The Annals of Statistics
1973, Vol. 1, No. 3, 545-550

ADMISSIBLE CONFIDENCE INTERVAL AND POINT ESTIMATION FOR TRANSLATION OR SCALE PARAMETERS

BY ARTHUR COHEN[1] AND WILLIAM E. STRAWDERMAN

Rutgers University

Sufficient conditions for admissibility of the best invariant confidence interval for a translation or scale parameter are given, for a very wide class of loss functions. These conditions result by adapting a theorem of L. D. Brown [2]. Simpler sufficient conditions are found for a subclass of loss functions of special interest. The subclass of losses involves three components. One concerned with coverage of the true value, another concerned with the distance from the interval end points to the true parameter, and a third concerned with length of the interval. Such a loss function unifies confidence interval and point estimation in the sense that if an optimality property holds for all loss functions in the subclass, then the optimality property holds for typical confidence interval problems and typical point estimation problems.

0. Introduction and summary. Sufficient conditions for admissibility of the best invariant confidence interval for a translation or scale parameter are given, for a very wide class of loss functions. These conditions result by adapting a theorem of L. D. Brown [2]. Simpler sufficient conditions are found for a subclass of loss functions of special interest. The subclass of losses involves three components; one concerned with coverage of the true value, another concerned with the distance from the interval end points to the true parameter, and a third concerned with length of the interval. Such a loss function unifies confidence interval and point estimation in the sense that if an optimality property holds for all loss functions in the subclass, then the optimality property holds for typical confidence interval problems and typical point estimation problems.

We start with preliminaries and the sufficient conditions for admissibility. In Section 2 we discuss the subclass of loss functions of special interest and give sufficient conditions for admissibility relevant to this subclass of losses. We conclude with examples.

1. Sufficient conditions for admissibility of the best invariant confidence interval or point estimate. Let X, Y be (jointly measurable) random variables with values in $Rx\mathscr{Y}(R = (-\infty, \infty))$ and

$$(1.1) \qquad P_\theta\{(X, Y) \in S\} = \iint_S p(x - \theta, y)\, dx\nu(dy)\,,$$

Received May 1972; revised October 1972.

[1] Research partially supported by N.S.F. Grant #-13709.

AMS 1970 *subject classifications*. 62C15, 62F10, 62G05.

Key words and phrases. Confidence intervals, admissibility, minimax, translation parameter, scale parameter.

where $\int p(x, y)\, dx = 1$ for almost all $y(\nu)$, θ is a real parameter in Ω, ν is a probability measure on the measurable sets in $\mathscr{Y}$. The problem is to obtain an interval estimator of θ and so an estimation procedure δ consists of two measurable functions $(\varepsilon_1(x, y), \varepsilon_2(x, y))$.

The loss functions considered may depend on three things: the difference between ε_1 and the true value of θ, ε_2 and the true value of θ and on y. The loss function will therefore be denoted by $W(t_1, t_2, y)$, where $t_i = \varepsilon_i - \theta$, $i = 1, 2$, and we will assume $0 \leqq W \leqq \infty$. The risk is

$$(1.2) \qquad R(\theta, \delta) = \int\int W(\varepsilon_1(x, y) - \theta, \varepsilon_2(x, y) - \theta, y) p(x - \theta, y)\, dx \nu(dy)\,.$$

We usually will write $\gamma_1(x, y) = \varepsilon_1(x, y) - x$, $\gamma_2(x, y) = \varepsilon_2(x, y) - x$ and so (1.2) often appears in the form

$$(1.3) \qquad R(\theta, \delta) = \int\int W(z + \gamma_1(z + \theta, y), z + \gamma_2(z + \theta, y), y) p(z, y)\, dz \nu(dy)\,.$$

An interval procedure is (translation) invariant if

$$(1.4) \qquad \varepsilon_1(x, y) = x + \gamma_1(y)\,, \qquad \varepsilon_2(x, y) = x + \gamma_2(y)\,.$$

Note from (1.3) that if δ is invariant $R(\theta, \delta) = R(\delta)$ (say) is independent of θ.

Let $\mathscr{I}$ be the class of invariant interval procedures. Unless otherwise stated it is assumed throughout that there exists at least one procedure $\delta_0 \in \mathscr{I}$ such that

$$(1.5) \qquad R_0 = R(\delta_0) = \inf_{\delta \in \mathscr{I}} R(\delta)\,.$$

The procedure δ_0 is called a best invariant interval estimator. Note that δ_0 is not necessarily uniquely determined. The symbols R_0 and $\delta_0(=(\varepsilon_{10}, \varepsilon_{20}))$, will sometimes be written $(\gamma_{10}, \gamma_{20})$. It is assumed that $R_0 < \infty$.

Now consider the following three assumptions:

$$(1.6) \qquad R(\delta_i) \to R_0 \qquad \delta_i \in \mathscr{I}, \; i = 1, 2, \cdots$$

implies $(\varepsilon_{1i}(x, y), \varepsilon_{2i}(x, y)) \to (x + \gamma_{10}(y), x + \gamma_{20}(y))$ (or $(\gamma_{1i}(y), \gamma_{2i}(y)) \to (\gamma_{10}(y), \gamma_{20}(y))$ in measure with respect to ν.

$$(1.7) \qquad J = \int_0^\infty d\lambda \{\sup_{\delta = (\gamma_1, \gamma_2) \in \mathscr{I}} \int \nu(dy) \int_{-\lambda}^{\lambda} [W(x + \gamma_{10}(y), x + \gamma_{20}(y), y) - W(x + \gamma_1(y), x + \gamma_2(y), y)] p(x, y)\, dx\} < \infty\,.$$

$$(1.8) \qquad W(x + t_1, x + t_2, y) \text{ is lower semi-continuous at } (t_1, t_2) = (\gamma_{10}(y), \gamma_{20}(y))\,, \quad \text{a.e. } (dx) X\nu\,.$$

We now state the

THEOREM. *Suppose assumption* (1.6), (1.7), *and* (1.8) *are satisfied and*

$$(1.9) \qquad \int \nu(dy) \int |x W(x + \gamma_{10}(y), x + \gamma_{20}(y), y)| p(x, y)\, dx < \infty\,.$$

Then the best invariant interval estimation procedure δ_0, is admissible.

PROOF. The proof requires an adaptation of the proof given by Brown [2]. We omit the details.

Note: After the author's proof of the theorem, Brown and Fox [3] developed a more general version of the above theorem.

Regarding condition (1.6), the reader is referred to conditions (a) and (b) of Brown and Fox [3], and Brown and Purves [4], for the analogue of Brown [2], Lemma 2.2.1. The analogue of Brown's Lemma 2.2.2 is

LEMMA 1. *Suppose there is a* $\delta \in \mathscr{I}$ *such that* $R(\delta) < \infty$. *Suppose* W *satisfies* (1.8) *and*

$$\lim_{r\to\infty} \inf_{(t_1,t_2)\notin U_r} W(t_1 t_2, y) \geqq R_0(y), \quad \text{a.e.} \quad \nu . \tag{1.10}$$

Then a best invariant procedure, δ_0, *say, exists. Assumption* (1.6) *is satisfied if and only if this interval is uniquely determined* a.e. ν.

Here $U_r = \{(v_1, v_2): v_1 \leqq v_2, ||v|| \leqq r^2\}$, and

$$R_0(y) = \inf_{(v_1,v_2)} R(v_1, v_2, y) \qquad \text{and}$$

$$R(v_1, v_2 | y) = \int W(z + v_1, z + v_2, y)p(z, y)\, dz .$$

REMARK. Minimaxity of δ_0 in Lemma 1 follows from Kiefer [8] page 263. Kiefer's proof requires closure of the class of decision procedures which follows from LeCam [11].

REMARK. For a counter example to the theorem, when (1.9) does not hold, see Perng [12].

2. Discussion of loss functions and some lemmas concerning assumption (1.7). We now choose to rewrite the loss function in a form appropriate for confidence interval estimation. That is, we write

$$W(t_1, t_2, y) = h(t_1, t_2, y) + (1 - I_{[t_1,t_2]}(0)), \qquad h \geqq 0 . \tag{2.1}$$

We would also like to distinguish a subclass of loss functions denoted as follows:

$$W(t_1, t_2, y) = c_0(1 - I_{[t_1,t_2]}(0)) + c_1 h_1(t_1, t_2, y) + c_2 h_2(t_1, t_2 y), \tag{2.2}$$

where $h_1 \geqq 0$, $h_2 \geqq 0$, and h_1 is an "estimation type" loss and h_2 is a "length type" loss. Consideration of (2.2) enables one to unify results in confidence estimation and point estimation. For example, suppose $h_1 = (t_1^2 + t_2^2)/2$ and $h_2 = (t_2 - t_1)$. Then if c_2 is "very large" relative to c_0 and c_1 we may be compelled to use a degenerate confidence interval i.e. $t_1 = t_2$. In the case of sampling from a continuous distribution then, the risk of such procedures reduces to a constant plus the mean squared error of a point estimate; which is equivalent to dealing with a point estimation problem with squared error loss. On the other hand if $c_1 = 0$, then the loss is a typical loss used in confidence interval estimation. Some specific losses will be treated in the next section on examples.

Now we offer some lemmas which give conditions that imply the validity of assumption (1.7) (i.e. $J < \infty$). The analogues of Brown [2], Lemma 2.3.1 and the first part of Lemma 2.3.2 are obvious. The following lemmas were designed to treat loss functions of the type (2.2).

Before we state the next lemma let us define the following sets:

$$Q_1' = \{(v_1, v_2): v_1 > 0, v_2 > 0, v_1 \leqq v_2\}$$
$$Q_2 = \{(v_1, v_2): v_1 \leqq 0, v_2 \geqq 0\}$$
$$Q_3' = \{(v_1, v_2): v_1 < 0, v_2 < 0, v_1 \leqq v_2\} .$$

We say that W satisfies condition (2.3), if there exists a bounded square B whose sides are of length 2β $(0 \leqq \beta < \infty)$ and whose center is at the origin, such that for each y,

(2.3) (i) $W(v_1, v_2, y)$ is non-decreasing along $(x + t_1, x + t_2)$ as x increases, for $(v_1, v_2) \in Q_1' \cap B^c$. ($B^c$ is the complement of B.)

(ii) $W(v_1, v_2, y)$ is non-increasing along $(x + t_1, x + t_2)$ as x increases, for $(v_1, v_2) \in Q_3' \cap B^c$.

(iii) $W(v_1, v_2 y)$ is non-decreasing in (v_1, v_2) for $(v_1, v_2) \in Q_2 \cap B^c$, as v_1 varies from 0 to $-\infty$ and v_2 varies from 0 to ∞.

(iv) $W(v_1, v_2, y)$ is non-decreasing in v_2 for fixed v_1 for $(v_1, v_2) \in Q_1' \cap B^c$.

(v) $W(v_1, v_2, y)$ is non-increasing in v_1 for fixed v_2 for $(v_1, v_2) \in Q_3' \cap B^c$.

Now we give

LEMMA 2. *Let W satisfy condition* (2.3) *and suppose*

$$\int \nu(dy) \int |x \sup_{|t_i|<2|x|+\beta, i=1,2} W(t_1, t_2, y)| p(x, y)\, dx < \infty . \tag{2.4}$$

Then (1.7) *is satisfied.*

PROOF. The proof is accomplished by essentially the same method used by Brown [2] in the proof of his Lemma 2.3.2. The analogue to Brown's $\hat{\gamma}'(y)$ is

$$(\hat{\gamma}_1'(y), \hat{\gamma}_2'(y)) \tag{2.5}$$
$$\begin{aligned}
&= (\beta + \lambda, \beta + \lambda) && \text{if } \gamma_1'(y) > \beta + \lambda \\
&= (\gamma_1'(y), \gamma_2'(y)) && \text{if } |\gamma_1'(y)| < \beta + \lambda , \quad |\gamma_2'(y)| < \beta + \gamma \\
&= (\gamma_1'(y), \beta + \lambda) && \text{if } |\gamma_1'(y)| < \beta + \lambda , \quad \gamma_2'(y) > \beta + \lambda \\
&= (-\beta - \lambda, \beta + \lambda) && \text{if } \gamma_1'(y) < -\beta - \lambda , \quad \gamma_2'(y) > \beta + \lambda \\
&= (-\beta - \lambda, \gamma_2'(y)) && \text{if } \gamma_1'(y) < -\beta - \lambda , \quad |\gamma_2'(y)| < \beta + \lambda \\
&= (-\beta - \lambda, -\beta - \lambda) && \text{if } \gamma_2'(y) < -\beta - \lambda .
\end{aligned}$$

REMARK. The conditions (2.3) are easier to interpret if we replace $h(t_1, t_2, y)$ in (2.1) by $h(t_2 - t_1)$ or replace $h_2(t_1, t_2 y)$ in (2.2) by $h_2(t_2 - t_1)$.

Next we give a lemma that will be used for a scale parameter problem. (See Example 2 of Section 3.) First we let $H = \{(v_1, v_2): v_2 < \beta\}$, $(0 \leqq \beta < \infty)$. We say that W satisfies condition (2.6) if, for each y,

(2.6) (i) $W(v_1, v_2, y)$ is non-decreasing along $(x + t_1, x + t_2)$ as x increases for $(v_1, v_2) \in Q_1' \cap H^c$.

(ii) $W(v_1, v_2, y)$ is non-decreasing in v_2 for fixed v_1 for $(v_1, v_2) \in (Q_1' \cup Q_2) \cap H^c$.

(iii) $W(v_1, v_2, y)$ is bounded in v_1 for fixed v_2 for $(v_1, v_2) \in (Q_2 \cup Q_3') \cap H^c$.

Now we give

LEMMA 3. *Let W satisfy condition* (2.6) *and suppose*

$$\int \nu(dy) \int |x \sup_{t_2<2|x|+\beta} W(t_1, t_2, y)| p(x, y)\, dx < \infty . \tag{2.7}$$

Then (1.7) *is satisfied.*

PROOF. Proceed as in proof of Lemma 2, only this time instead of (2.5) define the truncated versions of (γ_1, γ_2) as follows:

$$
\begin{aligned}
(2.8)\qquad &(\hat{\gamma}_1'(y), \hat{\gamma}_2'(y)) \\
&= (\gamma_1'(y), \gamma_2'(y)) \qquad \text{if } \gamma_2'(y) < \beta + \lambda \\
&= (\gamma_1'(y), \beta + \lambda) \qquad \text{if } \gamma_1'(y) < \beta + \lambda\,, \quad \gamma_2'(y) > \beta + \lambda \\
&= (\beta + \lambda, \beta + \lambda) \qquad \text{if } \gamma_1'(y) > \beta + \lambda\,.
\end{aligned}
$$

3. Examples.

EXAMPLE 1. Let X be a single observation (without loss of generality) from a normal distribution with mean θ and variance 1. Let the loss function be

$$(3.1)\qquad W(t_1, t_2, y) = (1 - I_{[t_1,t_2]}(0)) + c_1(t_1^2 + t_2^2)/2 + c_2(t_2 - t_1)\,.$$

Then the best invariant confidence interval is of the form $(x + \gamma_1, x + \gamma_2)$ where (γ_1, γ_2) are chosen to minimize

$$
\begin{aligned}
(3.2)\qquad \nu(\gamma_1, \gamma_2) &= E_0\, W(t_1, t_2, y) \\
&= 1 + (c_1/2)[2 + \gamma_1^2 + \gamma_2^2] + c_2(\gamma_2 - \gamma_1) - [\Phi(\gamma_2) - \Phi(\gamma_1)]\,.
\end{aligned}
$$

It is easily verified that the unique values minimizing (3.2) are as follows:
If $c_2 \geqq 1/(2\pi)^{\frac{1}{2}}$ then $\gamma_1 = \gamma_2 = 0$.
If $c_2 < 1/(2\pi)^{\frac{1}{2}}$ then $\gamma_1 = -\gamma_2$, and γ_2 is the unique positive solution of

$$(3.3)\qquad c_1\gamma_2 + c_2 - (e^{-\gamma_2^2/2}/(2\pi)^{\frac{1}{2}}) = 0\,.$$

Note if $c_2 \geqq 1/(2\pi)^{\frac{1}{2}}$ we have a point estimate, whereas if $c_1 = 0$ we have the confidence interval obtained by Joshi [7].
It is easily verified that the conditions of Lemma 2 are true, and also the other sufficient conditions of the theorem. Thus the best invariant procedure given above is admissible and minimax.

EXAMPLE 2. Let S be an observation from a population whose distribution is a scaled chi-square with n degrees of freedom. That is S/σ has a χ_n^2 distribution. We seek a confidence interval for σ when the loss function is

$$
\begin{aligned}
(3.4)\qquad &W((t_1 - \sigma)/\sigma, (t_2 - \sigma)/\sigma) \\
&= c_0(1 - I_{[t_1,t_2]}(\sigma)) + (c_1/\sigma^2)[(t_1 - \sigma)^2 + (t_2 - \sigma)^2] \\
&\quad + (c_2/\sigma)[t_2 - t_1]\,.
\end{aligned}
$$

We transform the above problem to the following translation parameter problem in the manner previously done by Farrell [6] and others. Observe $X = \log S$ so that the density of X is

$$(3.5)\qquad f(x - \theta) = \exp[(n/2)(x - \theta) - \tfrac{1}{2}e^{(x-\theta)}]\,,$$

where $\theta = \log \sigma$. Take the loss to be

$$(3.6)\qquad W'(t_1, t_2) = c_0(1 - I_{[t_1,t_2]}(0)) + c_1[(e^{t_1} - 1)^2 + (e^{t_2} - 1)^2] + c_2(e^{t_2} - e^{t_1})\,.$$

Once again it is easy to verify the conditions of Lemma 3 and the other sufficient conditions to conclude that the best invariant confidence interval is admissible and minimax. This example then gives the analogous result of Cohen [5] for confidence interval procedures. It is clear however that this result holds for more general loss functions and for many other scale parameter problems.

REMARK. Wolfowitz [14], Konijn [9], [10], Aitchison and Dunsmore [1], Winkler [13] evaluate confidence intervals for specific loss functions that combine length of the interval and, or some measure of the distance of the interval endpoints from the true parameter. The results of this paper are applicable to the loss functions they consider.

Acknowledgment. The authors are grateful to Professor L. D. Brown for corrections and very helpful comments.

REFERENCES

[1] AITCHISON, J. and DUNSMORE, I. R. (1968). Linear-loss interval estimation of location and scale parameters. *Biometrika* **55** 141-148.

[2] BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087-1136.

[3] BROWN, L. D. and FOX, M. (1973). Admissibility in statistical problems involving a location or scale parameter. *Ann. Statist.* To appear.

[4] BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. Submitted to *Ann. Statist.*

[5] COHEN, A. (1972). Improved confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **67** 382-387.

[6] FARRELL, R. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.* **35** 949-988.

[7] JOSHI, V. M. (1970). Admissibility of invariant confidence procedures for estimating a location parameter. *Ann. Math. Statist.* **41** 1568-1581.

[8] KIEFER, J. (1966). Multivariate optimality results, in *Multivariate Analysis*. Academic Press, New York.

[9] KONIJN, H. (1956). Some estimates which minimize the least upper bound of a probability together with the cost of observation. *Ann. Inst. Statist. Math.* **7** 143-158.

[10] KONIJN, H. (1964). Minimax interval estimates with a shortness criterion: a new formulation. *Ann. Inst. Statist. Math.* **15** 79-81.

[11] LECAM, L. (1955). An extension of Wald's theory of statistical decision function. *Ann. Math. Statist.* **26** 69-81.

[12] PERNG, S. K. (1970). Inadmissibility of various good statistical procedures which are translation invariant. *Ann. Math. Statist.* **41** 1311-1321.

[13] WINKLER, R. L. (1972). Decision-theoretic approach to interval estimation. *J. Amer. Statist. Assoc.* **67** 187-191.

[14] WOLFOWITZ, J. (1950). Minimax estimates of the mean of a normal distribution with known variance. *Ann. Math. Statist.* **21** 218-230.

STATISTICS CENTER
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903