CONDITIONAL CONFIDENCE LEVEL PROPERTIES[1]

By David L. Wallace

University of Chicago

1. Introduction. Some confidence region procedures have the property that, conditionally on the sample point lying in some subset of the sample space, the conditional confidence level (i.e. the conditional probability that the region covers the parameter) is less than the unconditional confidence level uniformly in the parameters. If confidence regions are interpreted as summarizing the knowledge of a parameter value obtained from an experiment, such behavior has been considered undesirable, particularly when the conditioning subset is in some sense irrelevant to the parameter of interest. ([2], [3], [4, Chap. IV], [9].) In many of these references, the issue is discussed in terms of an associated test of significance.) Buehler [1] has formalized this behavior and studied numerous examples. Tukey [9] has given a somewhat different formalization and obtained a number of results as part of a more complex framework for statistical inference.

In this paper, a class of conditional properties is defined that includes the Buehler and Tukey definitions. Sufficient conditions for a confidence procedure to possess various properties are obtained. The main result is that if a level α confidence procedure yields, for all samples, posterior probability α for some prior probability distribution on the parameter space, then there are no subsets of the sample space, with respect to which the conditional confidence is uniformly less (or greater) than α. A much more widely applicable, but slightly weaker, result is obtained if a sequence of prior distributions is used. The results apply to most of the classical confidence problems including discrete distribution problems and nuisance parameter problems as the Behrens-Fisher problem.

Confidence procedures for which no conditional confidence can be uniformly less (or greater) than and bounded away from the nominal level include the usual t, χ^2, F, Pitman conditional location and scale, and Behrens-Fisher procedures. The "uniformly less" conclusion applies to the one-sided binomial and Poisson procedures.

Definitions and terminology are given in Section two, results are stated in Section three and proved in Sections five and six, and examples are given in Section four.

2. Notation and definitions. Let Z be a sample space, Ω a parameter space, and $Y = Z \times \Omega$ their Cartesian product. For any set C in Y, let $C_{z\cdot} = \{\omega: (z, \omega) \varepsilon C\}$ and $C_{\cdot\omega} = \{z: (z, \omega) \varepsilon C\}$ denote the cross section sets. Let $(Z, \mathcal{A}, \mu)$ and $(\Omega, \mathcal{B}, \lambda)$ be measure spaces with σ-finite measures μ, λ. Let

Received December 29, 1958.

[1] Research carried out at the Statistical Research Center, University of Chicago, under the partial sponsorship of the Statistics Branch, Office of Naval Research. I wish to thank R. J. Buehler and F. J. Anscombe for helpful criticisms and suggestions.

p be a measurable function on $Z \times \Omega$, such that for each $\omega \varepsilon \Omega$, $p_\omega(\cdot) = p(\cdot \mid \omega)$ is a probability density on Z relative to μ. The function p is called by Tukey [9], a *specification*. The specification p will be fixed throughout this paper, though other related specifications will be used. Denote by P_ω and E_ω respectively the probability measure and expectation determined by p_ω on Z.

A density function ξ on Ω relative to the measure λ will be called a *prior density*. More generally, if ζ is any nonnegative function on Ω, not identically zero, ζ will be called a *prior quasi-density*. A prior quasi-density ζ will be said to be *admissible* with respect to the specification p, if

$$h_\zeta(z) \equiv \int_\Omega \zeta(\omega) p(z \mid \omega)\, d\lambda(\omega) < \infty$$

for all $z \varepsilon Z$ except for a set of μ measure zero. A prior quasi-density ζ will be said to be *admissible except on the set A with respect to the specification p* where A is any subset of Z, if $h_\zeta(z) < \infty$ for all $z \varepsilon Z - A$ except for a set of μ measure zero. Every prior quasi-density for which a constant multiple is a prior density, is admissible with respect to any specification.

For every prior quasi-density ζ and every $z \varepsilon Z$ for which $0 < h_\zeta(z) < \infty$, there is defined a density $g_\zeta(\cdot \mid z)$ on Ω relative to λ:

$$g_\zeta(\omega \mid z) = \frac{p(z \mid \omega)\zeta(\omega).}{h_\zeta(z)}$$

(That g_ζ is undefined for some z will not matter.) If ζ or some constant multiple of ζ is a prior density, $g_\zeta(\cdot \mid z)$ is the *posterior density* given by Bayes theorem. If not, $g_\zeta(\cdot \mid z)$ is still a probability density on Ω, but will be called here a *weak posterior density*. (Some useful simple properties of weak posterior densities are set forth in Section five.) If ξ is a prior density, then $h_\xi(\cdot)$ is a (marginal) density on Z relative to μ.

A *confidence procedure* is a measurable set C in the product space $Z \times \Omega$ with the interpretative rule that to each z, the confidence set $C_{z\cdot} = \{\omega\colon (z, \omega) \varepsilon C\}$ in Ω is assigned. Tukey calls C an *event*. No restrictions concerning confidence level will be placed in the definition of a confidence procedure.

A confidence procedure C is said to be *level α Bayes against ξ with respect to the specification p*—written C is $B(\alpha, \xi, p)$—if some constant multiple of ξ is a prior density on Ω and if, for each $z \varepsilon Z$ for which $g_\xi(\cdot \mid z)$ is defined, the set $C_{z\cdot}$ has probability α under the posterior density $g_\xi(\cdot \mid z)$.

A confidence procedure C is said to be *level α weak Bayes against ζ with respect to the specification p*—written C is $B^*(\alpha, \zeta, p)$—if ζ is an admissible prior quasi-density on Ω and no multiple of ζ is a prior density, and if, for each $z \varepsilon Z$ for which $g_\zeta(\cdot \mid z)$ is defined, the set $C_{z\cdot}$ has probability α under the weak posterior density $g_\zeta(\cdot \mid z)$.

A confidence procedure C is said to be *lower level α weak Bayes against ζ with respect to the specification* p—written C is $B^{**}(\alpha, \zeta, p)$—if ζ is a prior quasi-density on Ω admissible except for a set A (which may be empty), such that

$C_{z\cdot} = \Omega$ for all $z \varepsilon A$ and $C_{z\cdot}$ has a probability of at least α under the weak posterior density $g_\xi(\cdot \mid z)$ for all z for which g is defined.

Following Tukey [9], define a *selection* as a function k mapping Z into the unit interval such that $E_\omega(k) > 0$ for all $\omega \varepsilon \Omega$.[2] Let

$$p^{(k)}(z \mid \omega) = \frac{p(z \mid \omega)k(z)}{E_\omega(k)}.$$

$p^{(k)}$ is a specification and will be called the *selected (by k) specification.* Denote by $P_\omega^{(k)}, E_\omega^{(k)}, g_\xi^{(k)}$ the functions for the selected specifications corresponding to $P_\omega, E_\omega, g_\xi$.

Selection has the interpretation that in any conceptual infinite sequence of observation and parameter pairs, $\{(z_n, \omega_n); n = 1, 2, \cdots\}$, a new sequence is obtained as the subsequence in which the pair (z_n, ω_n) is retained according to the outcome of a chance process with retention probability $k(z_n)$. The process is assumed independent for each pair. If $k(z)$ takes only the values 0 and 1 (pure selection), the selection is according as z_n does or does not belong to the set $D = \{z: k(z) = 1\}$ and the selected specification consists of the family of densities $p_\omega(\cdot)$ truncated to the sample subspace D.

Define, now, a number of performance properties of a confidence procedure C.

1. C has property $c(\alpha)$ called *exact confidence* α if for all $\omega \varepsilon \Omega$, $P(C_{\cdot\omega}) = \alpha$.
2. C has property $\underline{c}(\alpha)$ called *lower confidence* α if

$$\inf_{\omega \varepsilon \Omega} P_\omega(C_{\cdot\omega}) = \alpha.$$

3. C has property $\bar{c}(\alpha)$ called *upper confidence* α if

$$\sup_{\omega \varepsilon \Omega} P_\omega(C_{\cdot\omega}) = \alpha.$$

4. C has *advance probability* α if it has exact confidence α, and if, for any selection k for which $P_\omega^{(k)}(C_{\cdot\omega}) = q$ for all $\omega \varepsilon \Omega$, $q = \alpha$.
5. C has *strong advance probability* α if it has advance probability α, and if, for any selections k_1, k_2 for which

$$P_\omega^{(k_1)}(C_{\cdot\omega}) \leqq P_\omega^{(k_2)}(C_{\cdot\omega})$$

for all $\omega \varepsilon \Omega$, equality holds for all $\omega \varepsilon \Omega$.

6. C has property $S_0(\alpha)$ if, for every selection k,

$$\alpha \leqq \sup_{\omega \varepsilon \Omega} P_\omega^{(k)}(C_{\cdot\omega}).$$

7. C has property $S_1(\alpha)$ if, for every selection k,

$$\inf_{\omega \varepsilon \Omega} P_\omega^{(k)}(C_{\cdot\omega}) \leqq \alpha \leqq \sup_{\omega \varepsilon \Omega} P_\omega^{(k)}(C_{\cdot\omega}).$$

---

[2] The restriction on positivity seems possibly too strong, except when the positive domain of $p(\cdot \mid \omega)$ is the same for all ω.

8. C has property $S_2(\alpha)$ if, for every selection k, there exist parameter values ω_1, ω_2, such that

$$P^{(k)}_{\omega_1}(C_{\cdot\omega_1}) \leqq \alpha \leqq P^{(k)}_{\omega_2}(C_{\cdot\omega_2}).$$

9. C has property $S_3(\alpha)$ if, for every selection k for which $P^{(k)}_{\omega}(C_{\cdot\omega}) \leqq \alpha$ for all ω or else $P^{(k)}_{\omega}(C_{\cdot\omega}) \geqq \alpha$ for all ω, equality holds for all ω.

10. C has property $S_4(\alpha)$ if it has property $S_3(\alpha)$ and if, for every pair of selections k_1, k_2 for which

$$P^{(k_1)}_{\omega}(C_{\cdot\omega}) \leqq P^{(k_2)}_{\omega}(C_{\cdot\omega})$$

for all ω, equality holds for all ω.

The properties have evident interrelations of which the most important are
strong advance probability $\alpha \Rightarrow$ advance probability $\alpha \Rightarrow c(\alpha)$.
$S_4(\alpha) \Rightarrow S_3(\alpha) \Rightarrow S_2(\alpha) \Rightarrow S_1(\alpha) \Rightarrow S_0(\alpha)$.
If C has property $c(\alpha)$, then
strong advance probability $\alpha \Leftrightarrow S_4(\alpha) \Rightarrow \cdots \Rightarrow S_1(\alpha) \Rightarrow$ advance probability α.

The ordinary term *confidence coefficient* α usually means exact confidence α, or sometimes lower confidence α. Tukey [9] introduced the sequence frequency (equivalent to exact confidence), advance probability, and strong advance probability to describe successively stronger properties of a confidence procedure in retaining "level α" under selections. The properties $S_1(\alpha)$ and $S_2(\alpha)$ for pure selections have been defined and studied by Buehler [1]. He names the selections violating the defining condition rather than the property. The principal reason for introducing the sequence $\{S_i(\alpha)\}$ is to permit differentiation of behavior of common confidence procedures. The unsymmetric property $S_0(\alpha)$ seems of interest in much the same "conservative" way that lower confidence $\underline{c}(\alpha)$ is of interest.

Buehler's examples, combined with the examples and results of this paper, seem to indicate the need for properties intermediate to advance probability and strong advance probability, and even suggest that strong advance probability may be so strong and rare as to be of little value.

3. Principal results.

THEOREM 1: *Let C be a confidence procedure which is level α Bayes against ξ with respect to the specification p for some ξ. Then C is $S_2(\alpha)$. If in addition, ξ is positive on Ω, C is $S_3(\alpha)$.*

COROLLARY 1: *A confidence procedure C which has lower (or upper) confidence α, but not exact confidence α, is never level α Bayes against any ξ positive on Ω.*

THEOREM 2: *Let C be a confidence procedure which is level α weak Bayes against ζ with respect to the specification p for some ζ. Then C is $S_1(\alpha)$.*

COROLLARY 2: *If C has exact confidence α and is level α weak Bayes against ζ, then C has advance probability α.*

COROLLARY 3: *If a fiducial distribution for ω for the sample point z has density*

$f(\cdot \mid z)$ which is a weak posterior density $g_\zeta(\cdot \mid z)$ with respect to some admissible prior quasi-density for all z, then any confidence procedure giving fiducial probability α for every z has the property $S_1(\alpha)$.

This specifically includes results of "integrating out" nuisance parameters. Such procedures will not, in general, have any of the confidence level properties: $c(\alpha)$, $\underline{c}(\alpha)$ or $\bar{c}(\alpha)$. The result is not at all dependent on the problems of construction or meaning of fiducial distributions and fiducial probability.

The results in examples (a), (b) of Section four could have been obtained using results of Fisher and of Jeffreys ([5], [6]) together with Corollary 3, but it seems preferable to derive directly the facts necessary to apply Theorem 2.

Confidence procedures for functions of ω with nuisance parameters are easily handled directly by Theorems 1 and 2 and Corollary 2. For a more explicit treatment in an important special case, suppose $\omega = (\theta, \phi)$ with $\Omega = \Theta \times \Phi$. A confidence procedure C will be called a *confidence procedure for θ* if C is a cylinder set with base C^* in $Z \times \Theta$. Assume that the measure λ on Ω is a product measure $\lambda_1 \times \lambda_2$ of measures on Θ and Φ.

COROLLARY 4: *If a confidence procedure C for θ with base C^* in Z has the property that*

$$C_{z.}^* = \{\theta: (z, \theta) \varepsilon C^*\}$$

has, for each z, probability α under the marginal distribution on Θ of a weak posterior density, then C is $S_1(\alpha)$.

THEOREM 3: *Let C be a confidence procedure which is lower level α weak Bayes against ζ with respect to the specification p for some ζ. Then C is $S_0(\alpha)$.*

Proofs are given in section six.

4. Examples. In examples (a), (b) and (c), Z is a Euclidean space with μ Lebesgue measure. In examples (d) and (e), Z is the nonnegative integers with counting measure. In all examples, Ω is a Euclidean space (or obvious subspace) with λ Lebesgue measure.

(a) *Normal.* Let Z be n-dimensional Euclidean space with coordinates independently and identically distributed as $N(\theta, \sigma^2)$. Let $\omega = (\theta, \sigma)$, $\bar{z} = \sum z_i/n$, $S = \sum (z_i - \bar{z})^2$.

(i) σ known. With admissible prior quasi-density $\zeta(\theta) \equiv 1$, the weak posterior density for θ is

$$g_\zeta(\theta \mid z) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n(\theta - \bar{z})^2}{2\sigma^2}}.$$

Hence any confidence procedure with confidence sets of the form

$$\{\theta: \theta - \bar{z} \varepsilon A_1\}$$

with A_1 a set on the real line with probability α under the distribution $N(0, \sigma^2/n)$ will have exact confidence α, be weak Bayes and $S_1(\alpha)$ and have advance probability α.

Not all such procedures are $S_2(\alpha)$. For let the set A_1 be any half-infinite interval, say $(-\infty, a)$. Let k be a pure selection retaining the point z if $\bar{z} < b$. The conditional confidence level

$$P_\theta\{z: \theta - \bar{z} < a \mid \bar{z} < b\} < \alpha$$

for all θ and b. The complementary selection gives conditional confidence greater than α for all θ. $S_1(\alpha)$ guarantees that the conditional confidence is not uniformly below (or above) and bounded away from α. I do not know if confidence procedures with A_1 a finite interval must have the property $S_2(\alpha)$.

(ii) σ unknown, $n \geqq 2$. With admissible prior quasi-density $\zeta(\theta, \sigma) = 1/\sigma$, the weak posterior density is

$$g_\zeta(\theta, \sigma \mid z) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n(\theta-\bar{z})^2}{2\sigma^2}} \cdot \frac{1}{\sigma^2\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{S}{2\sigma^2}\right)^{\frac{n-1}{2}} e^{-\frac{S}{2\sigma^2}}.$$

This can be best described as follows: (θ, σ) given z is distributed such that S/σ^2 is distributed as chi-square on $n - 1$ degrees of freedom and, conditional on σ, θ is $N(\bar{z}, \sigma^2/n)$. The marginal distribution of θ given z is such that $(\theta - \bar{z})\sqrt{n(n-1)/S}$ is distributed as Student's t on $n - 1$ degrees of freedom.

Any confidence procedure for θ with confidence sets of the form

$$\{\theta: (\theta - \bar{z})\sqrt{n(n-1)/S} \in A_2\}$$

with A_2 a set on the real line with probability α under the t_{n-1} distribution will have exact confidence α, and, using Corollary 4, will be weak Bayes level α with respect to ζ. The procedure is then $S_1(\alpha)$ and has advance probability α.

Buehler [1] has noted that *no* such procedure is $S_2(\alpha)$, a pure selection according as $S \geqq c$ giving conditional confidence uniformly less than α.

Any confidence procedure for σ with confidence sets of the form

$$\{\sigma: S/\sigma^2 \in A_3\}$$

with A_3 a set on the positive real line with probability α under the χ^2_{n-1} distribution, will have exact confidence α, be weak Bayes, and hence $S_1(\alpha)$ and have advance probability α.

These results are all special cases of example (c) on location and scale parameters.

(b) *The Behrens-Fisher problem.* Let Z be $n_1 + n_2$ dimensional Euclidean space, with coordinates independently distributed, the first n_1 identically as $N(\theta_1, \sigma_1^2)$, the last n_2 identically as $N(\theta_2, \sigma_2^2)$. Let $\omega = (\theta_1, \sigma_1, \theta_2, \sigma_2)$ and let $\bar{z}_1, \bar{z}_2, S_1, S_2$ be the means and sums of squares of deviations of the two sets of coordinates. Assume $n_1 \geqq 2$, $n_2 \geqq 2$.

With the admissible prior quasi-density $\zeta(\omega) = 1/\sigma_1\sigma_2$, the weak posterior distributions of (θ, σ_1) and (θ_2, σ_2) are independent and as obtained in example (a_{ii}) with appropriate $(n_i, \bar{z}_i, S_i)$.

The usual confidence procedures for σ_2/σ_1 with confidence sets of the form

$$\left\{ \sigma_2/\sigma_1 : \frac{S_1}{(n_1 - 1)\sigma_1^2} \cdot \frac{(n_2 - 1)\sigma_2^2}{S_2} \varepsilon A_1 \right\}$$

with A_1 having probability α under the F_{n_1-1, n_2-1} distribution will have exact confidence α, be weak Bayes and $S_1(\alpha)$ and have advance probability α.

The marginal weak posterior distribution for $\theta_1 - \theta_2$ is easily found to be such that

$$\frac{(\theta_1 - \theta_2) - (\bar{z}_1 - \bar{z}_2)}{\sqrt{a_1 + a_2}}$$

is distributed as the linear combination of independent Student's variates: $t_{n_1-1} \sin \theta - t_{n_2-1} \cos \theta$, where $a_i = S_i/[n_i(n_i - 1)]$ and $\theta = \tan^{-1}[a_1/a_2]^{\frac{1}{2}}$. This distribution can usefully be called the Behrens-Fisher distribution with parameters $n_1 - 1$, $n_2 - 1$ and θ — written BF$(n_1 - 1, n_2 - 1; \theta)$. (The usefulness of this terminology is illustrated in the paper [10] in which a more detailed related treatment of the Behrens-Fisher problem is given.)

Any confidence procedure for $\theta_1 - \theta_2$ with confidence sets of the form

$$\left\{ \theta_1 - \theta_2 : \frac{(\theta_1 - \theta_2) - (\bar{z}_1 - \bar{z}_2)}{\sqrt{a_1 + a_2}} \varepsilon A_2 \right\}$$

with A_2 a set having probability α under the distribution BF$(n_1 - 1, n_2 - 1, \theta)$ will be weak Bayes level α_1 and have the property $S_1(\alpha)$. Since the marginal weak posterior density for $\theta_1 - \theta_2$ is exactly the fiducial density for $\theta_1 - \theta_2$ under the Behrens-Fisher solution, and fiducial procedure with fiducial probability α has the property $S_1(\alpha)$. Such procedures are known not to have exact confidence α, but at least for n_1 and n_2 sufficiently large to have lower confidence α and not be $S_2(\alpha)$. The behavior for small n_1 and n_2 is unclear. (C.f. [10].)

The Welch asymptotic procedure with asymptotically exact confidence α does not possess property $S_1(\alpha)$. Fisher's criticism ([3]) of this procedure amounts effectively to showing that, for $n_1 = n_2 = 7$, a pure selection with retention if $|(S_1/S_2) - 1| < \delta$ for δ small gives conditional confidence uniformly below and bounded away from α. For n_1 and n_2 sufficiently large, asymptotic theory suffices to show that a selection with retention if

$$|[S_1 n_2(n_2 - 1)^2/S_2 n_1(n_1 - 1)^2] - 1| < \delta$$

for δ small has a similar effect. Calculations for small and moderate values of n_1 and n_2 indicate that the effect holds fairly generally.

(c) *Location and scale parameter families.* Let Z be n-dimensional Euclidean space, let $\omega = (\theta, \sigma)$ with θ a (real) location parameter, σ a (positive) scale parameter for the family p of distributions. Let $\epsilon = (1, \cdots, 1)$. Then

$$p_\omega(z) = \frac{1}{\sigma^n} q\left(\frac{z - \theta\epsilon}{\sigma}\right)$$

for a fixed density q.

(i) $\sigma = 1$, known. The prior quasi-density $\zeta_1(\theta) = 1$ is admissible ($\int q(z - \theta\epsilon)\, d\theta < \infty$ except on sets of measure zero) for all q, and the weak posterior density of θ is:

$$g_{\zeta_1}(\theta \mid z) = \frac{q(z - \theta\epsilon)}{\int_{-\infty}^{\infty} q(z - \psi\epsilon)\, d\psi}.$$

Let C be a confidence procedure which is level α weak Bayes with respect to ζ_1 and which also possesses the translation property: $(z, \theta) \varepsilon C$ if and only if, $(z + a\epsilon, \theta + a) \varepsilon C$ for all a, or equivalently, if and only if $z - \theta\epsilon \varepsilon C_{.0}$. The translation property guarantees that the confidence level is constant; for

$$\int_{C_{.\theta}} q(z - \theta\epsilon)\, d\mu(z) = \int_{C_{.0}} q(z)\, d\mu(z)$$

and since C is $S_1(\alpha)$, then it must have exact confidence α, and have advance probability α. Such a procedure is a conditional procedure with translation property as constructed by Pitman [7], by choosing a set with conditional probability α under the conditional distribution of $z - \theta\epsilon$ on each configural line determined by the differences $\{z_i - z_1,\, i = 2, \cdots, n\}$ of the sample point. Buehler [1] proved the $S_1(\alpha)$ result when $n = 1$.

(ii) $\theta = 0$, known. The prior quasi-density $\zeta_2(\sigma) = 1/\sigma$ is admissible for all q, and the weak posterior density of σ is

$$g_{\zeta_2}(\sigma \mid z) = \frac{\sigma^{-(n+1)} q(z/\sigma)}{\int_{-\infty}^{\infty} \tau^{-(n+1)} q(z/\tau)\, d\tau}.$$

Again, a confidence procedure which is level α weak Bayes with respect to ζ_2 and has the natural property under scale change, has exact confidence α, is $S_1(\alpha)$ and has advance probability α. It is a Pitman scale procedure, with conditional confidence α on each configural ray from the origin.

(iii) For $n \geqq 2$, the prior quasi-density $\zeta_3(\theta, \sigma) = 1/\sigma$ is admissible for all q and the weak posterior density of (θ, σ) is

$$g_{\zeta_3}(\theta, \sigma \mid z) = \frac{\sigma^{-(n+1)} q\left(\dfrac{z - \theta\epsilon}{\sigma}\right)}{\int_{-\infty}^{\infty} d\psi \int_0^{\infty} d\tau \left[\tau^{-(n+1)} q\left(\dfrac{z - \psi}{\tau}\right)\right]}.$$

Let C be a confidence procedure which is level α weak Bayes with respect to ζ_3 and which has the translation scale change property that $(z, (\theta, \sigma)) \varepsilon C$ if and only if $(z - \theta\epsilon)/\sigma \varepsilon C_{.(0,1)}$. Any such procedure will be $S_1(\alpha)$, have exact confidence α and advance probability α. Included are confidence procedures for θ and σ jointly and for θ or σ separately. For the latter, it suffices to determine each $C_{z.}$ according to the marginal posterior densities. Again, the procedures are just those of Pitman.

In all examples, the prior quasi-densities were those of the Haar measure on the appropriate group of translation and/or scale changes. If Ω is any σ-compact group of transformations on Z, and if $\zeta(\omega)\, d\lambda(\omega)$ is Haar measure on Ω, then a confidence procedure C can be obtained which is weak Bayes level α and which satisfies the standard invariance condition under the group Ω. The latter condition insures that the procedure has exact confidence not depending on ω and, from the former, the procedure is $S_1(\alpha)$, hence has exact confidence α and advance probability α. To prove that the Haar measure is admissible and to show that the weak posterior density is for each fixed ω the conditional density on each orbit in Z under the group Ω, seems to require slightly more structure.

(d) *Binomial.* Let p_ω be the binomial (n, ω) density. The usual one-sided confidence interval $(l_1(z), 1)$ for ω with lower confidence α is obtained for $z > 0$ as the root of the equation

$$\sum_{i=0}^{z-1} p(i \mid l_1) = \sigma$$

or, equivalently, of the incomplete beta equation

$$\frac{1}{B(z, n - z + 1)} \int_{l_1}^{1} \omega^{z-1}(1 - \omega)^{n-z}\, d\omega = \alpha.$$

For $z = 0$, $l_1(0) = 0$. The weak posterior density with respect to the prior quasi-density $\zeta_1(\omega) = \omega^{-1}$ (i.e., a beta quasi-distribution with parameters $(0, 1)$) is

$$g_{\zeta_1}(\omega \mid z) = \frac{\omega^{z-1}(1 - \omega)^{n-z}}{B(z, n - z + 1)}.$$

Since ζ_1 is admissible except for $z = 0$, for which $C_{0\cdot} = \Omega$, the confidence procedure is lower level α weak Bayes and is $S_0(\alpha)$. Selection by the set $\{z = 0\}$ shows that the procedure is not $S_1(\alpha)$.

Similarly, the usual confidence interval $(0, l_2(z))$ with lower confidence α is lower level α weak Bayes with respect to the prior quasi-density $\zeta_2(\omega) = (1 - \omega)^{-1}$, and is $S_0(\alpha)$ but not $S_1(\alpha)$.

The usual two-sided confidence interval $(l_1(z), l_2(z))$ combining the two one-sided lower confidence α procedures is a much more complex procedure having lower confidence depending on n and α, but lying between $2\alpha - 1$ and α. The nominal lower level $2\alpha - 1$ is achieved only for rare combinations of n and α, so that, *a fortiori*, the two-sided procedure is not usually $S_1(2\alpha - 1)$

(e) *Poisson*: Let p_ω be the Poisson density with mean ω. As with the binomial, the usual one-sided procedure for ω with intervals of the form $(l_1(z), \infty)$ and with lower confidence α is lower level α weak Bayes against the prior quasi-density $\zeta_1(\omega) = \omega^{-1}$, and the procedure is $S_0(\alpha)$ Since $l_1(0) = 0$, it is not $S_1(\alpha)$. However, the other one-sided procedure does not suffer from the end effect and the interval $(0, l_2(z))$ determined for all z from the equation

$$\sum_{z+1}^{\infty} p(i \mid l_2) = \alpha$$

or from the equation

$$\frac{1}{\Gamma(z+1)} \int_0^{l_2} \omega^z e^{-\omega}\, d\omega = \alpha$$

gives a confidence procedure with lower confidence α which is level α weak Bayes against the prior quasi-sensity $\zeta_2(\omega) = 1$ and hence has the property $S_1(\alpha)$.

5. A property of prior quasi-densities and weak posterior densities.

THEOREM 3: *If ζ is a prior quasi-density with $\int \zeta(\omega)\, d\lambda = \infty$, with corresponding weak posterior density $g_\zeta(\cdot \mid z)$, then there exists a sequence of prior densities $\{\xi_n\}$ with corresponding posterior densities $g_n(\cdot \mid z)$ such that for all $\omega \varepsilon \Omega$ and all z for which g_ζ is defined,*

$$\lim_{n\to\infty} g_n(\omega \mid z) = g_\zeta(\omega \mid z).$$

Further, if $\{\xi_n\}$ is any sequence of prior densities with corresponding prior densities $\{g_n(\cdot \mid z)\}$ such that there exist constants K and $\{a_n ; n = 1, 2, \cdots\}$ such that for all ω,

$$\lim_{n\to\infty} a_n \xi_n(\omega) = \zeta(\omega) \tag{5.1}$$

and

$$a_n \xi_n(\omega) \leqq K\zeta(\omega) \tag{5.2}$$

then

$$\lim_{n\to\infty} g_n(\omega \mid z) = g_\zeta(\omega \mid z).$$

In the second part of the theorem, condition (5.2) is necessary in that sequences $\{\xi_n\}$ can be found that satisfy condition (5.1) but for which $\{g_n(\omega \mid z)\}$ does not converge, or which converges but not to a probability density.

The second part will be proved first. For all z for which the right-hand denominator is finite and positive,

$$g_n(\omega \mid z) = \frac{a_n p(z \mid \omega) \xi_n(\omega)}{a_n \int_\Omega p(z \mid u) \xi_n(u)\, d\lambda(u)} .$$

Under conditions (5.1) and (5.2), both numerator and denominator converge respectively to the numerator and denominator of $g_\zeta(\omega \mid z)$ for all ω and for every z for which g_ζ is defined.

The first part of the theorem will be proved by exhibiting a sequence $\{\xi_n\}$ satisfying conditions (5.1) and (5.2). Since λ is σ-finite, there exists an increasing sequence of sets in Ω:$\{B_n ; n = 1, \cdots\}$ such that $\lim B_n = \Omega$ and $\lambda(B_n) < \infty$.

Define

$$\xi_n(\omega) = \begin{cases} \dfrac{\min [n, \zeta(\omega)]}{\displaystyle\int_{B_n} \min [n, \zeta(u)]\, d\lambda(u)} & \omega \varepsilon B_n \\ 0 & \omega \varepsilon B_n \end{cases}$$

ξ_n clearly satisfies the two conditions with $K = 1$,

$$a_n = \int_{B_n} \min [n, \zeta(u)]\, d\lambda(u).$$

6. Proofs of results of section three.

LEMMA 1: *If a confidence procedure C is $B(\alpha, \xi, p)$ then for every selection k, C is $B(\alpha, \xi_k, p^{(k)})$, with prior density*

$$\xi_k(\omega) = \frac{\xi(w) \cdot E_\omega(k)}{\int \xi(u) \cdot E_u(k)\, d\lambda(u)}.$$

LEMMA 2: *If a confidence procedure C is $B^*(\alpha, \zeta, p)$ (or $B^{**}(\alpha, \zeta, p)$), then for every selection k, C is $B^*(\alpha, \zeta_k, p^{(k)})$ (or $B^{**}(\alpha, \zeta_k, p^{(k)})$) with prior quasi-density*

$$\zeta_k(\omega) = \zeta(\omega) \cdot E_\omega(k).$$

By assumption in Lemma 1,

$$\frac{\displaystyle\int_{C_{z.}} \xi(\omega) p(z \mid \omega)\, d\lambda(\omega)}{\displaystyle\int \xi(u) p(z \mid u)\, d\lambda(u)} = \alpha.$$

But

$$\xi_k(\omega) p^k(z \mid \omega) = a(z) \xi(\omega) p(z \mid \omega)$$

so that

$$\frac{\displaystyle\int_{C_{z.}} \xi_k(\omega) p^{(k)}(z \mid \omega)\, d\lambda(\omega)}{\displaystyle\int_{\Omega} \xi_k(u) p^{(k)}(z \mid u)\, d\lambda(u)} = \alpha.$$

The proof of Lemma 2 is the same with the exceptional set for admissibility unchanged for the selected specfication.

Let χ_C denote the set characteristic function of the set C in $Z \times \Omega$. By the usual conditional expectation interchange of order of integration, for any prior density ξ and specification p,

$$
\begin{aligned}
\int_\Omega P_\omega(C, \omega)\xi(\omega)\, d\lambda(\omega) &= \int_{Z\times\Omega} \chi_C(z, \omega)p(z \mid \omega)\xi(\omega)\, d[(\mu \times \lambda)(z, \omega)] \\
&= \int_Z h_\xi(z)\left[\int_\Omega \chi_C(z, \omega)g_\xi(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z) \\
&= \int_Z h_\xi(z)\left[\int_{C_{z.}} g_\xi(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z).
\end{aligned}
\tag{6.1}
$$

Note that the lack of definition of g_ξ for those z for which $h_\xi(z) = 0$ is of no consequence.

If C is $B(\alpha, \xi, p)$, then for any selection k, C is $B(\alpha, \xi_k, p^{(k)})$ by Lemma 1, and applying equation (6.1) to ξ_k and specification $p^{(k)}$ yields

$$\int_\Omega P_\omega^{(k)}(C_{.\omega})\xi_k(\omega)\, d\mu(z) = \alpha.$$

It follows immediately that $P_w^{(k)}(C_{\cdot\omega}) < \alpha(>\alpha)$ for all ω is impossible so C is $S_2(\alpha)$. If ξ is positive on Ω, so is ξ_k, and C is $S_3(\alpha)$ and Theorem 1 is proved.

If C is $B^*(\alpha, \zeta, p)$, then for any selection k, C is $B^*(\alpha, \zeta_k, p^{(k)})$ by Lemma 2. Let $\{\xi_n\}$ be the sequence of prior densities guaranteed by Theorem 4 and let $\{g_n(\cdot \mid z)\}$ be the corresponding posterior densities under $p^{(k)}$, converging to $g_{\zeta_k}^{(k)}(\cdot \mid z)$. Since these are probability densities on Ω, it follows from Scheffé's theorem [8] that

$$\lim_{n\to\infty} \int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega) = \int_{C_{z.}} g_{\zeta_k}(\omega \mid z)\, d\lambda(\omega) \tag{6.2}$$

uniformly in z with the right hand side identically equal to α by hypothesis. Then with

$$h_n(z) = \int \xi_n(\omega)p^{(k)}(\omega \mid z)\, d\lambda(\omega),$$

$$\lim_{n\to\infty} \int_Z h_n(z)\left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z) = \alpha,$$

and this, together with equation (6.1) applied to ξ_n and $p^{(k)}$, yields

$$\lim_{n\to\infty} \int_\Omega P_\omega^{(k)}(C_{.\omega})\xi_n(\omega)\, d\lambda(\omega) = \alpha.$$

Hence, for no k is it possible that

$$\sup_\omega P_\omega^{(k)}(C_{\cdot\omega}) < \alpha$$

or

$$\inf_\omega P_\omega^{(k)}(C_{\cdot\omega}) > \alpha$$

so that C is $S_1(\alpha)$ and Theorem 2 is proved.

If C is $B^{**}(\alpha, \zeta, p)$ with ζ admissible except for the set $A \subset Z$ then, using the same notation as in the preceding proof, equation (6.2) holds uniformly in z for $z \not\in A$. The right hand side is now not less than α. Then

$$\int_Z h_n(z) \left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z) - \alpha$$
$$= (1 - \alpha) \int_A h_n(z)\, d\mu(z) + \int_{Z-A} h_n(z) \left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega) - \alpha\right] d\mu(z)$$

so that

$$\liminf_{n\to\infty} \int_Z h_n(z) \left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z) \geqq \alpha$$

and hence

$$\liminf_{n\to\infty} \int_\Omega P_\omega^{(k)}(C_{.\omega})\xi_n(\omega)\, d\lambda(\omega) \geqq \alpha.$$

Then for no k is it possible that

$$\sup_\omega P^{(k)}(C_{\cdot\omega}) < \alpha$$

and C is $S_0(\alpha)$ and Theorem 3 is proved.

Corollary 1 follows immediately from Theorem 1, and Corollaries 2 and 3 from Theorem 2. Corollary 4 follows from Theorem 2, by noting that $C_{z\cdot}$ is a cylinder set with base $C^*_{z\cdot}$ in $Z \times \Theta$.

REFERENCES

[1] R. J. BUEHLER, "Some validity criteria for statistical inferences," *Ann. Math. Stat.*, Vol. 30 (1959) pp. 845–863.
[2] D. R. COX, "Some problems connected with statistical inference," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 357–372.
[3] R. A. FISHER, "On a test of significance in Pearson's Biometrika Table (No. 11)," *J. Roy. Stat. Soc., Series B*, Vol. 18 (1956), pp. 56–60.
[4] R. A. FISHER, *Statistical Methods and Scientific Inference*, Edinburgh and London, Oliver and Boyd, 1956.
[5] H. JEFFREYS, *The Theory of Probability, Second Edition*, Oxford, Oxford University Press, 1948, pp. 120–122.
[6] H. JEFFREYS, "Note on the Behrens-Fisher formula," *Ann. Eugenics*, Vol. 10 (1940), pp. 48–51.
[7] E. J. G. PITMAN, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1938), pp. 391–421.
[8] H. SCHEFFÉ, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.
[9] J. W. TUKEY, "Fiducial inference," 1958 Wald Lectures, Cambridge, Massachusetts, meeting of the Institute of Mathematical Statistics, August 1958.
[10] D. L. WALLACE, "The Behrens-Fisher problem—a critical review and a subjective approach," in preparation.