r$ with $r > 0$ and $d(x) = xb$ if $r < 0$. Moreover, let $d(e) = a$. The interval $C(x; \theta)$ which equals 1 if $\theta = d(x)$, 0 otherwise, is a 25%-level Bayes credible interval (invariant prior), but $\psi_G(a) = \int_G (C - \frac{1}{4}) f(h; a)\, d\mu(h) = (\sum Cf) - \frac{1}{4} = P(d(x) = a; \theta = a) - \frac{1}{4} = \frac{3}{4}$. Similarly, $\psi_G(e) = \frac{3}{4}$, and $\psi_G(\theta) = \frac{1}{2}$ for $\theta \neq a$ or e. Therefore $\psi_G(\theta) \geqq \frac{1}{2}$ for all θ. Stone observes that as a Neyman–Pearson confidence interval, this has level $\geqq 75\%$, very different from its Bayes level. Of course, this G is not amenable.

Acknowledgments. To G. K. Robinson for much discussion; to University College, London for hospitality. Some knowledgeable and patient referees and associate editors have greatly improved the paper.

REFERENCES

BARNDORFF-NIELSEN, O. (1973). On *M*-ancillarity. *Biometrika* **60** 447–455.

BIRNBAUM, A. (1969). Concepts of statistical evidence. In *Essays in Honor of Ernest Nagel* (Sidney Morgenbesser, P. Suppes, and M. White, eds.). St. Martin's, New York.

BONDAR, J. (1976). On Borel cross-sections and maximal invariants. *Ann. Statist.* **4** 866–877.

BONDAR, J. and MILNES, P. (1976). Hunt-Stein and related conditions on groups. Submitted to *Ann. Statist.*

BROWN, L. (1967). The conditional level of Student's test. *Ann. Math. Statist.* **38** 1068–1071.

BUEHLER, R. J. (1959). Some validity criteria for statistical inferences. *Ann. Math. Statist.* **30** 845–863.

BUEHLER, R. J. and FEDDERSEN, A. (1963). Note on a conditional property of Student's *t*. *Ann. Math. Statist.* **34** 1098–1100.

COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.

DAWID, A. P. and STONE, M. (1972). Expectation consistency of inverse probability distributions. *Biometrika* **59** 486–489.

DAWID, A. P., STONE, M. and ZIDEK, J. (1973). Marginalisation paradoxes in bayesian and structural inference. *J. Roy. Statist. Soc. B* **35** 189–233.

DURBIN, J. (1969). Inferential aspects of the randomness of sample size in survey sampling. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, Jr., eds.) 629–651. Wiley, New York.

EMERSON, W. and GREENLEAF, F. (1967). Covering properties and Følner conditions for locally compact groups. *Math. Z.* **102** 370–384.

FRASER, D. A. S. (1964). On the definition of fiducial probability. *Bull. Inst. Internat. Statist.* **40** 842–856.

FRASER, D. A. S. (1968). *The Structure of Inference.* Wiley, New York.

GREENLEAF, F. (1969). *Invariant Means on Locally Compact Groups.* Van Nostrand, Princeton.

MILLER, R. G., JR. (1966). *Simultaneous Statistical Inference.* McGraw-Hill, New York.

NACHBIN, L. (1976). *The Haar Integral.* Reprinted by R. Krieger, Huntington, New York.

ROBINSON, G. K. (1975). Conditional properties of statistical procedures for location and scale parameters. Submitted to *Ann. Statist.*

ROBINSON, G. K. (1976). Properties of Student's t and of the Behren's—Fisher solution to the two means problem. *Ann. Statist.* **4** 963-971.

SAVAGE, L. J. (1976). On rereading R. A. Fisher. *Ann. Statist.* **4** 441-500.

STEIN, C. M. (1961). Wald lectures. Unpublished.

STEIN, C. M. (1965). The approximation of improper prior measures by prior probability measures. *Bernoulli-Bayes-Laplace Festschrift* 217-240. Springer, Berlin.

STONE, M. (1972). Review of the Waterloo Symposium on Statistical Inference. *Biometrika* **59** 236.

STONE, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.* **71** 114-116.

STONE, M. and VON RANDOW, R. (1968). Statistically inspired conditions on the group structure of invariant experiments and···. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 70-80.

WALLACE, D. (1959). Conditional confidence level properties. *Ann. Math. Statist.* **30** 864-876.

WIJSMAN, R. A. (1967). Cross-sections of orbits and their application to densities of maximal invariants. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 389-400, Univ. of California Press.

WIJSMAN, R. A. (1972). Examples of exponentially bounded stopping time of invariant tests···. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 109-127, Univ. of California Press.

RESEARCH BRANCH
CANADIAN RADIO-TELEVISION AND
TELECOMMUNICATIONS COMMISSION
100 METCALFE STREET
OTTAWA, ONTARIO K1A ON2
CANADA