

ADMISSIBILITY IN STATISTICAL PROBLEMS INVOLVING A LOCATION OR SCALE PARAMETER

BY LAWRENCE D. BROWN¹ AND MARTIN FOX²

Cornell University and Kibbutz Givat Oz; Michigan State University

The purpose of this short paper is to point out how a slight modification of the formulation of the problem in Brown (1966) yields a variety of admissibility results for problems involving an unknown location or scale parameter. The results sketched, but not proved, include some applications of the analogous two parameter results in Brown and Fox (1974). The applications involve testing problems, estimation problems, and confidence set problems. We treat here only fixed sample size situations; although there also exist similar sequential applications. The proof is essentially that contained in Brown (1966).

The purpose of this short paper is to point out how a slight modification of the formulation of the problem in Brown (1966) yields a variety of admissibility results for problems involving an unknown location or scale parameter. The results sketched below include some applications of the analogous two parameter results in Brown and Fox (1974). The applications involve testing problems, estimation problems, and confidence set problems. We treat here only fixed sample size situations; although there also exist similar sequential applications.

The formulation. The formulation is a one dimensional analog of the two dimensional formulation in Brown and Fox (1974) with some topological additions. We sketch it here; omitting some (measure theoretic) details which are contained in that paper.

X, Y are (jointly measurable) random variables with values in $\mathcal{R} \times \mathcal{Y}$ ($\mathcal{R} = (-\infty, \infty)$)

$$(1) \quad \Pr_{\theta} \{X, Y \in S\} = \int \int_S p(x - \theta, y) dx \nu(dy)$$

where $\int p(x, y) dx = 1$.

The decision space, \mathcal{D} , is a locally compact σ -compact topological space. It has defined on it a group of transformations $\{g_z : z \in \mathcal{R}\}$ which is a homeomorphic image of the translation group in \mathcal{R} . The (measurable) loss function, L , is assumed to be invariant, i.e.

$$(2) \quad L(d, \theta, x, y) = W(g_{-\theta}d, x - \theta, y).$$

The basic result. With the above formulation, and subject to certain other regularity conditions described below, the best invariant decision procedure is admissible. Let \mathcal{I} denote the class of nonrandomized invariant procedures.

Received May 1972; revised September 1973.

¹ This author's research was supported in part by NSF research grant number GP 24438.

² This author's research was supported in part by NSF research grant number GP 31123X.

AMS 1970 subject classifications. Primary 62C15; Secondary 62A05.

Key words and phrases. One dimensional location parameter, invariant, admissible.

Sufficient regularity conditions are:

(a) There exists some best invariant procedure. Call it δ_0 and let $R_0 = R(\theta, \delta_0)$. For all $\delta \in \mathcal{S}$, $R(\theta, \delta) = R_0$ implies $\delta(x, y) = \delta_0(x, y)$ a.e. $(dx \nu(dy))$. (In other words, δ_0 is essentially uniquely determined.)

(b) If $\delta_i \in \mathcal{S}$ and $R(\theta, \delta_i) \rightarrow R_0$, then $\delta_i(0, \cdot) \rightarrow \delta_0(0, \cdot)$ in measure (ν) . (Note: $\delta_i(x, y) = g_x \delta_i(0, y)$.) Conversely, if $\delta_i \in \mathcal{S}$ and $\delta_i(0, \cdot) \rightarrow \delta_0(0, \cdot)$ in measure (ν) then

$$\int \nu(dy) \int (W(\delta_0(x, y), x, y) - W(\delta_i(x, y), x, y))^+ p(x, y) dx \rightarrow 0 .$$

(It is usually the case that the topology on \mathcal{D} is locally compact and σ -compact and $L(\cdot, \theta, x, y)$ is a lower semi-continuous function on \mathcal{D} for each θ, x, y . Then the above assumption is implied by (a) and $\lim_{d \rightarrow \infty} L(d, \theta, x, y) = \infty$. Weaker conditions on $\lim_{d \rightarrow \infty} L(d, \theta, x, y)$ will also suffice.)

(c)
$$\int |x| W(\delta_0(x, y), x, y) p(x, y) dx \nu(dy) < \infty .$$

(d)
$$\int_0^\infty d\lambda \{ \sup_{\delta \in \mathcal{S}} \int \nu(dy) \int_{-\lambda}^\lambda [W(\delta_0(x, y), x, y) - W(\delta(x, y), x, y)] p(x, y) dx \} < \infty .$$

[The proof of the above result follows exactly the proof of Theorem 2.1.1 of Brown (1966). The main differences are notational. Two additional points to note are:

(i) In Brown (1966) it was true that $\mathcal{D} = \mathcal{R}$, while here \mathcal{D} is allowed to be a more general topological space. Hence those statements in Brown (1966) which concern the topology on \mathcal{D} need to be altered in the obvious manner in order to accommodate this additional generality.

(ii) The converse statement in (b) suffices to guarantee that the analog of the conclusion (2.1.19) of Brown (1966) is valid.]

(I) We note as in Brown (1966, pages 1093–1094) that if the expression $P(d(x - \theta) | y)$ replaces $p(x - \theta, y) dx$ on the right of (1) then it is possible to conclude that δ_0 is almost admissible under the appropriate regularity conditions. It is not, in general, possible to conclude that δ_0 is admissible.

(II) There is another almost-admissibility version of the theorem which is sometimes of interest. If the unicity in Assumption (a) fails, but the other assumptions are satisfied, then δ_0 is almost admissible. (The topology on \mathcal{D} need not be Hausdorff for this result to be valid; but it still must be first countable in order for the topological analog of the statement of (2.1.18) of Brown (1966) to be used to imply the analog of (2.1.21) of Brown (1966). For more details see Brown and Fox (1974).)

Condition (c) states that X has one more moment than what is needed for $R(\theta, \delta_0) < \infty$. For the problem of estimating θ , Perng (1970) has given an example in which (c) alone is violated and δ_0 is inadmissible. A similar example has been given by Fox and Perng (1969) for the testing problem treated by Lehmann and Stein (1953).

Problems with additional parameters. Suppose the parameter space is $\{(\theta, \varphi) : \theta \in \mathcal{R}, \varphi \in \Phi\}$ and the distribution in (1) is replaced by

$$(3) \quad \Pr_{\theta, \varphi} \{X, Y \in S\} = \iint_S p_\varphi(x - \theta, y) dx h_\varphi(y) \nu_1(dy)$$

where p and h are both jointly measurable functions. Suppose the loss function—call it L' —is a location invariant function, i.e.

$$(4) \quad L'(d, (\theta, \varphi), x, y) = W'(g_{-\theta}d, \varphi, x - \theta, y).$$

The basic result does not necessarily characterize all admissible invariant procedures under suitable regularity conditions but it does characterize some, as described below.

Let P be any prior probability distribution on Φ . Let δ_0^P denote a procedure which is Bayes relative to P with respect to the class of (location) invariant procedures. Without loss of generality we assume that ν_1 has been chosen so that $\int h_\varphi(y)P(d\varphi) > 0$ a.e. (ν_1) . Under suitable measurability conditions (see Brown and Purves (1973)) δ_0^P will satisfy

$$(5) \quad R'(P, \delta_0^P | y) = \inf_{d \in \mathcal{D}} R'(P, g_x d | y)$$

where

$$R'(P, g_x d | y) = \iint W'(g_x d, \varphi, x, y) p_\varphi(x, y) dx P(d\varphi).$$

δ_0^P will then be essentially uniquely determined if and only if the infimum in (5) is uniquely determined a.e. (ν_1) .

Set

$$(6) \quad p(x, y) = \int p_\varphi(x, y) h_\varphi(y) P(d\varphi) / \int h_\varphi(y) P(d\varphi)$$

and

$$(7) \quad \nu(A) = \int_A (\int h_\varphi(y) P(d\varphi)) \nu_1(dy)$$

and

$$(8) \quad L(d, \theta, x, y) = \frac{\int L'(d, (\theta, \varphi), x, y) p_\varphi(x - \theta, y) h_\varphi(y) P(d\varphi)}{\int p_\varphi(x - \theta, y) h_\varphi(y) P(d\varphi)}.$$

$p, \nu,$ and L as given above define a location invariant problem like that treated in our basic result. Hence under appropriate regularity conditions the best invariant procedure for this problem will be admissible for this problem. Elementary computations show that this best invariant procedure is in fact δ_0^P , and that δ_0^P admissible in the problem defined by (7), (8), (9) implies that δ_0^P is admissible in the original problem defined by (3), (4); and conversely. Hence the basic result may be used to conclude admissibility of invariant procedures which are Bayes relative to the class of invariant procedures. Note that one of the regularity conditions is that δ_0^P be uniquely determined a.e. (ν) .

Certain cases of the above perhaps demand some special comment. If L' does not depend on φ then φ is a “nuisance parameter.” More interestingly, if L' does not depend on θ then θ is a “nuisance location parameter.” Assuming that

L' does not depend on x —which is virtually always the case in applications—this means that the group $\{g_z\}$ does not really play any role at all since the formulation (4) then demands that $g_z d = d$. Thus $\{g_z\}$ always acts trivially in such situations.

Suppose Φ is a compact topological space and the risk functions of all admissible invariant procedures are *continuous* on Φ (these risk functions are of necessity independent of $\theta \in \mathcal{R}$). Then any invariant procedure which is admissible relative to the class of invariant procedures must be Bayes relative to that class for some prior P on Φ . Hence all admissible invariant procedures will be of the form δ_0^P for some distribution P on Φ .

The above considerations may be somewhat generalized. For example we have made the assumption here that the marginal distributions of Y given φ are absolutely continuous with respect to some measure $-\nu_1$. This assumption is unnecessary, and we have made it here only to avoid having to define and use the appropriate (measurable) conditional probability distributions in (6), (7), (8).

Following the next general paragraph we sketch some more specific problems to which the above results apply or to which its two dimensional analog in Brown and Fox (1974) applies. The appropriate regularity conditions are satisfied in all these examples, but we will not pause here to explicitly state or check conditions on the loss functions involved which would suffice to imply fulfillment of the appropriate regularity conditions.

Scale parameter problems. It was pointed out in Farrell (1964) how scale parameter problems may be transformed to location parameter problems. Since many of the results given below require this transformation we remind the reader how it goes. Suppose $X' > 0$, Y have the distribution

$$P_{\theta'}\{X', Y \in S\} = (\theta')^{-1} \int \int_S p(x'/\theta', y) dx' \nu(dy), \quad \theta' > 0$$

and the loss function—call it L'' —is scale invariant where the group acting on \mathcal{D} is denoted as $\{g_{\theta'}\}$. Let $X = \ln X'$, $\theta = \ln \theta'$, and $g_{\theta} = g'_{\theta'}$. Then X, Y is a location family with location parameter θ and with invariant loss function L defined in the natural way from L'' by $L(d, \theta, x, y) = L''(d, e^{\theta}, e^x, y)$. Evidently the risk function of a procedure in the original scale problem equals the risk function of its transform in the equivalent location problem. Invariant procedures are transformed into invariant procedures. Hence a best invariant procedure is admissible in the scale problem if and only if a best invariant procedure is admissible in the transformed problem.

Estimating a scale parameter with an unknown location parameter. (The following problem was also treated by Portnoy (1971) using different techniques.) Suppose the variables U, V, Z ($U \in (0, \infty)$, $V \in \mathcal{R}$) have joint distribution

$$\Pr_{\mu, \sigma}(U, V, Z \in S) = \int \int_S \int \sigma^{-2} q(u/\sigma, v - \mu/\sigma, z) du dv \nu_2(dz), \\ \sigma > 0, -\infty < \mu < \infty.$$

Suppose $\mathcal{D} = (0, \infty)$ and the loss function is of the form $L(d, (\mu, \sigma)) = W'(d/\sigma)$. Assume W' is "reasonable" so that assumptions (a)—(d) will be satisfied. Let P be a probability distribution on $\{\mu/\sigma\} = \mathcal{R}$. Making the transformation $X = \ln U, Y = (U/V, Z)$ results in a location parameter problem with a nuisance parameter μ/σ . It follows from the previous remarks that in the original location-scale problem the scale invariant procedure which is Bayes relative to P with respect to the class of scale invariant procedures is admissible.

Confidence set procedures, geometric criteria. Suppose X, Y are as in (1) and suppose the elements of \mathcal{D} are closed subsets of \mathcal{R} . \mathcal{D} can be made into a metric space by providing it with the Hausdorff distance. (See for example, Engelking (1969, pages 201, 163).) Assume \mathcal{D} is locally compact and σ -compact in this topology. Three among many possible choices for \mathcal{D} are the collection of all closed subsets of \mathcal{R} , the collection of all non-empty closed subsets of \mathcal{R} , or the collection of all closed intervals $[v_1, v_2] \subset \mathcal{R}, v_1 \leq v_2$ (or $v_1 < v_2$). In the latter case, according to the given topology, $d_i \rightarrow d$ if and only if $v_{1i} \rightarrow v_1$ and $v_{2i} \rightarrow v_2$.

Suppose the loss function is

$$L(d, \theta, x - \theta, y) = (1 - \chi_d(\theta)) + h(d)$$

where $h(d)$ is lower semi-continuous on \mathcal{D} and "reasonable." Then the best invariant (confidence interval) procedure is admissible if Assumptions (a)—(d) are satisfied, which they generally will be. Two possible choices for h are the diameter of the set $d (= \sup \{|x_2 - x_1| : x_1, x_2 \in d\})$ or the Lebesgue measure of d . Both of the resulting loss functions are lower semi-continuous if $\mathcal{D} = \{[v_1, v_2] : v_1 \leq v_2\}$, but only the first of these is lower semi-continuous if \mathcal{D} is the collection of all non-empty closed subsets of \mathcal{R} . Some further comments and generalizations are contained in Cohen and Strawderman (1973).

Similar remarks apply to two-dimensional location-parameter confidence set problems. In such problems a reasonable choice for \mathcal{D} for which Lebesgue measure on \mathcal{R}^2 is a lower semi-continuous function is the collection of all compact convex subsets of \mathcal{R}^2 . Of course in two dimensions it is the regularity conditions of Brown and Fox (1974) which apply rather than (a)—(d) above. The above comments contain several of the results which were first proved in Joshi (1966).

Various generalizations of the above are possible. For example one could add an unknown nuisance parameter to the problem—for example, a scale parameter. It would then follow that invariant procedures which are Bayes for some prior on the nuisance parameter relative to the class of invariant procedures are admissible.

Another type of confidence problem for real location parameters has $\mathcal{D} = \{i \ni i : \mathcal{R} \rightarrow [0, 1]\}$. Here $i(\theta)$ is to be interpreted as the probability that the point θ is included in the confidence interval. Typically $L(d, \theta, x, y) = (1 - i(\theta)) + \int_{-\infty}^{\infty} i(t) dt$. Such a problem was treated in Joshi (1969), (1970). Give \mathcal{D} the

weak topology derived from the L_∞ pseudo-norm $\|i\| = \text{ess sup}_\theta |i(\theta)|$, (i.e. $j_i \rightarrow k$ if, and only if, $\int_a^b j_i(t) dt \rightarrow \int_a^b k(t) dt$ for all a, b such that $-\infty < a < b < \infty$.) Almost admissibility of the best invariant procedure can then be deduced in this type of problem from Statement II following the general theorem if Assumptions (b)—(d) are satisfied, which will generally be the case.

Multiple decision problems with a nuisance location parameter. Suppose X, Y have the distribution (3), \mathcal{D} is any measurable space, and the loss function is a function of d, φ , and y only. Then, as previously noted, under suitable regularity conditions any invariant procedure which is Bayes for some prior on Φ relative to the class of invariant procedures is admissible. We want only to point out here that if $\Phi = \{0, 1\}$ and $\mathcal{D} = \{0, 1\}$ then this problem is a location invariant testing problem, and the results derived here imply those in Lehmann and Stein (1953). Similarly the "multiple decision problem" where Φ and \mathcal{D} are finite sets was treated in Fox (1971). In both the above cases Φ is compact, etc., and hence (again subject to the regularity conditions) one may characterize all of the admissible invariant procedures. Assumptions (a)—(d) in the case of finite \mathcal{D} and Φ reduce to (b) and $\int |x|L'(d, \varphi, y)p_\varphi(x, y)h_\varphi(x)\nu(dy) < \infty$ for each $\varphi \in \Phi$, $d \in \mathcal{D}$.

Two particularly important special cases follow.

Univariate analysis of variance. As in the usual canonical analysis of variance setup let $Z_i, i = 1, \dots, p$ be independent normal variables with means $\mu_i, i = 1, \dots, p$, and variance σ^2 . It is known that $\mu_i = 0$ for $i = r + 1, \dots, p$ and it is desired to test the null hypothesis $H_0: \mu_i = 0$ for $i = 1, \dots, r$. This is the case where *no nuisance means are present*. Consider the (restricted) alternative $H_1^{(k)}: \sum_{i=1}^r \mu_i^2/\sigma^2 = k$ where k is a fixed constant, $k \neq 0$. Sufficiency reduces the problem to the variables $X' = (\sum_{i=r+1}^p Z_i^2)^{1/2}$ and $Y = (Z_1/X', \dots, Z_r/Z')$. This is a scale parameter situation with additional parameters $\varphi_1 = \mu_1/\sigma, \dots, \varphi_r = \mu_r/\sigma$. The usual procedure—i.e. to reject H_0 if and only if $\sum_{i=1}^r Y_i^2 = (\sum_{i=1}^r Z_i^2)/(\sum_{i=r+1}^p Z_i^2) > K$ —is Bayes relative to the class of scale invariant procedures for the prior for which under H_0 $\varphi_1, \dots, \varphi_r$ is of course concentrated on the point $0, \dots, 0$; and under H_1 $\varphi_1, \dots, \varphi_r$ has the spherically symmetric distribution on the sphere $\sum \varphi_i^2 = k$.

Hence the usual procedure is admissible against the restricted alternative $H_1^{(k)}$. It is thus also admissible against the more usual alternative $H_1: \sum_{i=1}^r \mu_i^2 \neq 0$. This latter admissibility result has been proved by other methods; for example Stein (1956) and Kiefer and Schwarz (1965).

Testing the location of a quantile. This example was suggested by the referee. Let the Z_i be as in the previous example with $r = 1$. We wish to test $H_0': \mu_1 - \eta\sigma = k$ versus $H_1': \mu_1 - \eta\sigma \neq k$ (η known). Replacing Z_1 by $Z_1 - k$ we reduce the problem to testing $H_0: \mu_1/\sigma = \eta$ versus $H_1: \mu_1/\sigma \neq \eta$ using observations $Z_1 - k, Z_2, \dots, Z_p$.

For the one-sided alternative, Lehmann and Stein (1953) proved the one-sided

noncentral t -test is admissible. We will show that, in our case, the two-sided noncentral t -test is admissible.

The test statistic is $T = (Z_1 - k)/(\sum_{i=2}^p Z_i^2)^{1/2}$. See Lehmann (1959, page 230) for the details of the test. For $\eta_1 < \eta < \eta_2$ consider the restricted alternative $H_1^{(\eta_1, \eta_2)}: \mu/\sigma = \eta_1$ or η_2 . Proceeding as in the previous example, for an appropriate prior distribution on φ_1 over the values η, η_1, η_2 the noncentral t -test is Bayes relative to the class of scale invariant procedures and, hence, is admissible for the problem of testing H_0 versus $H_1^{(\eta_1, \eta_2)}$. Admissibility in the original problem then follows.

The methods of this paper can be applied to a variety of other problems involving location or scale parameters and additional nuisance parameters. In addition to the above applications and to the basic classical problem of estimating a location or scale parameter such as treated in Stein (1959), Farrell (1964), and Brown (1966) and the basic testing situation treated in Lehmann and Stein (1953), Fox (1971) and Brown and Fox (1974) some possible applications are the following: estimating the largest mean (e.g. in a multivariate normal situation with known variance), estimating the p th percentile point (of a distribution in a location parameter family), and estimation and testing problems involving a parameter of a geometric or exponential density with unknown location parameter. Versions of these problems are formulated in Blumenthal and Cohen (1968), Cohen and Sackrowitz (1970), Zidek (1971), and Klotz (1970) and various results concerning them are given. Mainly the given results are different from those obtainable by the methods of this paper.

REFERENCES

- [1] BLUMENTHAL, S. and COHEN, A. (1968). Estimation of the larger translation parameter. *Ann. Math. Statist.* **39** 502-516.
- [2] BROWN, L. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087-1136.
- [3] BROWN, L. and FOX, M. (1974). Admissibility of procedures in two-dimensional location parameter problems. *Ann. Statist.* **2** 248-266.
- [4] BROWN, L. and PURVES, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1** 902-912.
- [5] COHEN, A. and SACKROWITZ, H. (1970). Estimation of the last mean of a monotone sequence. *Ann. Math. Statist.* **41** 2021-2034.
- [6] COHEN, A. and STRAWDERMAN, W. (1973). Admissible confidence interval and point estimation for translation or scale parameters. *Ann. Statist.* **1** 545-550.
- [7] ENGELKING, R. (1968). *Outline of General Topology*. Wiley, New York.
- [8] FARRELL, R. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.* **35** 949-998.
- [9] FOX, M. (1971). Admissibility of certain location invariant multiple decision procedures. *Ann. Math. Statist.* **42** 1553-1561.
- [10] FOX, M. and PERNG, S.K. (1969). Inadmissibility of the best invariant test when the moment is infinite under one of the hypotheses. *Ann. Math. Statist.* **40** 1483-1485.
- [11] JOSHI, V. M. (1966). Admissibility of confidence intervals. *Ann. Math. Statist.* **37** 629-638.
- [12] JOSHI, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40** 1042-1067.

- [13] JOSHI, V. M. (1970). Admissibility of invariant confidence procedures for estimating a location parameter. *Ann. Math. Statist.* **41** 1568–1581.
- [14] KIEFER, J. and SCHWARZ, R. (1965). Admissible Bayes character of T^2 , R^2 and other fully invariant tests for classical multivariate normal problems. *Ann. Math. Statist.* **36** 747–770.
- [15] KLOTZ, J. (1970). The geometric density with unknown location parameter. *Ann. Math. Statist.* **41** 1078–1083.
- [16] LEHMANN, E. and STEIN, C. (1953). The admissibility of certain invariant statistical tests involving a translation parameter. *Ann. Math. Statist.* **24** 473–479.
- [17] PERNG, S. K. (1970). Inadmissibility of various “good” statistical procedures which are translation invariant. *Ann. Math. Statist.* **41** 1311–1321.
- [18] PORTNOY, S. L. (1971). Formal Bayes estimation with application to a random effects analysis of variance model. *Ann. Math. Statist.* **42** 1379–1402.
- [19] STEIN, C. (1956). The admissibility of Hotelling’s T^2 -test. *Ann. Math. Statist.* **27** 616–623.
- [20] STEIN, C. (1959). The admissibility of Pitman’s estimator of a single location parameter. *Ann. Math. Statist.* **30** 970–979.
- [21] ZIDEK, J. (1971). Inadmissibility of a class of estimators of a normal quantile. *Ann. Math. Statist.* **42** 1444–1447.
- [22] ZIDEK, J. (1973). Estimating the scale parameter of the exponential distribution with unknown location. *Ann. Statist.* **1** 264–278.

STATISTICS CENTER
RUTGERS UNIVERSITY
NEW BRUNSWICK, N. J. 08903

DEPARTMENT OF STATISTICS
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48823