SHORTER CONFIDENCE INTERVALS FOR THE MEAN OF A NORMAL DISTRIBUTION WITH KNOWN VARIANCE[1]

BY JOHN W. PRATT

Harvard Graduate School of Business Administration

1. Introduction and summary. This paper obtains and explores a family of confidence procedures for the mean of a normal distribution which are, in a certain sense, more efficient than the usual procedure.

Let X (possibly a sample mean) have a normal density $\varphi(x;\theta,\sigma^2)$ with unknown mean θ, and known variance σ^2. Let $R(X)$ be a confidence region for θ at level $1-\alpha$. Let $m(R)$ be the length of R if R is an interval; more generally for any region R let

$$m(R) = \int_R d\theta, \tag{1}$$

which we will also call the length of R. Then, following [3], we have

$$\begin{aligned} E_{\theta'}\{m(R(X))\} &= \iint_{\theta\varepsilon R(x)} d\theta\varphi(x;\theta',\sigma^2)\,dx = \int P_{\theta'}\{\theta\,\varepsilon\,R(X)\}\,d\theta \\ &= \int_{\theta\neq\theta'} P_{\theta'}\{\theta\,\varepsilon\,R(X)\}\,d\theta. \end{aligned} \tag{2}$$

Thus the expected length of the confidence region $R(X)$ may also be interpreted as the integral over all false values θ of the probability of covering θ, where the expected length and the probability are both computed under the true value θ'. Whether we are interested in length or in the probability of covering false values, we would like to make (2) small.

For a particular θ', we can minimize (2) as follows. Let $A(\theta)$ be the acceptance region of the family of tests corresponding to $R(X)$, that is

$$X\,\varepsilon\,A(\theta) \quad \text{if and only if} \quad \theta\,\varepsilon\,R(X). \tag{3}$$

Substituting (3) in (2) gives

$$E_{\theta'}\{m(R(X))\} = \int P_{\theta'}\{X\,\varepsilon\,A(\theta)\}\,d\theta = \int_{\theta\neq\theta'} P_{\theta'}\{X\,\varepsilon\,A(\theta)\}\,d\theta. \tag{4}$$

For $\theta\neq\theta'$, $1-P_{\theta'}\{X\,\varepsilon\,A(\theta)\}$ is the power of the test of the null hypothesis value θ against the alternative θ'. Thus we see that the expected length of the confidence region is minimized, when θ' is the true value, by choosing the test of each null hypothesis value θ which is most powerful against the alternative θ'.

Received July 5, 1962; revised November 27, 1962.

[1] This research was supported by the National Science Foundation. Reproduction in whole or in part is permitted for any purpose of the United States Government.

This gives the confidence interval

$$\min\{\theta', X - \xi_\alpha\sigma\} \leqq \theta \leqq \max\{\theta', X + \xi_\alpha\sigma\}, \tag{5}$$

where ξ_α is the upper α-point (not $\frac{1}{2}$ α-point) of the standard normal distribution.

Table 1 shows what happens if we guess $\theta = \theta'$ and use the foregoing confidence procedure: our expected length will be considerably less than that of the usual procedure if we guess correctly, but greater if we are wrong by much over 2σ (for $\alpha = .05$).

Since we do not know the true value of θ, we may prefer to minimize not the expected length under a particular θ', but a weighted average of this. Consider then, for some $W \geqq 0$, the weighted average

$$\int W(\theta')E_{\theta'}\{m(R(X))\}\,d\theta' = \iint P_{\theta'}\{X \in A(\theta)\}W(\theta')\,d\theta'\,d\theta, \tag{6}$$

where the equality follows from (4). If W is interpreted as the prior density of θ, then (6) is the prior (marginal) expected length of the confidence region $R(X)$, but this interpretation need not be made. The procedure minimizing (6) also corresponds to a most powerful test of each null hypothesis value θ against a certain alternative distribution, as we shall see explicitly in Section 2.

We are concerned with two kinds of question:

(A) If we use the minimizing confidence procedure for some W, what is the effect of W on the expected length and on the efficiency to be gained by giving up the usual procedure? In particular, how diffuse does W have to get before the gain over the usual procedure is small in terms of the weighted average of the expected length?

(B) Are the minimizing confidence intervals for W more like posterior probability intervals obtained from the prior W than are the usual intervals? Does the use of W in selecting a confidence procedure largely eliminate the difference between confidence intervals and posterior probability intervals?

Specifically, we will introduce a normal weight function $W(\theta') = \varphi(\theta'; \theta_0, \omega^2)$ and obtain the minimizing confidence procedure $R_\omega(X)$, which reduces to (5) with $\theta' = \theta_0$ when $\omega = 0$ and to the usual procedure when $\omega = \infty$ and is given by Figure 1 or Table 5 when $\omega = \sigma$ and $\alpha = .05$. Then to answer (A), Table 2 gives the expected length of the minimizing procedure for $\alpha = .05$ and various values of ω. Notice that, as ω increases, so does the value of $\theta - \theta_0$ at which the minimizing procedure has the same expected length as the usual procedure: for $\omega = 4\sigma$, $\theta - \theta_0$ can be about $1.5\omega = 6\sigma$ before the usual procedure is better; for $\omega = 2\sigma$, about $1.6\omega = 3.2\sigma$; for $\omega = 0$, about 2σ (from Table 1). Table 2 also gives the weighted average (6) of the expected length of the minimizing procedure. Thus for $\omega = 4\sigma$, the minimizing confidence procedure has weighted average expected length 1.5% less than the usual procedure, and in this sense saves 3% of the observations, so that the usual procedure is 97% efficient. For $\omega \leqq 2\sigma$, however, the usual procedure wastes more than 10% of the observations, in the same sense.

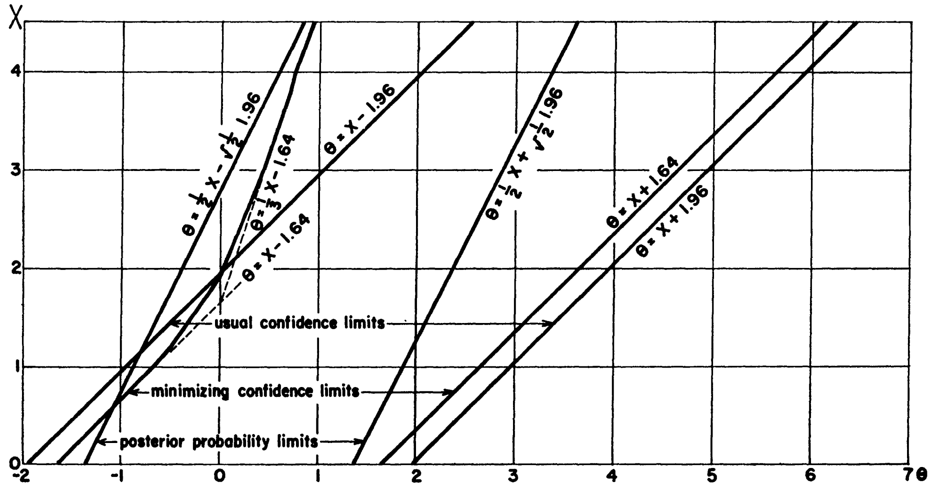

FIG. 1. Regions for θ at level $1 - \alpha = .95$.

To answer (B), Table 3 gives, for $\alpha = .05$ and $\omega = \sigma$, the posterior probability of the usual interval and the minimizing confidence interval when the prior density is normal with variance $\omega^2 = \sigma^2$. The X-scale reflects the fact that, with this prior, the prior (marginal) variance of X is $2\sigma^2$. Even for a priori probable X's, the confidence level .95 is a poor approximation to the posterior probability of either interval, though much poorer for the usual interval. Thus, by using a prior distribution, one can obtain a confidence procedure with prior expected length substantially less than that of the usual procedure (11% less in this instance, corresponding to a 23% smaller sample size), but considerable discrepancy remains between confidence and posterior probability.

In Section 2, we obtain the minimizing procedure and the formulas used in calculating the tables. Section 3 concerns conditioning on the event that the confidence region covers the true value of θ. Section 4 consists of remarks, which are largely independent of Sections 2 and 3.

I am grateful to the referee for a number of suggestions and comments.

2. The minimizing procedure. To minimize (6) for a normal weighting $W(\theta') = \varphi(\theta'; \theta_0, \omega^2)$, write the inner integral on the right-hand side in the form

$$\int P_{\theta'}\{X \varepsilon A(\theta)\} W(\theta')\, d\theta' = \iint_{x \varepsilon A(\theta)} \varphi(x; \theta', \sigma^2)\varphi(\theta'; \theta_0, \omega^2)\, dx\, d\theta' \tag{7}$$

$$= \int_{A(\theta)} \varphi(x; \theta_0, \sigma^2 + \omega^2)\, dx.$$

The second equality follows from the fact that if X is $N(\theta', \sigma^2)$ given θ' and θ' is $N(\theta_0, \omega^2)$, then X is marginally $N(\theta_0, \sigma^2 + \omega^2)$. Of course the calculation in no way depends on interpreting the weight function as a prior distribution.

The right-hand side of (7) is the probability that the test of the null hypothesis value θ would "accept" if X were distributed $N(\theta_0, \sigma^2 + \omega^2)$. Accordingly (7) and therefore (6) are minimized by choosing the most powerful test of each null hypothesis value θ against the alternative $N(\theta_0, \sigma^2 + \omega^2)$. By the Neyman-Pearson Lemma, this test "accepts" for

$$\exp\{-\tfrac{1}{2}(X - \theta)^2/\sigma^2\} > k(\theta, \theta_0, \sigma^2, \omega^2, \alpha) \exp\{-\tfrac{1}{2}(X - \theta_0)^2/(\sigma^2 + \omega^2)\}, \tag{8}$$

and hence for

$$|(X - \theta)/\sigma - (\sigma/\omega^2)(\theta - \theta_0)| < c_\alpha[(\sigma/\omega^2)(\theta - \theta_0)], \tag{9}$$

where $c_\alpha(z)$ is defined in terms of the standard normal cumulative Φ by

$$\Phi(z + c_\alpha(z)) - \Phi(z - c_\alpha(z)) = 1 - \alpha. \tag{10}$$

Clearly $c_\alpha(-z) = c_\alpha(z)$, $c_\alpha(0) = \xi_{\alpha/2}$, and $c_\alpha(z) = z + \xi_\alpha + o(1)$ as $z \to \infty$, where $o(1)$ denotes a term approaching 0. Furthermore, $z + c_\alpha(z)$ and $z - c_\alpha(z)$ are increasing functions of one another, whence each is an increasing function of z. A table of $z + c_\alpha(z)$ as a function of $z - c_\alpha(z)$ can be read directly from a table of the normal cumulative, and this yields a table of $c_\alpha(z) - z$ as a function of z. Table 4 was so constructed.

The minimizing confidence region for θ given by (9) is an interval $R_\omega(X)$ whose endpoints give equality in (9). For $\theta_0 = 0$, $\sigma^2 = 1$, this equation for the endpoints becomes

$$|X - \theta - \omega^{-2}\theta| = c_\alpha(\omega^{-2}\theta). \tag{11}$$

From the properties of $c_\alpha(z)$ it follows that the solutions $\underline{\theta}_\alpha(X, \omega)$ and $\bar{\theta}_\alpha(X, \omega)$ of (11) are increasing functions of X and:

$$\underline{\theta}_\alpha(x, \omega) = -\bar{\theta}_\alpha(-x, \omega), \qquad \underline{\theta}_\alpha(\xi_{\alpha/2}, \omega) = 0, \tag{12}$$

$$\bar{\theta}_\alpha(x, \omega) = x + \xi_\alpha + o(1) \quad \text{as} \quad x \to \infty, \tag{13}$$

$$\underline{\theta}_\alpha(x, \omega) = [\omega^2/(\omega^2 + 2)](x - \xi_\alpha) + o(1) \quad \text{as} \quad x \to \infty. \tag{14}$$

Table 4 and (11) yield x as a function of $\underline{\theta}_\alpha$ or $\bar{\theta}_\alpha$. From this, one can obtain $\underline{\theta}_\alpha$ and $\bar{\theta}_\alpha$ as a function of x. Alternatively one can use

$$\Phi(x - \underline{\theta}_\alpha(x, \omega)) - \Phi((1 + 2\omega^{-2})\underline{\theta}_\alpha(x, \omega) - x) = 1 - \alpha, \tag{15}$$

which follows from (10) and (11). The corresponding equation for $\bar{\theta}_\alpha$ is not needed because of (12). For arbitrary θ_0 and σ^2, the minimizing confidence interval $R_\omega(X)$ has endpoints $\theta_0 + \sigma\underline{\theta}_\alpha((X - \theta_0)/\sigma, \omega/\sigma)$ and

$$\theta_0 + \sigma\bar{\theta}_\alpha((X - \theta_0)/\sigma, \omega/\sigma).$$

Table 5 gives $\underline{\theta}_{.05}(x, 1)$ and $\bar{\theta}_{.05}(x, 1)$ and thus gives $R_\omega(X)$ for $\omega^2 = \sigma^2$, $\alpha = .05$. The posterior probabilities of Table 3 follow immediately from Table 5 and the fact that if X is $N(\theta', \sigma^2)$ given θ' and θ' is a priori $N(\theta_0, \omega^2)$ then θ' is a posteriori $N((\sigma^2 + \omega^2)^{-1}(\sigma^2\theta_0 + \omega^2 X), (\sigma^{-2} + \omega^{-2})^{-1})$ given X.

This section concludes with the evaluation required for Tables 1 and 2 of the expected length and weighted average expected length of $R_\omega(X)$. For convenience we take $\theta_0 = 0$ and $\sigma^2 = 1$ henceforth. Substituting the acceptance region (9) into (4) then gives

$$\begin{aligned} E_{\theta'}\{m(R_\omega(X))\} &= \int P_{\theta'}\{|X - \theta - \omega^{-2}\theta| < c_\alpha(\omega^{-2}\theta)\}\, d\theta \\ &= \omega^2 \int P_{\theta'}\{|X - \omega^2\lambda - \lambda| < c_\alpha(\lambda)\}\, d\lambda \end{aligned} \tag{16}$$

Letting $\omega^2 + 1 = a$, this may be written

$$\begin{aligned} E_{\theta'}\{m(R_\omega(X))\} &= \omega^2 \int_{-\infty}^{\infty} [\Phi(a\lambda - \theta' + c_\alpha(\lambda)) - \Phi(a\lambda - \theta' - c_\alpha(\lambda)]\, d\lambda \\ &= \omega^2 \int_0^{\infty} [G(a\lambda - \theta' - c_\alpha(\lambda)) - G(a\lambda - \theta' + c_\alpha(\lambda)) \\ &\quad - G(a\lambda + \theta' + c_\alpha(\lambda)) + G(a\lambda + \theta' - c(\lambda))]\, d\lambda, \end{aligned} \tag{17}$$

where $G = 1 - \Phi$ is introduced to make the integrals of the individual terms finite. Expression (17) was evaluated by using a piece-wise linear approximation to $c_\alpha(\lambda)$ and the relation

$$\int_u^{\infty} G(z)\, dz = \varphi(u) - uG(u) = L(u), \tag{18}$$

say. Raiffa and Schlaifer [4] have tabled the function $L(u)$, because it is the unit-normal linear-loss integral $\int_u^\infty (z - u)\varphi(z)\, dz$.

As $\theta \to \infty$ we have

$$E_\theta\{m(R_\omega(X))\} = 2\,\frac{\theta + (\omega^2 + 1)\xi_\alpha}{\omega^2 + 2} + o(1), \tag{19}$$

since, by (13) and (14), as $x \to \infty$,

$$m(R_\omega(x)) = \bar{\theta}_\alpha(x, \omega) - \underline{\theta}_\alpha(x, \omega) = 2\,\frac{x + (\omega^2 + 1)\xi_\alpha}{\omega^2 + 2} + o(1). \tag{20}$$

The weighted average of (17) was evaluated by using a piecewise quadratic approximation to (17).

For $\omega = 0$, the minimizing interval $R_0(X)$ is

$$\min(0, X - \xi_\alpha) \leqq \theta \leqq \max(0, X + \xi_\alpha);$$

then, by (4) or direct computation, we have the following formula, from which Table 1 was computed:

$$E_\theta\{m(R_0(X))\} = L(-\theta - \xi_\alpha) + L(\theta - \xi_\alpha). \tag{21}$$

3. Conditioning on coverage. It can be argued that one is really interested in short intervals when the true value is covered but not necessarily otherwise, and similarly in small probability of covering false values conditional on the true value being covered. The basic relation between expected length and probability of covering false values continues to hold when both are conditioned on coverage [3]. In the problem at hand, the conditional density of X given that $\theta' \varepsilon R(X)$ is just $(1 - \alpha)^{-1}\varphi(x; \theta', \sigma^2)$ for those x such that $\theta' \varepsilon R(x)$ and 0 otherwise, so (2) becomes

$$\begin{aligned} &E_{\theta'}\{m(R(X)) \mid \theta' \varepsilon R(X)\} \\ &= (1 - \alpha)^{-1} \int_{\theta' \varepsilon R(x)} \int_{\theta \varepsilon R(x)} d\theta \varphi(x; \theta', \sigma^2)\, dx \\ &= (1 - \alpha)^{-1} \int P_{\theta'}\{\theta \varepsilon R(X), \theta' \varepsilon R(X)\}\, d\theta \\ &= \int P_{\theta'}\{\theta \varepsilon R(X) \mid \theta' \varepsilon R(X)\}\, d\theta. \end{aligned} \tag{22}$$

Unfortunately the method of minimization used earlier no longer applies, but one can still use (22) to calculate the conditional expected length of $R_\omega(X)$ just as (2), in the form (4), was used at (16).

Conditioning on coverage, of course, makes no difference to the expected length of the usual procedure, which has constant length, and it turns out to make little difference to the expected length of $R_\omega(X)$ and virtually none to the weighted average or efficiency. Specifically, for $\theta_0 = 0$ and $\sigma^2 = 1$ (which we assume hereafter for convenience) and $\alpha = .05$, the expected length of $R_\omega(X)$ conditional on coverage is smaller than the unconditional expected length by about .02 for $\theta = 0$ and $\omega \leqq 2$, while it is larger by .11, .07, .04, .02, and .01 for large θ and $\omega = 0, 1, 2, 3$, and 4 respectively. The approach to the values obtaining for large θ is fairly rapid, but not rapid enough to make the weighted average larger conditionally than unconditionally. The formula which applies as $\theta \to \infty$, obtained from (13), (14), and (20), is

$$\begin{aligned} &E_\theta\{m(R_\omega(X)) \mid \theta \varepsilon R_\omega(X)\} \\ &= 2(\omega^2 + 2)^{-1}[\theta + (\omega^2 + 1)\xi_\alpha + (1 - \alpha)^{-1}\varphi(\xi_\alpha)] + o(1), \end{aligned} \tag{23}$$

where α is now arbitrary. This is to be compared with (19).

$E_\theta\{m(R_\omega(X)) \mid \theta \varepsilon R_\omega(X)\}$, which we will denote temporarily by $E(\theta, \omega)$, is discontinuous at $\theta = 0$, $\omega = 0$, and $E(0, 0)$ itself depends on the test of $\theta = 0$ corresponding to $R_0(X)$, which has not mattered previously. We have

$$\begin{aligned} &E(0, \omega) \to 2\xi_\alpha + (1 - \alpha)^{-1}[L(\xi_\alpha) + \varphi(\xi_\alpha) - 2\varphi(\xi_{\frac{1}{2}\alpha})] \quad \text{as} \quad \omega \to 0, \\ &E(\theta, 0) \to 2\xi_\alpha + (1 - \alpha)^{-1}L(\xi_\alpha) \quad \text{as} \quad \theta \to 0. \end{aligned} \tag{24}$$

The weighted average of $E(\theta, \omega)$ approaches the second limit as $\omega \to 0$. The first limit, which is slightly smaller, is the value of $E(0, 0)$ if the test of $\theta = 0$ corresponding to $R_\omega(X)$ is equal-tailed at level α for $\omega = 0$ as it is for all $\omega > 0$. The second limit is the value of $E(0, 0)$ if the test corresponding to $R_0(X)$ is one-tailed at level α for $\theta = 0$ as it is for $\theta \neq 0$. The test corresponding to (5) as given (with $\theta' = 0$) never rejects $\theta = 0$.

It should be emphasized that, while conditioning on coverage makes little difference to the expected length of $R_\omega(X)$, it has not been proved that $R_\omega(X)$ minimizes the weighted average expected length conditional on coverage. Indeed it does not in the case $\omega = 0$, $\alpha < \frac{2}{3}$, provided regions which are not intervals are permitted. This may be seen as follows. The version of $R_0(X)$ of minimum expected length conditional on covering 0, when $\theta = 0$, is the one corresponding to an equal-tailed level α test of $\theta = 0$ and hence covering 0 if and only if $|X| \leqq \xi_{\frac{1}{2}\alpha}$. Alter the acceptance regions for $\theta \neq 0$ by including all values of X with $|X| > \xi_{\frac{1}{2}\alpha}$ and removing some values of X with $|X| \leqq \xi_{\frac{1}{2}\alpha}$. For $|X| \leqq \xi_{\frac{1}{2}\alpha}$ this never adds points to $R_0(X)$ and sometimes removes points, so it must reduce the expected length conditional on covering 0. The weighted average must also be reduced, since all the weight is at 0. (If $\alpha \geqq \frac{2}{3}$, $R_0(X)$ consists of the point 0 alone for $|X| \leqq \xi_{\frac{1}{2}\alpha}$ and hence already has expected length 0 conditional on covering 0.)

4. Remarks.

4.1. *The effect of varying α.* For $\omega = 0$, the efficiency (as in Table 2) of the usual interval increases as α decreases, being, for instance .65 for $\alpha = .10$ and .82 for $\alpha = .99$ ([3], Table 1). This presumably gives some idea of the effect of varying α even for $\omega \neq 0$.

4.2. *The effect of changing the measure of size or reparameterizing.* If length is replaced by any other measure of size, the minimizing procedure for a given weighting function W remains the same [3]. Changing the measure of size, therefore, does not change the posterior probabilities of Table 3. It does, however, change the expected lengths of Tables 1 and 2; in particular, it changes the efficiency of the usual procedure with respect to the minimizing procedure. Of course, changing the weighting function W changes the minimizing procedure and, therefore, changes everything.

It follows that if θ is transformed monotonically to a new parameter μ, then the minimizing procedure for W is transformed into the minimizing procedure for the transformation (by the rule for densities) of W, regardless of the measure of size, and Table 3 transforms properly. The expected length comparisons of Tables 1 and 2, however, apply to that measure of size in μ which corresponds to length in θ.

4.3. *Direct comparison with posterior probability intervals.* If θ is a priori normal with mean θ_0 and variance ω^2, then it is a posteriori normal with mean

$$(\omega^2 X + \sigma^2 \theta_0)/(\omega^2 + \sigma^2)$$

and variance $\omega^2\sigma^2/(\omega^2 + \sigma^2)$. From this, various posterior probability intervals can be obtained, the shortest for a given probability having as midpoint the posterior mean.

One might attempt to answer question (B) by direct comparison of the usual confidence procedure and the minimizing procedure $R_\omega(X)$ with a posterior probability interval. For each of the confidence procedures, the posterior probabilities of the two tails are unequal, as well as varying. This means there is no natural posterior probability interval for comparison, so the interpretation of any such comparison is problematical. In addition, even when a posterior probability interval has been chosen, it seems to me harder to assess discrepancies in position than in posterior probability. Accordingly, it seems to me most natural and meaningful to compare the probabilities of coverage, as in the first two lines of Table 3. Nevertheless, the usual procedure, the minimizing procedure, and the shortest interval with posterior probability $1 - \alpha$ are given in Figure 1 for $\alpha = .05$ and $\omega^2 = \sigma^2$ and in Table A for arbitrary α and ω. In Table A, $R_\omega(X)$ is given only approximately. This approximation, which is derived in Section 2, becomes exact as $\omega \to 0$ but poor as $\omega \to \infty$. In both Figure 1 and Table A, we have taken $\theta_0 = 0$ and $\sigma^2 = 1$ for convenience.

4.4. *One-sided problems.* A weight function can also be used to select a one-sided confidence procedure, with length replaced by a measure which becomes zero when the confidence bound lies on the wrong side of the true value, as seems appropriate in one-sided problems [3]. This avoids the difficulties connected with unequal tails and the question of whether to condition on coverage (because the measure does so automatically, in effect). However, for the present problem, or any other with a monotone likelihood ratio, the minimizing one-sided procedure is just the usual one. This means that (a) the discrepancy between posterior probability and confidence is greater for the minimizing one-sided procedure than the minimizing two-sided procedure, but (b) one must look at the two-sided problem to see whether this discrepancy is reduced by using the minimizing procedure in preference to the usual procedure.

4.5. The usual procedure takes into account only the sample information, which is a proportion $\omega^2/(\omega^2 + \sigma^2)$ of all the information if W is prior information. From this point of view it is perhaps surprising the usual procedure is as efficient as it is. The explanation lies in the nature of the task set, which places complicated restrictions on what is permitted. In fact, it is not obvious one can make any use of prior information for this task: in the one-sided problem, for instance, one cannot.

4.6. If the weight function W represents one's prior judgment, then

$$\omega^{-1}|\theta - \theta_0| \geqq 2$$

is surprising and $\omega^{-1}|\theta - \theta_0| \geqq 3$ is astonishing, so that it would be astonishing if any of the large expected lengths in Table 2 occurred. A similar remark applies to Table 3.

4.7. In testing problems, the use of a weight function to select a procedure can essentially eliminate any difference between the actions or working conclusions

TABLE A

Comparison of Three Intervals

	Upper Limit	Lower Limit	Midpoint	Half-length
Shortest posterior probability interval	$\frac{\omega^2 X}{\omega^2+1} + \frac{\omega \xi_{\frac{1}{2}\alpha}}{(\omega^2+1)^{\frac{1}{2}}}$	$\frac{\omega^2 X}{\omega^2+1} - \frac{\omega \xi_{\frac{1}{2}\alpha}}{(\omega^2+1)^{\frac{1}{2}}}$	$\frac{\omega^2 X}{\omega^2+1}$	$\left(\frac{\omega^2}{\omega^2+1}\right)^{\frac{1}{2}} \xi_{\frac{1}{2}\alpha}$
Usual confidence interval	$X + \xi_{\frac{1}{2}\alpha}$	$X - \xi_{\frac{1}{2}\alpha}$	X	$\xi_{\frac{1}{2}\alpha}$
$R_\omega(X)$: $X \leqq -\xi_\alpha$	$\frac{\omega^2(X+\xi_\alpha)}{\omega^2+2}$	$X - \xi_\alpha$	$\frac{(\omega^2+1)X - \xi_\alpha}{\omega^2+2}$	$\frac{\lvert X\rvert + (\omega^2+1)\xi_\alpha}{\omega^2+2}$
$R_\omega(X)$: $\lvert X\rvert \leqq \xi_\alpha$	$X + \xi_\alpha$	$X - \xi_\alpha$	X	ξ_α
$R_\omega(X)$: $X \geqq \xi_\alpha$	$X + \xi_\alpha$	$\frac{\omega^2(X-\xi_\alpha)}{\omega^2+2}$	$\frac{(\omega^2+1)X + \xi_\alpha}{\omega^2+2}$	$\frac{\lvert X\rvert + (\omega^2+1)\xi_\alpha}{\omega^2+2}$

$\theta_0 = 0, \sigma^2 = 1$. Values for $R_\omega(X)$ are approximate.

of "orthodox" and "Bayesian" statisticians, though not the difference in their interpretations of data as evidence. In confidence theory, however, the use of a weight function does not eliminate the difference, because there the interpretation is paramount and posterior probability distributions are not ordinarily confidence distributions.

4.8. *Prior probability of coverage.* It is worth noting, however, that the prior expected value of the posterior probability of coverage is exactly $1 - \alpha$ for any confidence procedure $R(X)$ having exact level $1 - \alpha$ for all values of the parameter θ. In fact, regarding θ and X as jointly distributed random variables, we have, for any system of regions $R(X)$,

$$E(P\{\theta \varepsilon R(X) \mid X\}) = P\{\theta \varepsilon R(X)\} = E(P\{\theta \varepsilon R(X) \mid \theta\}), \tag{25}$$

that is, the prior expected value of the posterior probability of coverage equals the prior (marginal) probability of coverage, which equals the prior expected value of the confidence level. Thus if $R(X)$ has confidence level

$$P\{\theta \varepsilon R(X) \mid \theta\} = 1 - \alpha$$

for all θ, then each of the three quantities (25) is $1 - \alpha$, in particular, the prior expected value of the posterior probability of coverage is $1 - \alpha$.

By the same token, if $R(X)$ has posterior probability of coverage

$$P\{\theta \varepsilon R(X) \mid X\} = 1 - \alpha$$

for all X, then again each of the three quantities (25) is $1 - \alpha$, in particular' the prior expected value of the confidence level is $1 - \alpha$. All this is true in any problem, for arbitrary θ and X (not necessarily real), and remains true if

"$\theta \varepsilon R(X)$"

is replaced by

"$\tau(\theta) \varepsilon R(X)$"

throughout, i.e., if we are only interested in making statements about some function of θ. In this specific sense, then, a confidence region is an approximate posterior probability region and vice versa.

4.9. *Definitions of Bayesian shortness.* The confidence procedures discussed here and in [3] are "Bayes shortest" in the sense of minimizing the (prior) expected length of the confidence interval. This would be appropriate for a Bayesian faced with the task of producing a confidence interval, with a loss of utility proportional to the length of the interval realized. (From the Bayesian point of view, of course this task is very artificial, specifically the requirement that the probability of coverage given θ be at least $1 - \alpha$ for every θ.)

Starting from the concept of "shortest" in the sense of Neyman (= "most selective" or "most accurate"), Borges [1] calls a confidence procedure "subjectivtrennscharfe" if it minimizes

$$\int_{\theta' \neq \theta} P_{\theta'}\{\theta \varepsilon R(X)\} W(\theta')\, d\theta' \tag{26}$$

for every θ. This is equivalent to the concept under discussion here, as is indicated at (6) and shown generally in another paper [3], which also discusses more fully the relation between shortness in the Neyman sense and ordinary length.

TABLES

$\alpha = .05$. X is $N(\theta, \sigma^2)$, σ^2 known. Usual confidence interval has length 3.92σ. Among confidence intervals $R(X)$ at level $1 - \alpha$, $R_\omega(X)$ minimizes the weighted average of $E_\theta\{m(R(X))\}$, where $m(R)$ is the length of R and the weighting of θ is $N(\theta_0, \omega^2)$.

TABLE 1

$E_\theta\{m(R(X))\}$ for $R(X)$ minimizing $E_{\theta'}\{m(R(X))\}$

$\sigma^{-1}\|\theta - \theta'\|$	0	1	2	3	4	5	6	7	8
$\sigma^{-1}E_\theta\{m(R(X))\}$	3.33	3.45	3.89	4.68	5.65	6.65	7.64	8.64	9.64

TABLE 2

$\sigma^{-1}E_\theta\{m(R(X))\}$ and its weighted average

Eff. = efficiency of usual procedure with respect to $R_\omega(X)$ = (wtd. ave./3.92)^2

ω/σ \ $\omega^{-1}\|\theta - \theta_0\|$	0	1	2	3	4	5	6	7	8	Wtd. Ave.	Eff.
0	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	.72
1	3.35	3.44	3.75	4.24	4.87	5.53	6.20	6.86	7.53	3.45	.77
2	3.46	3.65	4.14	4.75	5.41	6.08	6.74	7.41	8.08	3.71	.89
3	3.62	3.76	4.15	4.64	5.18	5.72	6.27	6.81	7.36	3.78	.93
4	3.72	3.82	4.09	4.47	4.89	5.33	5.78	6.22	6.67	3.86	.97
∞	3.92	3.92	3.92	3.92	3.92	3.92	3.92	3.92	3.92	3.92	1.00

TABLE 3

Posterior probabilities for usual interval and $R_1(X)$

Prior $N(\theta_0, \sigma^2)$. X is marginally $N(\theta_0, 2\sigma^2)$

$2^{-\frac{1}{2}}\sigma^{-1}(X - \theta_0)$		0	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Inside interval	usual	.994	.988	.962	.898	.780	.607	.409	.233	.109	.042	.013
	$R_1(X)$	.980	.964	.928	.918	.930	.947	.962	.974	.982	.989	.993
Left tail	usual	.003	.012	.038	.102	.220	.393	.591	.767	.891	.958	.987
	$R_1(X)$	.010	.034	.072	.082	.070	.053	.038	.026	.018	.011	.007
Right tail	usual	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	$R_1(X)$	.010	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000

Godambe [2] defines a "Bayes shortest" confidence procedure as one such that, for each x, the interval has maximum posterior probability among intervals of the same length. He shows such a procedure is "admissible" in the sense that no other procedure both (i) has intervals as short for every x and (ii) has probability of coverage, given θ, as great for all θ and greater for some. (Condition (ii) would lead to greater marginal probability of coverage and hence to greater posterior probability of coverage given x for some x, which is impossible by (i) and Godambe's definition of Bayes shortest.)

In the present situation, with a normal prior, the confidence procedure minimizing the prior expected length of the interval is not Bayes shortest in Godambe's sense, since the interval is not symmetric around the posterior mean for every x (as may be seen from its behavior as $x \to \infty$. See also Table 5.) In fact, there is no confidence procedure having level exactly $1 - \alpha$ for all θ which is Bayes shortest in Godambe's sense, at least if $\alpha < .5$ and the prior variance ω^2 is sufficiently small. For suppose that $(\underline{\theta}(X), \bar{\theta}(X))$ is Bayes shortest in Godambe's sense and, for convenience, that $\theta_0 = 0$ and $\sigma^2 = 1$. Then

$$\tfrac{1}{2}[\bar{\theta}(X) + \underline{\theta}(X)] = \omega^2 X/(\omega^2 + 1),$$

the posterior mean. If the procedure has level at least $1 - \alpha$, then $\bar{\theta}(X) \geqq$

TABLE 4

$c_{.05}(z)$

$c_{.05}(z)$ satisfies $\Phi(z + c) - \Phi(z - c) = .95$, where Φ = unit normal cdf

z	0	.02	.04	.06	.08	.10	.12	.14	.16	.18	.20
$c_{.05}(z) - z$	1.960	1.940	1.921	1.903	1.886	1.870	1.854	1.839	1.825	1.811	1.798
z	.20	.25	.30	.35	.40	.45	.50	.60	.80	1.00	∞
$c_{.05}(z) - z$	1.798	1.770	1.745	1.724	1.707	1.693	1.682	1.665	1.650	1.646	1.645

TABLE 5

Endpoints of $R_1(X)$

Endpoints are $\theta_0 + \sigma\, \underline{\theta}_{.05}((x - \theta_0)/\sigma, 1)$ and $\theta_0 + \sigma\bar{\theta}_{.05}((x - \theta_0)/\sigma, 1)$. $\underline{\theta}_{.05}(x, 1)$ satisfies $c_{.05}(\underline{\theta}) = x - 2\underline{\theta}$ or equivalently $\Phi(x - \underline{\theta}) - \Phi(3\underline{\theta} - x) = .95$. $\bar{\theta}_{.05}(x, 1) = -\underline{\theta}_{.05}(-x, 1)$ satisfies $c_{.05}(\bar{\theta}) = 2\bar{\theta} - x$ or equivalently $\Phi(3\bar{\theta} - x) - \Phi(x - \bar{\theta}) = .95$.

x	$\leqq .4$	.4	.6	.8	1.0	1.2	1.4	1.6	1.8	2.0	
$\underline{\theta}_{.05}(x, 1)$	$x - 1.645$	−1.245	−1.046	−0.849	−0.660	−0.485	−0.332	−0.199	−0.083	+0.020	
x	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
$\underline{\theta}_{.05}(x, 1)$	.020	.114	.201	.282	.360	.434	.507	.577	.647	.715	.783
x	4.0	4.2	4.4	$\geqq 4.4$							
$\underline{\theta}_{.05}(x, 1)$	.783	.850	.918	$\frac{1}{2}(x - 1.645)$							

$X + \xi_\alpha$. Therefore

$$(27)\quad \underline{\theta}(X) = [2\omega^2 X/(\omega^2+1)] - \bar{\theta}(X) \leqq [(\omega^2-1)/(\omega^2+1)]X - \xi_\alpha .$$

For $\omega^2 \leqq 1$, it follows $\underline{\theta}(X) < 0$ for $X \geqq 0$; and $\underline{\theta}(X) < 0$ for $X < 0$ since $\underline{\theta}(X) \leqq X - \xi_\alpha$. Thus $\underline{\theta}(X) < 0$ for all X. Similarly $\bar{\theta}(X) > 0$ for all X. Hence for $\theta = 0$ the probability of coverage is 1. For $\omega^2 > 1$, (27) and a similar inequality for $\bar{\theta}(X)$ give

$$(28)\qquad P_\theta\{\bar{\theta}(X) > 0 > \underline{\theta}(X)\} \geqq P_\theta\{|X| \leqq [(\omega^2+1)/(\omega^2-1)]\xi_\alpha\}.$$

The left-hand side is the probability of coverage if $\theta = 0$, while the right-hand side exceeds $1 - \alpha$ if $(\omega^2+1)\xi_\alpha/(\omega^2-1) > \xi_{\frac{1}{2}\alpha}$. Thus $(\underline{\theta}(X), \bar{\theta}(X))$ cannot have level exactly $1 - \alpha$ at $\theta = 0$ if $\omega^2 < (\xi_{\frac{1}{2}\alpha} + \xi_\alpha)/(\xi_{\frac{1}{2}\alpha} - \xi_\alpha)$.

REFERENCES

[1] Borges, Rudolph (1962). Subjectivtrennscharfe Konfidenzbereiche. *Z. Wahrscheinlichkeitstheorie* **1** 47–69.

[2] Godambe, Vidyadhar P. (1961). Bayesian shortest confidence intervals and admissibility. *Bull. Inst. Internat. Statist.* 33e Session (Paris).

[3] Pratt, John W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 549–567.

[4] Raiffa, Howard and Schlaifer, Robert (1961). *Applied Statistical Decision Theory.* Graduate School of Business Administration, Harvard University, Boston.